UC Irvine
UC Irvine Previously Published Works

Title

Incorporating testing volume into estimation of effective reproduction number dynamics.

Permalink

Journal

ISSN

Authors

Goldstein, Isaac
Wakefield, Jon
Minin, Volodymyr

Publication Date

DOI

Peer reviewed

Incorporating testing volume into estimation of effective reproduction number dynamics

Isaac H. Goldstein[1], Jon Wakefield[2] and Volodymyr M. Minin[1] (ID)

[1]Department of Statistics, University of California, Irvine, CA, USA
[2]Departments of Biostatistics and Statistics, University of Washington, Seattle, WA, USA

Address for correspondence: Volodymyr M. Minin, Donald Bren Hall 2068 Box 354322 Irvine, CA 92697-1250, United States.
Email: vminin@uci.edu

Abstract

Branching process inspired models are widely used to estimate the effective reproduction number—a useful summary statistic describing an infectious disease outbreak—using counts of new cases. Case data is a real-time indicator of changes in the reproduction number, but is challenging to work with because cases fluctuate due to factors unrelated to the number of new infections. We develop a new model that incorporates the number of diagnostic tests as a surveillance model covariate. Using simulated data and data from the SARS-CoV-2 pandemic in California, we demonstrate that incorporating tests leads to improved performance over the state of the art.

Keywords: Bayesian inference, COVID-19, infectious diseases, nowcasting

1 Introduction

In an infectious disease epidemic, the effective reproduction number is the average number of people a newly infected person will subsequently infect. When the effective reproduction number is above 1, an epidemic is out of control and will continue to grow, vice versa if it is below 1. This makes the effective reproduction number a useful summary of the state of an epidemic which can provide guidance to policy makers. As such, estimates of the effective reproduction number based on observed data can be an important part of any public health response during an epidemic. Recent examples from the SARS-CoV-2 pandemic include work by Mishra et al. (2022) in Scotland, as well as efforts by Swiss National Covid-19 Science Task Force (2020).

An early effort of using a likelihood-based approach to estimate the effective reproduction number is that of Wallinga and Teunis (2004), which is based on modelling transmission trees. A recently popular class of estimators for the effective reproduction number (used in both Mishra et al., 2022; Swiss National Covid-19 Science Task Force, 2020) is inspired by stochastic branching process models, where infectious individuals infect a random number of new individuals at random points in time. The most widely used model in this class is available in the `EpiEstim` R package (Cori et al., 2013; Thompson et al., 2019), which is based on ideas put forth by Fraser (2007). `EpiEstim` assumes all new infections (incidence) are observed, and uses a time series of observed cases as data. During the SARS-CoV-2 pandemic, a number of methods in this class of estimators have been developed. The methods of Parag (2021) and Capistrán et al. (2022) continue to assume incidence (or incidence up to a constant) are observed, and focus on improving how changes in R_t are modelled over time, while avoiding Markov chain Monte Carlo (MCMC)-based methodologies. The methods of Abbott et al. (2020), Huisman et al. (2022), Scott et al. (2021), and Bhatt et al. (2023) use more computationally intensive approaches which model observed data as functions of latent incidence, either through explicit Bayesian models

(Abbott et al., 2020; Bhatt et al., 2023; Scott et al., 2021) or through a pipeline that first bootstraps latent incidence which is then used as input into EpiEstim. The methods of Teh et al. (2022), Scott et al. (2021), and Bhatt et al. (2023) also begin to tackle the problem of how to estimate R_t across spatial locations. Many of these methods have not been scrutinised via extensive simulation studies under model mis-specification, or in some cases, not probed at all, making it difficult to understand the strengths and weaknesses of this class of methods.

This gap in knowledge is particularly relevant when it comes to applying such methods to observed case counts of an infectious disease. As the SARS-CoV-2 pandemic has demonstrated, observed cases of an infectious disease are often circuitously related to the true number of new infections, due to constraints in testing supply, asymptomatic infections, testing eligibility, and reporting delays. These factors can make estimating the effective reproduction number from cases quite difficult in real world situations. This is a widely recognised challenge; a recent survey of papers using EpiEstim found the most common challenge for users was dealing with the quality of observed case data (Nash et al., 2022). One sensible approach to resolving this issue is to use other sources of data. For instance, Flaxman et al. (2020) used a model similar to those available in epidemia to assess the effects of non-pharmaceutical interventions by fitting a model to death counts rather than case counts, while Mishra et al. (2022) incorporated data sources such as deaths and sero-prevalence data in addition to case data. Turning to other data sources is an appealing strategy for retrospective analyses, but during an ongoing epidemic it is often desirable to provide real-time estimates of the effective reproduction number, a task called now-casting. When now-casting, case data is one of the earliest available data sources to indicate a change in the effective reproduction number. It behooves us, then, to develop reasonable methods for using case data when estimating the effective reproduction number, despite the difficulties involved.

Our study has two main contributions. First, we develop our own model for estimating the effective reproduction number making different modelling choices than other available methods. The most significant of these is that we incorporate the number of diagnostic tests administered (both positive and negative) as a covariate in our model. Second, to increase understanding of the broader class of branching process inspired methods, we conduct simulation studies comparing our new model to EpiEstim and a model constructed using the epidemia package developed by Scott et al. (2021). The latter approach builds on the EpiEstim framework by allowing for more flexible and complex models that treat new infections as unobserved variables, with various time series such as cases or deaths modelled as noisy realisations of unobserved infections used as data (Bhatt et al., 2023; Scott et al., 2021). In particular, we explore scenarios with differing diagnostic test availability. We also fit our model to real data from the SARS-CoV-2 pandemic in 15 California counties. Our results show that our new model outperforms existing methodologies under a variety of different testing scenarios and provides novel insights when applied to real data, highlighting the utility of incorporating tests when using case data as well as distributional choices made in the modelling process.

2 Methods

2.1 Available data

Consider an outbreak observed for a total of T time intervals. We restrict ourselves to two kinds of infectious disease outbreak data. The first is the time series of observed cases, $\mathbf{O} = (O_1, O_2, O_3, \ldots, O_T)$, where O_u is the number of newly observed cases of an infectious disease during time interval u. The second is the time series of diagnostic tests, $\mathbf{M} = (M_1, M_2, \ldots, M_T)$, where M_u is the total number of diagnostic tests administered during time interval u. For this study, we assume tests are perfectly accurate. We also do not model the total number of tests performed, but rather model the number of positive tests conditioned on the total number of tests. We assume that O_u is a noisy realisation of recent latent unobserved new infections (incidence); denoted by I_u during time interval u.

2.2 Modelling incidence

We first differentiate between incidence during the observation period, when case data is available, and incidence prior to the observation period. It is rare in practice to begin analysis of an infectious disease epidemic at the exact start of the epidemic. We follow Scott et al. in modelling a number of

unobserved incidence values (often called seeded incidence) drawn from a hierarchical exponential model (Bhatt et al., 2023; Scott et al., 2021). That is, for $t = -n, \, -n-1, \, \ldots, 0$,

$$\lambda \sim \text{Exponential}(\eta),$$
$$I_t \sim \text{Exponential}(\lambda).$$

We model latent incidence during the observation period as a latent gamma random variable:

$$I_t \mid \mathbf{L}_{-n:t}, R_t \sim \text{gamma}\left(R_t \sum_{u=-n}^{t-1} g_{t-u} I_u v, \, v \right), \quad t = 1, \, \ldots, T,$$

where $\mathbf{L}_{-n:t}$ is the set of all previous incidences between times $-n$ and t, g_t is the discretised probability density function of the generation time (the time from an individual becoming infected to infecting someone else) distribution for the interval t, and R_t is the effective reproduction number at time interval t. Parameter v, describing the proportional mean–variance relationship of the above gamma distribution, receives its own prior:

$$\log(v) \sim N(\mu_v, \sigma_v^2).$$

We assume g_t to be known. Svensson (2007), Champredon and Dushoff (2015), and Champredon et al. (2018) have highlighted that in a closed population, the generation time distribution depends on population dynamics, i.e. it changes over time depending on the number of susceptibles available, somewhat similarly to the effective reproduction number. This is not taken into account in our model (nor, to our knowledge, in any model in this class of estimators). Instead, we use the intrinsic generation time distribution which assumes a fully susceptible population.

Note that under this model,

$$\text{E}(I_t \mid \mathbf{L}_{-n:t}, R_t) = R_t \sum_{u=-n}^{t-1} I_u g_{t-u}, \tag{1}$$

$$\text{Var}(I_t \mid \mathbf{L}_{-n:t}, R_t) = R_t \sum_{u=-n}^{t-1} I_u g_{t-u} / v. \tag{2}$$

The assumed mean relationship lies at the heart of branching process inspired methods for estimating the effective reproduction number (Fraser, 2007). Pakkanen et al. (2023) show that equation (1) is justified under a formulation of disease transmission modelled as a variation on the Crump–Mode–Jagers branching process. Regardless of the underlying model, we think it is beneficial to allow for incidence to change stochastically. To this end, we model incidence as an auto-regressive gamma process while preserving the branching process inspired mean model (1). By modelling incidence as a continuous random variable, we are able to use Hamiltonian Monte Carlo (HMC) to approximate the posterior distribution of our model parameters. The mean–variance relationship of the gamma distribution is also somewhat convenient, as it allows for over-dispersion in the variance of incidence through parameter v.

To allow for the effective reproduction number to change over time, we model it as a random walk on the log scale:

$$\log R_1 \sim \text{Normal}(\mu_{r1}, \sigma_{r1}^2),$$

$$\log R_t \mid \log R_{t-1}, \mathbf{L}_{-n:t-1} \sim \text{Normal}\left(\log R_{t-1}, \frac{\sigma^2}{T-1} \right), \quad t = 2, \, \ldots, T.$$

The prior distribution of σ, $\log(\sigma) \sim N(\mu_\sigma, \sigma_\sigma^2)$, is chosen to reflect beliefs about the total amount of possible variation in the effective reproduction number over the course of the observed period.

2.3 Modelling observed cases

Depending on the context of an infectious disease, the relationship between observed cases and incidence can be complex. One challenge relates to testing supply. The number of cases observed is always a function of the number of diagnostic tests administered. In the context of a novel infectious disease, testing supplies may change rapidly as new technologies are developed, approved, and deployed. Thus, we model observed cases (O_t) conditioned on previous and current incidence (I_t, $\mathbf{I}_{-n:t-1}$,) as a negative binomial random variable, where the mean of the negative-binomial random variable is a function of incidence (as in Abbott et al., 2020; Bhatt et al., 2023; Scott et al., 2021), the number of tests administered, and a detection parameter ρ, with over-dispersion parameter κ:

$$\kappa \sim \text{Truncated-Normal}(\mu_\kappa, \sigma_\kappa^2),$$
$$\log \rho \sim \text{Normal}(\mu_\rho, \sigma_\rho),$$
$$D_t = \sum_{j=-n}^{t} I_j d_{t-j}, \tag{3}$$
$$O_t \mid I_t, \mathbf{I}_{-n:t}, \rho, \kappa, M_t \sim \text{Neg-Binom}(\rho \times M_t \times D_t, \kappa), \quad t = 1, \ldots T,$$

where $\rho \times M_t \times D_t$ is the mean of the negative-binomial distribution. As defined above, M_t is the total number of diagnostic tests administered during time interval t. As a result, the detection rate for time t is $\rho \times M_t$, which allows the detection rate to change over time as a function of the number of tests available. With a detection rate which depends on tests, the model can discern between situations where cases increase because of increases in latent incidence, as opposed to increases in the number of tests administered. The weights d_{t-j} are discretised weights of the delay period distribution, that is, the time from infection to detection. Delays occur for a variety of reasons, based on when the difference between when individuals are infected and when they test, as well as delays in reporting the results of the test. In our study, we will use simulations and data where the only delay is caused by our assumption that cases represent individuals transitioning from the latent stage of infection to the infectious stage. Thus, for this study, d_{t-j} are discretised weights of the latent period distribution. Note that we allow for cases observed at time t to come from incidence observed at time t as well; this can be adjusted depending on how quickly a particular disease spreads and at what granularity observations are recorded.

It is difficult to choose generic priors for κ and ρ, as they both depend in some way on properties of the surveillance system used to collect data. We address this challenge in the sections below.

2.4 Prior for case over-dispersion

In our experience, some choices of the prior distribution for κ result in poor MCMC convergence. To overcome this issue, we developed an approach for choosing the prior distribution for κ inspired by Empirical Bayes methods. We fit a Bayesian thin plate regression spline to the time series of cases, assuming a negative-binomial distribution with the mean number of cases being a non-parametricaly estimated function of time, then use the posterior estimate for the over-dispersion parameter to construct the prior for our model (Wood, 2017). We use `brms` (version 2.15.0) to fit the regression spline to observed cases (Bürkner, 2017). This method has drawbacks from a theoretical perspective, because the spline-based model is fit to the same data that is then analysed with our semi-mechanistic model. For simulations, this is easily overcome by fitting the spline to a simulated data set that is then not analysed by our model, this is the approach we took for our simulation study. For real data analysis, one solution is to fit a spline to data from an outbreak occurring in a similar location to the one being analysed. For this study, we put aside theoretical concerns and fit a spline-based model to each real data set used in this study to derive the prior for κ and then applied our model. We choose the parameters of the prior by minimising a squared loss function, searching for prior parameters which minimised the

squared difference between the quantiles of the spline posterior, and the empirical quantiles of the candidate prior distribution.

2.5 Prior for the case detection rate

Choosing the prior for the case detection parameter ρ likewise requires some care, because the meaning of ρ depends on the number of diagnostic tests in the data. We propose the following procedure: first construct a plausible range for what proportion of incidence has been observed. Then, using the the 50% quantile of tests in the observed test time series, construct a prior for ρ which matches the prior for the overall mean case detection rate. In practice, we can construct the prior for ρ using other quantiles as part of sensitivity analyses. For simulations, we use a ρ prior derived using the 50% quantile, and for real data anlysis, we use the 25% quantile which we found improved MCMC convergence.

2.6 Bayesian inference

Let $\mathbf{R} = (R_1, R_2, \ldots, R_T)$ denote the vector of effective reproduction numbers and $\mathbf{I} = (I_{-n}, \ldots, I_T)$ the vector of latent incidence counts. We are interested in the posterior distribution of our model parameters:

$$P(\mathbf{I}, \mathbf{R}, \rho, \kappa, v, \lambda, \sigma \mid \mathbf{O}) \propto P(\mathbf{O} \mid \mathbf{I}, \rho, \kappa)P(\mathbf{I} \mid \mathbf{R}, v, \lambda)P(\mathbf{R} \mid \sigma)\pi(\rho, \kappa, v, \lambda, \sigma).$$

Here, $P(\mathbf{O} \mid \mathbf{I}, \rho, \kappa)$ defines the emissions model, $P(\mathbf{I} \mid \mathbf{R}, v, \lambda)$ defines the latent case model, $P(\mathbf{R} \mid \sigma)$ the random walk prior for the effective reproduction number, and $\pi(\rho, \kappa, v, \lambda, \sigma)$ the prior on all other model parameters.

We use Hamiltonian Monte Carlo, implemented in the R package `rstan` (version 2.21.2), to approximate the above posterior distribution (Stan Development Team, 2020). For the remainder of this study, we will refer to our effective reproduction number estimation method as Rt-estim-gamma.

2.7 State-of-the-art methods

`EpiEstim` models observed cases as incidence, and assumes that

$$I_t \mid I_1, \ldots, I_{t-1}, R_t \sim \text{Poisson}\left(R_t \sum_{u=1}^{t-1} I_u g_{t-u}\right).$$

To facilitate smooth estimates, the effective reproduction number is assumed to be fixed for a given period of time, and then repeatedly estimated for all such periods in the data set. We choose a period size of 1 week, and allow for an uncertain generation time, re-fitting the model using different values for g_{t-u} (see Cori et al., 2013 for details). The prior on the effective reproduction number for each window is a gamma distribution with shape parameter 1 and scale parameter 5.

Using R package `epidemia` (version 1.0.0), we created the Rt-estim-normal model. In this model, latent incidence is an auto-regressive normal random variable with variance equal to the mean multiplied by an over-dispersion parameter so the mean–variance relationship is the same as in our autoregressive gamma model. We model cases as a negative-binomial random variable, using the latent period distribution as the delay distribution (though in `epidemia` it is assumed cases cannot be generated from the current latent incidence). The case detection prior is chosen to reflect a range of plausible values for case detection depending on the simulation scenario and real data. For observed cases, we attempted to use a prior for the over-dispersion parameter that had similar values to the prior used in our model for the over-dispersion parameter of the negative-binomial distribution, but found this led to issues with MCMC convergence. As such, we use the default prior for the inverse of the over-dispersion parameter implemented in `epidemia`. All other priors used are default priors from the `epidemia` package. For a full description of Rt-estim-normal, see online supplementary Section 1.3.

2.8 `EpiEstim` as an autoregressive generalised liner model

Under the basic `EpiEstim` modelling framework, the only value in equation (1) which is random is R_t. Consequently, `EpiEstim` can be mimicked via Poisson regression with an identity link and no intercept. This raises the possibility of assessing the presence of over-dispersion in case data using standard statistical methods. To be more explicit, we can rewrite equation (1) in the style of a generalised linear model (GLM):

$$E[I_t \mid \mathbf{I}_{-n:t}, R_t] = \eta = \beta_1 x_1.$$

In this construction, $\beta_1 = R_t$ and $x_1 = \sum_{u=1}^{t-1} I_u g_{t-u}$ is the weighted sum of previous incidence. After choosing an arbitrary number of previous incidences to include in x_1, we can construct x_1 manually for every observed incidence at time t with the requisite number of observed previous incidences. To estimate the effective reproduction number over time, we use Poisson regression repeatedly on subsets of data, where each subset has observations equal to the length of the smoothing period used in `EpiEstim`. For example, we can implement Poisson regression on data sets with four observations, estimating a β_1 which is fixed for those 4 observations. This is equivalent to using a period of 4 in `EpiEstim`. We assign the estimated effective reproduction number to the last date among the 4 observations, and change the settings of `EpiEstim` to match its estimates to the last observation as well. In addition to mimicking `EpiEstim` with a Poisson GLM, we mimic `EpiEstim` using a quasi-Poisson GLM. Using the quasi-Poisson's estimated over-dispersion parameter, we can assess how well the assumed mean variance relationship of the Poisson GLM matches the empirical variance seen in the observed data. We implement both GLM versions of `EpiEstim` and compare to the simplest version of `EpiEstim` using a fixed generation time in order to motivate the use of more complex models. All code and data needed to reproduce the results are available on GitHub at https://github.com/igoldsteinh/improving˙rt.

3 Results

3.1 GLM `EpiEstim` applied to the SARS-CoV-2 outbreak in Orange County, CA

To motivate the use of more complex models for estimating the effective reproduction number, we applied `EpiEstim` and our two GLM mimics of `EpiEstim` to case data from the SARS-CoV-2 outbreak in Orange County, CA, from 17 May 2020 to 15 January 2021. We used a window size of 4 for `EpiEstim` and corresponding data sets with 4 observations for the GLM mimics. Data and effective reproduction number estimates are displayed in Figure 1. The Poisson GLM closely tracks the effective reproduction number trajectory estimated by `EpiEstim`. However, the quasi-Poisson estimate of the effective reproduction number has much wider confidence intervals than the Poisson GLM. This is because the estimated over-dispersion parameter in the quasi-Poisson model ranges from 1.01 to 26851.84. This shows that the Poisson model for incidences may be inadequate, resulting in overconfidence of R_t inference.

3.2 Simulation protocol

Simulated data for this study was generated from a stochastic Susceptilbe, Exposed, Infected Removed (SEIR) model in R (version 4.0.4) using the `stemr` package (version 0.2.0) (Fintzi et al., 2022; R Core Team, 2020). SEIR models generate an infectious disease outbreak at a population level, with the population divided into four compartments: susceptible, exposed (infected but not yet infectious), infectious, and removed (neither infectious nor susceptible). The changes in these compartments are governed by rate parameters which depend on the populations in the compartments. In our simulations, the mean latent period was 4 days, the mean infectious period was 7.5 days. Daily case data was generated from transitions from the E to the I compartment on day t, using a fixed number of tests and a negative-binomial distribution. For all simulations, ρ was set to be 9×10^{-5} and κ was set to be 5.

The basic reproduction number R_0 was given a fixed trajectory, leading to similar R_t trajectories for each realisation of the simulation. More details on the stochastic SEIR model used for simulation are available in online supplementary Subsection S-1.1. Note that the SEIR models used for
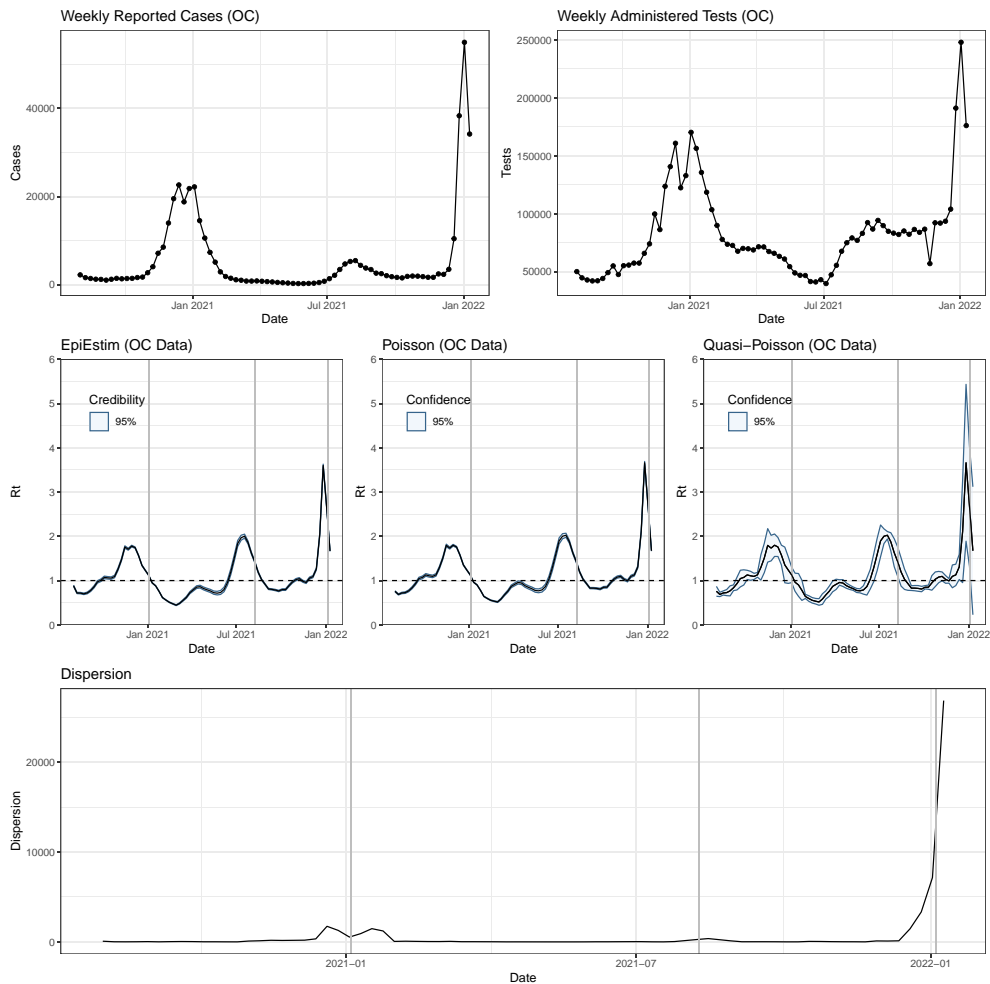
Figure 1. Estimation of the effective reproduction number of SARS-CoV-2 in Orange County, CA, from 2 August 2020 through 15 January 2022. The top row displays the observed cases and total diagnostic tests administered for the period. The middle row displays estimates of the effective reproduction number from `EpiEstim`, and from two GLM-based mimics of `EpiEstim` using Poisson, and quasi-Poisson regression. Shaded regions represent 95% credible or Wald confidence intervals, while lines represent posterior median or point estimates of the effective reproduction number. The final row displays the estimated over-dispersion parameter from the quasi-Poisson regression model. Vertical lines mark the date maximum statewide cases were reported for the original winter 2020 wave, the summer 2021 wave, and the winter 2021 wave. GLM = generalised linear model.

the simulations do not match any of the models used for inference of the R_t trajectories. In other words, all our simulation results are produced in the presence of model mis-specification—a desirable feature for a realistic simulation protocol.

We simulated three separate scenarios lasting 28 weeks, where all parameters were the same except for the number of tests at each time step. In Scenario 1, weekly tests were drawn from a normal distribution with parameters that remained constant over time. In Scenario 2, tests were held constant for the first 6 weeks of the simulation, then increased at varying rates over the next 11 weeks of the simulation. Scenario 3 was similar to Scenario 2, except that testing was held constant for the first 8 weeks, and increased more quickly than in Scenario 2. All simulations were done on a daily time scale, then aggregated into weeks for analysis. The true effective reproduction number for a single week was taken to be the true effective reproduction number of the third day of that week. In all simulations, the first 11 weeks were not analysed, leaving 17 weeks of data for analysis. For each scenario, we generated 100 simulations. Realisations of all three simulations are
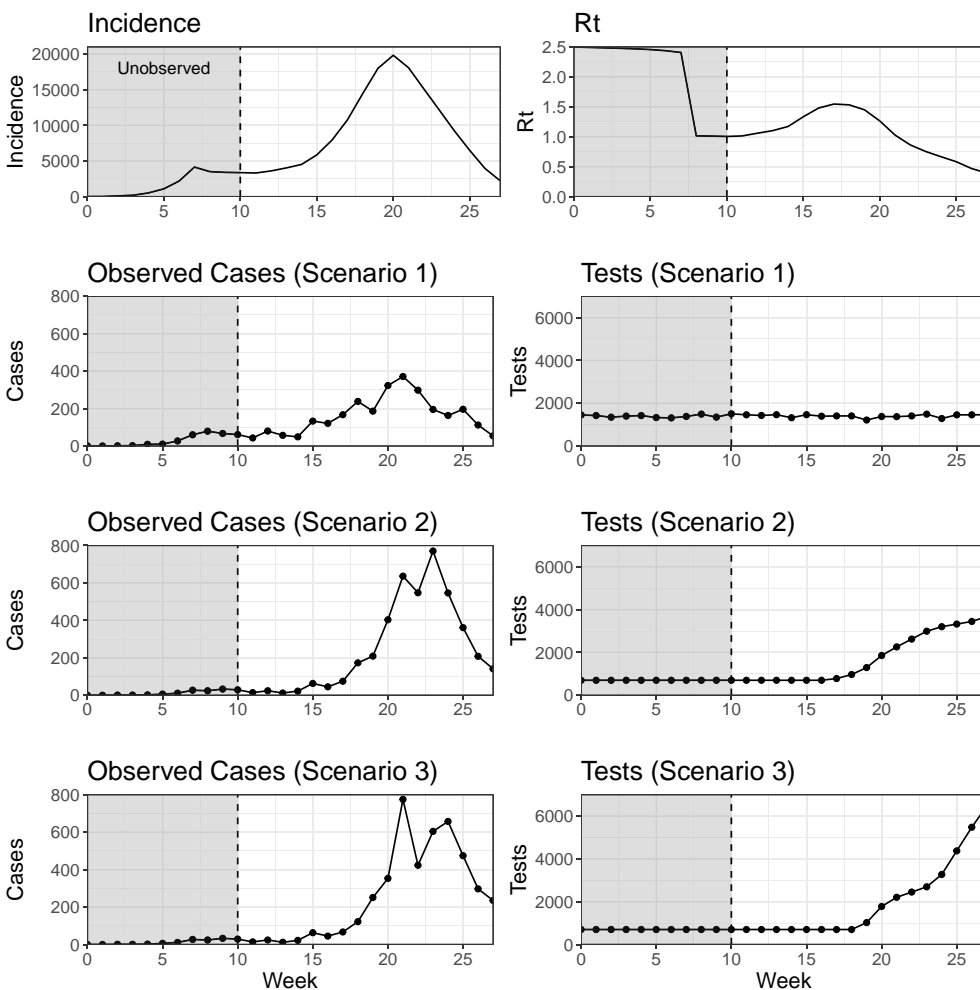
Figure 2. Simulated epidemic data generated from an SEIR model. Cases are generated using an emissions model which includes total diagnostic tests administered as a covariate. Three different testing scenarios are considered, all with the same underlying R0 trajectory. Included are underlying incidence and effective reproduction number trajectories. While these may vary slightly across simulations, they will be very similar due to identical infectious disease dynamics. In Scenario 1, tests are randomly sampled from a normal distribution. In Scenarios 2 and 3, tests stay flat and then increase at varying rates. Simulated epidemics start with 10 individuals and last for 28 weeks. The first 11 weeks are discarded, and are not used when simulated data are analysed by the three effective reproduction number estimation methods.

displayed in Figure 2. The priors used for the Rt-estim-gamma method in the simulation study are displayed in Table 1.

3.3 Simulation results

For each model fit using `rstan`, we sampled 2,000 posterior draws, discarding the first half as burn-in. Figure 3 visualises the estimates for the effective reproduction number from `EpiEstim`, Rt-estim-normal and Rt-estim-gamma for the three data sets visualised in Figure 2. We checked convergence diagnostics for Rt-estim-normal and Rt-estim-gamma for all simulations and ensured adequate convergence of all models. More details are in the online supplementary Subsection S-1.4. Since `EpiEstim` does not provide estimates for the first time point in the series, we report only time points for which all three methods have estimates.

Credible intervals for `EpiEstim` frequently miss the true R_t values (covering between 6 and 9 of the 16 true values), while credible intervals for Rt-estim-normal and Rt-estim-gamma cover most
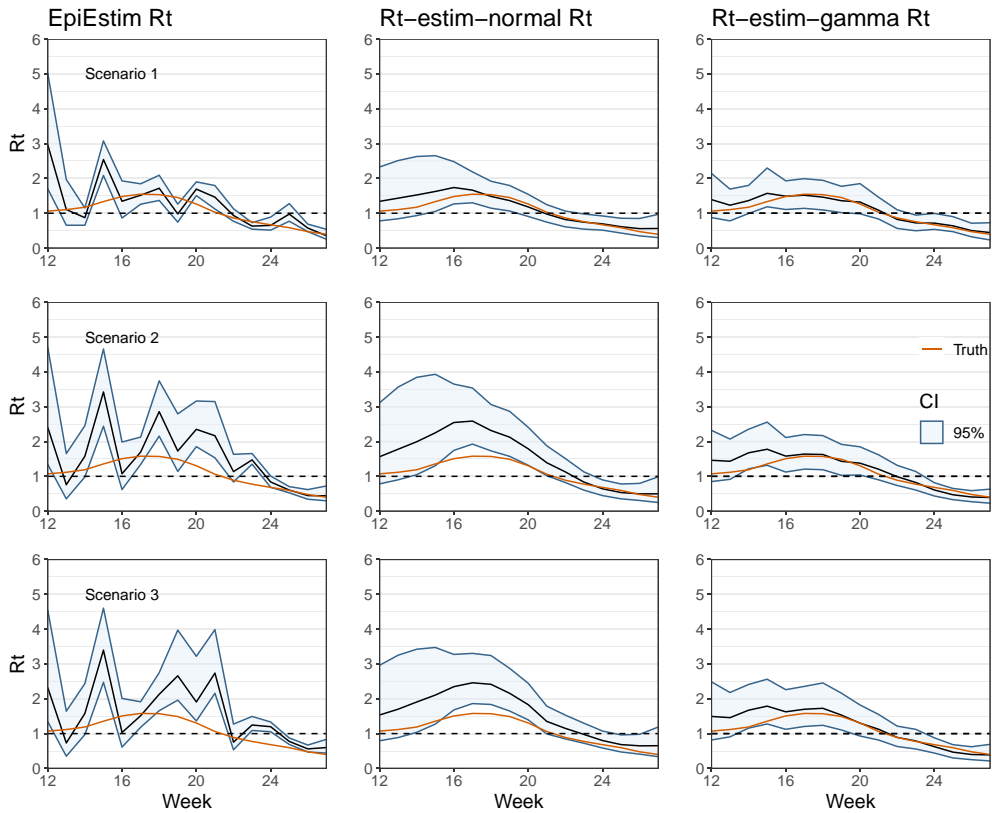
Figure 3. R_t estimation using three different methods for three simulated data sets under different testing scenarios. True R_t trajectories are coloured in red, black lines represent median estimates from the posterior distribution, shaded areas are 95% credible intervals.

true values across simulations. However, Rt-estim-gamma covers more true values than Rt-estim-normal, with narrower credible intervals (ranging between 11 and 16 values for Rt-estim-normal, and 16 values for every scenario for Rt-estim-gamma).

Figure 4 visualises estimates of latent incidence from Rt-estim-normal and Rt-estim-gamma for the three data sets visualised in Figure 2. Rt-estim-normal credible intervals rarely cover the true incidence (covering from 0 to 5 to true values), while Rt-estim-gamma credible intervals generally do (covering 11 to 16 true values).

Posterior predictive distributions for cases for both Rt-estim-normal and Rt-estim-gamma are displayed in online supplementary Figure S2 (the posterior predictive distribution for EpiEstim is not readily available). For all three scenarios, for both models, 95% credible intervals from the posterior predictive distributions cover all observed data points. Rt-estim-gamma had generally narrower credible intervals than Rt-estim-normal.

Because we are using a stochastic SEIR model to generate simulations, each simulation has a different, though similar in shape, true effective reproduction number curve (despite having the same true basic reproduction number curve). The range of true effective reproduction number curves is visualised in online supplementary Figure S1. We report frequentist metrics in order to summarise performance across a variety of different epidemic curves. Model performance on simulated data sets for each of the three models is summarised in Figure 5. For each metric, we summarise results in boxplots where solid lines represent medians, hinges are upper and lower quartiles and whiskers are at most 1.5 times the inter-quartile range from the median. Envelope is a measure of coverage. For each simulation the envelope is the proportion of time points for which a 95% credible interval from the posterior distribution captured the true value of interest. Mean credible interval width (MCIW) is the mean of credible interval widths across time points
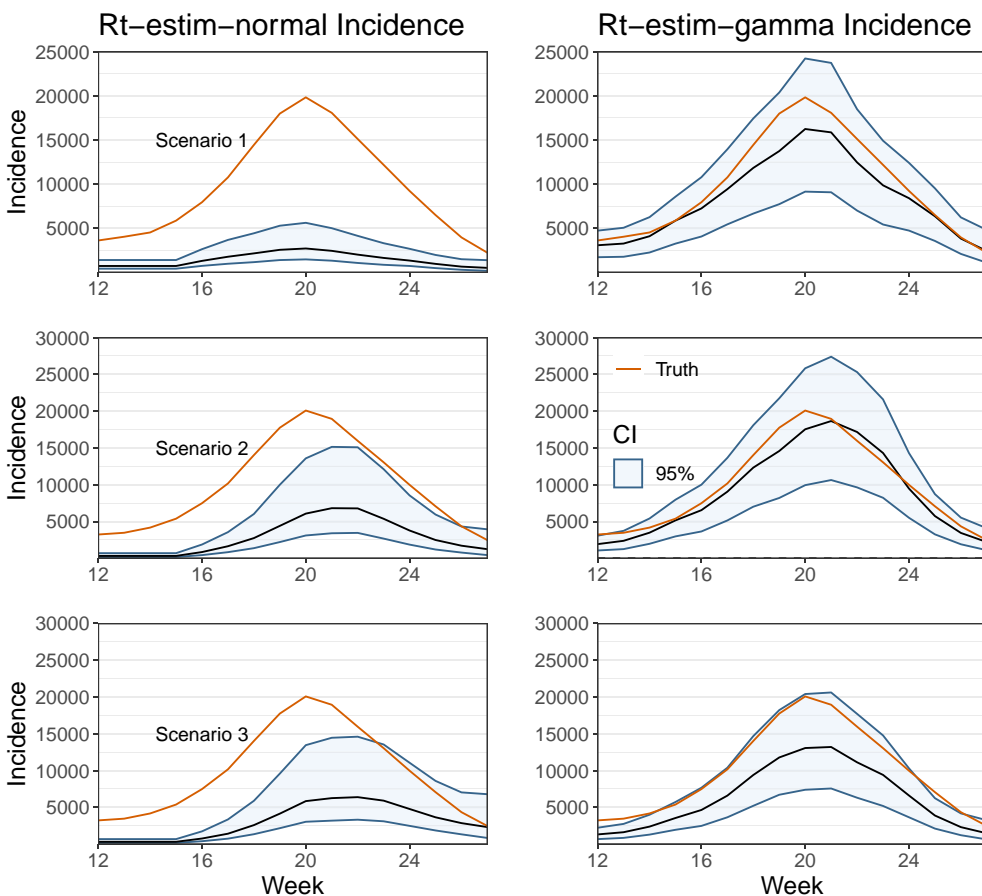
Figure 4. Incidence estimation using two different R_t estimation methods for three simulated data sets under different testing scenarios. True incidence trajectories are coloured in red, black lines represent median estimates from the posterior distribution, shaded areas are 95% credible intervals.

within a simulation. Absolute deviation is a measure of bias, and is the mean of the absolute difference between the posterior median and the true value at each time point. Finally, mean absolute sequential variation (MASV) measures how well each method captured the variation in the effective reproduction number across time by computing the mean of the absolute difference between the posterior median at time point t and the posterior median at time point $t-1$. We compare this to the true MASV in each simulation. EpiEstim had the lowest envelope in all simulation scenarios. Rt-estim-normal had high envelope in Scenario 1 but dropped to lower values in scenarios with time-varying testing supply. Rt-estim-normal had the largest MCIW in all three scenarios, while Rt-estim-gamma had the smallest MCIW in all three scenarios. EpiEstim and Rt-estim-normal had relatively similar values for absolute deviation, Rt-estim-gamma had the smallest absolute deviation in all scenarios. Finally, EpiEstim had the largest MASV in all scenarios, while Rt-estim-normal and Rt-estim-gamma had relatively comparable MASV. For two of three scenarios, Rt-estim-gamma was closer to the true MASV than Rt-estim-normal. We ran three additional experiments using the data sets from Scenario 3 to better understand our model. All results are displayed in online supplementary Figure S4, with the results from Figure 5 included as a baseline comparison. In the first experiment, we halved each parameter in the hypo-exponential distribution and refit the model to the data sets from Scenario 3. This led to narrower credible intervals and lower envelope (see online supplementary Figure S4 for results). In the second experiment, we used a spline fit to the same data being analysed in order to choose a prior for κ. We found
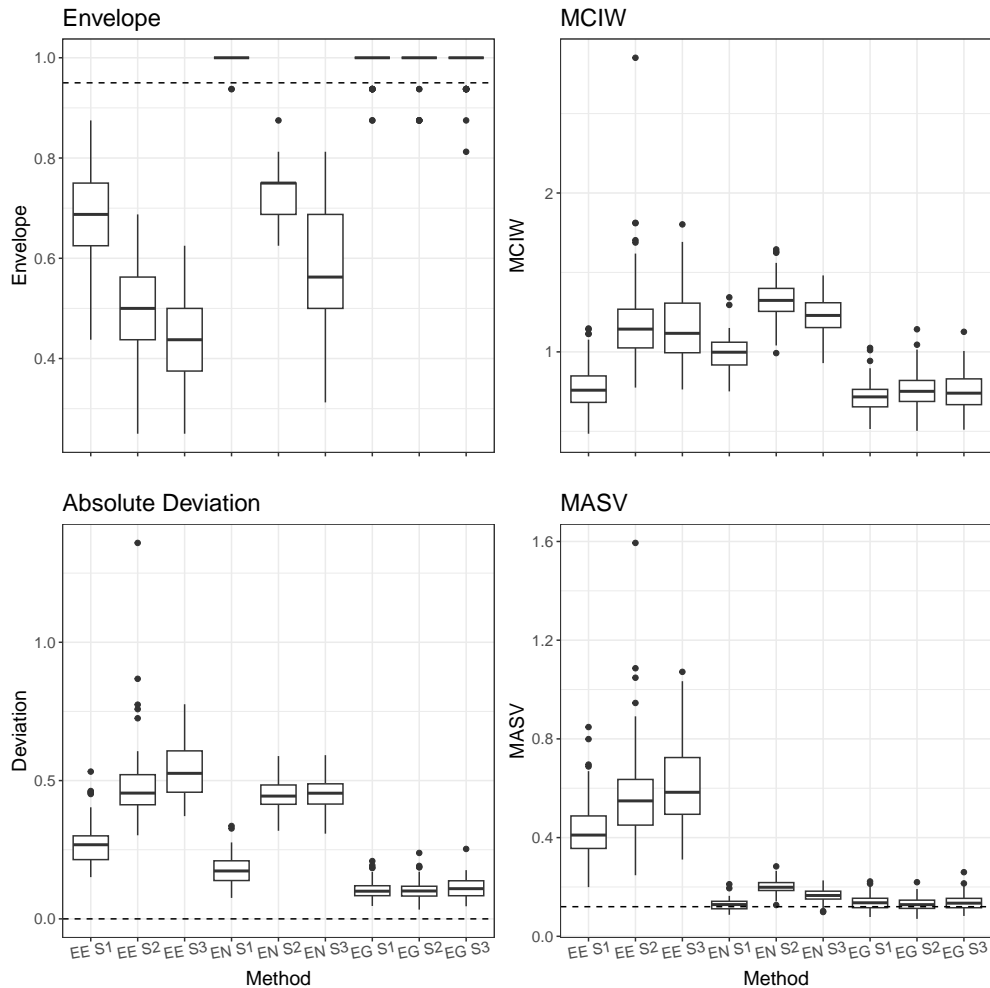
Figure 5. Frequentist metrics for `EpiEstim`, Rt-estim-normal, and Rt-estim-gamma applied to three different simulated epidemics. Envelope is a measure of coverage, taking the average coverage of 95% intervals over the time series. MCIW is the average mean credible interval width. Absolute deviation is the mean of the absolute value of the difference between the median R_t at each time point and the true value. Mean absolute standard deviation is the difference between the current median point estimate for R_t and the previous point estimate for R_t. The dashed line in the bottom right panel represents the true absolute standard deviation. Solid lines represent medians, hinges are upper and lower quartiles and whiskers are at most 1.5 times the inter-quartile range from the median. `EpiEstim` is denoted by EE, Rt-estim-normal by EN, and Rt-estim-gamma by EG, while scenarios are marked S1, S2, and S3. EE S1 describes results of using `EpiEstim` in simulation scenario 1.

no meaningful difference in performance. In the third experiment, we used a prior for ρ derived from the 25% quantile of tests, rather than the 50% quantile, this again led to no meaningful difference in performance with regard to estimating the effective reproduction number, though we expect it to change estimates of incidence. Overall, we find that our Rt-estim-gamma model outperforms the `EpiEstim` and Rt-estim-normal in all metrics, surprisingly even in Scenario 1, where the number of diagnostics tests did not vary appreciably over time (Rt-estim-gamma and Rt-estim-normal have similarly high envelope values in this case).

3.4 Estimating the effective reproduction number of SARS-CoV-2 in 15 California counties

We analysed SARS-CoV-2 reported case data from 15 California counties representing Northern California (Alameda, Sacramento, San Francisco, Santa Clara, Contra Costa), Central California

(Fresno, Merced, Monterey, Stanislaus, Tulare), and Southern California (Los Angeles, Orange, Riverside, San Bernardino, San Diego). These counties represent more than 75% of the population of California, and differ widely along demographic, economic, and political characteristics. We analysed data from 2 August 2020 through 15 January 2022. Data are publicly available from the California Open Data Portal (California Open Data Portal, 2023). Positive cases are associated with the date of their test, rather than the date they were reported.

To estimate the effective reproduction number of SARS-CoV-2, we must choose a generation time distribution to use in our models. Estimating intrinsic generation times from observed data is non-trivial (Park et al., 2021). Early efforts from Ferretti et al. (2020) and Ganyani et al. (2020) estimate the mean generation time to be between 5.5 and 5.2 days, respectively. A more recent estimate of the mean intrinsic generation interval for the original version of SARS-CoV-2 estimated it to be 9.7 days (Sender et al., 2021), but the issue of optimal generation time inference seems far from settled. An additional complication is that a number of important variants of SARS-CoV-2 have spread over the course of the pandemic, and the generation times for the variants may differ from that of the original viral strain. Hart et al. (2022) found it is likely that the intrinsic mean generation time of the delta variant is shorter than that of the alpha variant, likewise a preliminary study by Abbott et al. (2022) suggests the intrinsic mean generation time of the omicron variant is shorter than that of the delta variant. We find the methodology of Sender et al. somewhat persuasive, and use their point estimate of the generation time (a log-normal distribution with mean 9.7 days) as the default generation time for the original SARS-CoV-2 strain and alpha variant versions of SARS-CoV-2. We compare these default findings to results using the Ferretti et al. point estimate distribution (a Weibull distribution with mean 5.5 days) in online supplementary Section S7.

We then created an alternative version of our model which allowed for changing the generation time distribution due to the delta and omicron variants. We changed the generation time distribution starting in July 2021, reflecting our assumption that delta variant dominated new cases by this point, and changed it again in December 2021, reflecting the same assumption about the omicron variant. Hart estimates the median reduction in the mean generation time for the delta variant is 15% as compared to alpha (we assumed alpha and wild-type had the same generation time) (Hart et al., 2022), while Abbott estimates the median reduction in the mean generation time for omicron is 28% as compared to delta (Abbott et al., 2022). We created generation time distributions for these variants by minimising a squared loss function to search for parameters such that the new distributions had the appropriate new mean generation time, while preserving the standard deviation of the original distribution (see online supplementary Subsection S-1.7 for complete details). We tested whether this new model was needed by calculating the Bayes factor of the two models using data from Alameda County the `bridgesampling` package in R (Gronau et al., 2020; Meng & Wong, 1996), running both models for 26,000 iterations with the first 1,000 iterations discarded as burn-in on 3 chains. The point estimates for the marginal likelihood had error of 7% for the constant generation time model and 6% for the varying generation time model, with a reported Bayes Factor of 1.58 in favour of the model with variant-specific generation times. Even accounting for the margin of error, it is hard to conclude the varying generation time model was decisively superior to the constant generation time model, so we used the constant generation time model in this paper. Because we were testing a characteristic of the infectious disease which should generalise across locations, and because of the computational cost involved, we did not calculate Bayes factors for all 15 counties.

Finally, we used the point estimate of the latent period distribution from Xin et al. (2022) as the delay distribution in our model, with a mean latent period of 5.5 days using a gamma distribution. For the alternative analysis using the Ferretti et al. distribution, we scaled this distribution by 0.5 to halve the mean latent period.

We fit `EpiEstim`, Rt-estim-normal (using the priors from the simulations), and Rt-estim-gamma (see online supplementary Table S1) to this data. The posterior summaries for the effective reproduction number as calculated by `EpiEstim` are displayed in online supplementary Figure S5 and those calculated by Rt-estim-gamma are displayed in Figure 6. Accompanying incidence posterior distributions and case posterior predictive distributions for Rt-estim-gamma are displayed in online supplementary Figures S6 and S7, respectively. Visualisations of the priors and posteriors for non-time-varying parameters for Rt-estim-gamma fit to Los Angeles County data are displayed in
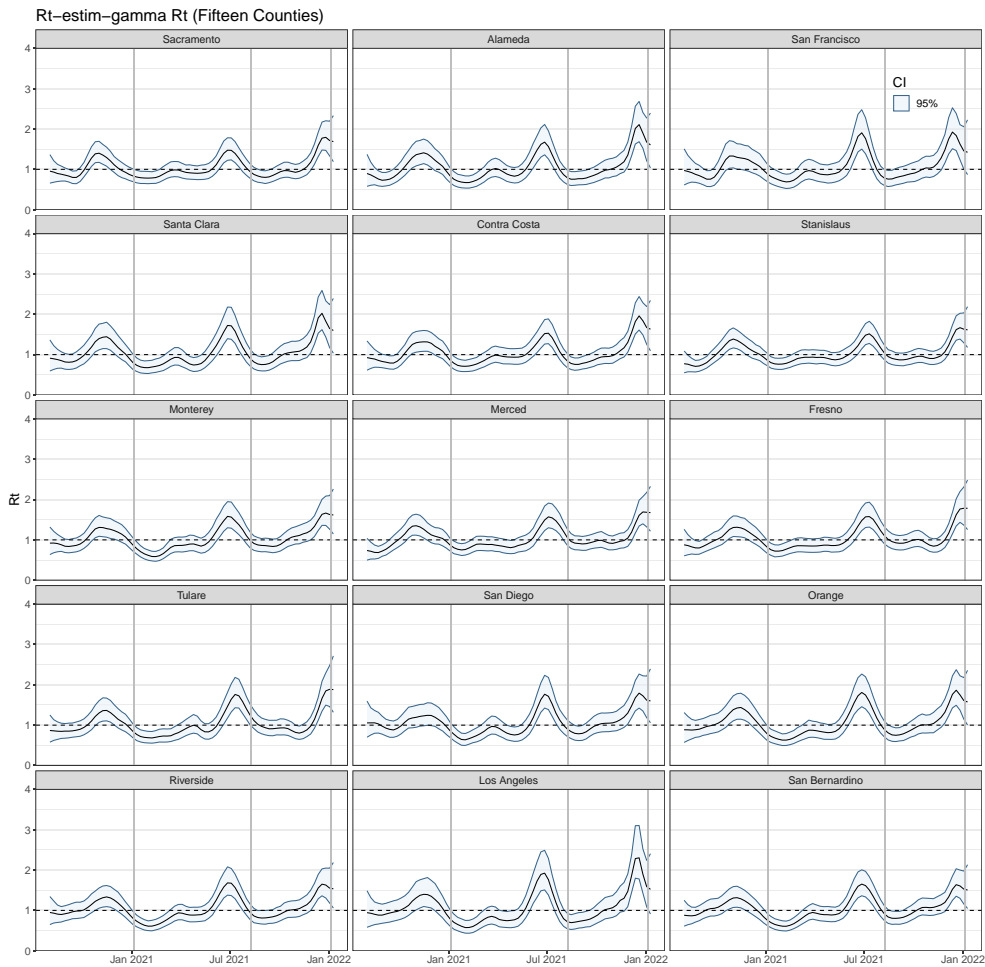
Figure 6. Estimates of the effective reproduction number of SARS-CoV-2 from Rt-estim-gamma applied to 15 counties in California, USA, from 2 August 2020 through 15 January 2022. Blue shaded regions are 95% posterior credible intervals. Black lines are medians. Grey vertical lines mark the date maximum statewide cases were reported for the original winter 2020 wave, the summer 2021 wave, and the winter 2021 wave.

online supplementary Figure S14. After running into convergence issues with Rt-estim-normal, we reduced the data set to 2 August 2020 through 6 November 2021 and fit Rt-estim-normal to this data set. The Rt-estim-normal results were generated using R version 4.2.2.

Priors for Rt-estim-gamma were the same as in the simulations, except that the prior σ had a mean of −0.61 (the range of plausible values was similar), and the priors for ρ and κ were chosen for each county individually using the protocols described in the *Methods* section. We assumed the overall median proportion of observed incidence was 0.066. An example of posterior predictive intervals from the thin plate spline used to choose the prior for κ and from Rt-estim-gamma fit to SARS-CoV-2 case data from Alameda County, CA, are visualised in Figure 3.

Comparisons with Rt-estim-normal are displayed in online supplementary Figures S8–S10. The Rt-estim-gamma results were generated in R version 4.1.2, but all packages were the same as those used to generate simulation results except for Rcpp which was version 1.0.8 rather than version 1.0.7. Overall, Rt-estim-gamma estimates were smoother and more uncertain than estimates from EpiEstim, but less smooth and uncertain than those produced by Rt-estim-normal. This behaviour is consistent with model performance in the simulation scenarios. Rt-estim-gamma estimates tended to estimate less extreme magnitudes than Rt-estim-normal estimates, and while the two models produced broadly similar estimates of the trajectory of the effective reproduction number,

they differed in some counties in significant ways. For example, in San Diego County, the median estimate from Rt-estim-normal is always above 1 before January 2021, while the median estimate from Rt-estim-gamma is below 1 for parts of this period. Additionally, in all counties the median estimate from Rt-estim-gamma crossed below 1 before the median estimate for Rt-estim-normal in fall 2021. Rt-estim-normal and Rt-estim-gamma produced different estimates of the latent incidence (online supplementary Figure S9), but both produced 95% posterior predictive intervals for the observed cases which had good coverage in all counties (online supplementary Figure S10).

The results using Rt-estim-gamma with a mean generation time of 5.5 days are displayed in online supplementary Figures S11–S13. Using a shorter generation time led to generally smaller estimates of the peak effective reproduction number with narrower credible intervals. However, the trajectories using either generation time were similar, and the estimated trajectories agreed on when the median reproduction number was above or below one.

Median estimates for the effective reproduction number were larger during the summer 2021 wave than during the winter 2020 wave. The estimate of the reproduction number during the winter 2021 wave was similar to that of summer 2021 wave except in a few counties where it was larger, such as Los Angeles and Alameda counties. Trajectories were similar across counties, but varied in timing and magnitude from county to county. For instance, the peak reproduction number in the Winter 2020 wave was estimated to occur in the week of 1 November in Sacramento County, and the week of 22 November in Los Angeles County.

4 Discussion

We presented a model for estimating the effective reproduction number using time series of observed cases and diagnostic tests, as well as methods for choosing key priors for the model. We tested the model on simulated data sets, showing it can successfully estimate the true effective reproduction number when data is generated from a stochastic compartmental model. We also tested other models used for estimating the effective reproduction number, demonstrating that when testing supply is relatively constant, a case observation model which ignores testing is reasonable, but when testing supply changes rapidly, ignoring testing leads to poor model performance. Using data from the SARS-CoV-2 epidemic in California, we have shown how using a model fit to case observation data that incorporates testing data leads to different conclusions about the trajectory and magnitude of the effective reproduction number in real world epidemics.

We found that `EpiEstim` had poor performance across all simulation scenarios. In contrast, an assessment of `EpiEstim` by Gostic et al. (2020) found it had reasonable performance on simulated data and recommended it over other existing methodologies (models available in `epidemia` were not assessed in this study). However, Gostic et al. only tested `EpiEstim` on simulated data sets where the true incidence was known. In our study, we tested `EpiEstim` on data sets where cases were noisy realisations of unobserved incidence, a much more realistic scenario for many diseases, such as SARS-CoV-2. The performance of our GLM versions of `EpiEstim` on data from the SARS-CoV-2 epidemic in Orange County, CA, provides one reason for this poor performance. Modelling reported cases as a Poisson random variable assumes a stringent mean–variance relationship which is likely to under-estimate uncertainty. We do not recommend using `EpiEstim` to estimate the effective reproduction number when there is reason to believe reported cases do not reflect true incidence.

When testing was relatively constant, the Rt-estim-normal model, which assumes latent incidence but ignores tests, still performed well. Even in this scenario, the Rt-estim-gamma model we developed for this study had smaller mean credible interval widths and smaller absolute deviations. This suggests that our other modelling choices beyond including tests as a covariate, such as the use of the gamma distribution to model latent incidence, and our process for choosing the prior for the case over-dispersion parameter, had positive effects on model performance. As the SARS-CoV-2 pandemic has unfolded, a number of modelling groups have developed similar techniques for estimating the effective reproduction number. We have demonstrated how modifying distributional assumptions and developing protocols for choosing priors can have significant impact on model performance. We hope these findings motivate the larger community of researchers focused on modelling the effective reproduction number to revisit their work and establish best practices for this class of models.

In simulation scenarios where testing supply increased dramatically, we were still able to successfully estimate the effective reproduction number using the Rt-estim-gamma model. Our findings suggest incorporating testing data is a viable strategy for using case data, which should improve the accuracy of efforts at now-casting the effective reproduction number. It is worth noting that we avoided using a delay distribution in Rt-estim-gamma which incorporated reporting delays, instead using data where cases and tests were tied to the date of the test. This should not prevent the use of Rt-estim-gamma for up-to-date now-casting even though counts of the most recent cases and tests will inevitably be under-counts, so long as the proportion of positive to total tests is independent of reporting delays. We assumed this was the case when applying Rt-estim-gamma to the SARS-CoV-2 data from California. If this assumption proves not to be true, then approaches which do not use testing data and incorporate more elaborate delay distributions, such as those of Abbott et al. (2020) and Bhatt et al. (2023) are probably a better choice. Similarly, because our model relies on the proportion of positive to total tests, rather than the raw counts of positive tests, it should be robust to changes in types of tests available, so long as reported positive cases used the same kinds of tests recorded in total diagnostic tests. This allows us to avoid any problems which arise from the availability of rapid tests for SARS-CoV-2 during the omicron wave.

In a representative set of simulations, even when Rt-estim-normal posteriors captured the effective reproduction number, its posterior estimates for latent incidence did not capture the true latent incidence. To a lesser extent, the same was true of the Rt-estim-gamma model. We have not yet seen any discussion as to the accuracy of incidence estimation for this class of models. Our findings suggest incidence estimates should not be trusted, as there are many values for incidence which lead to the same observed cases and the same reproduction number estimates. Estimates where a population size are taken into account, such as in Mishra et al. (2022), may be more trustworthy, but we recommend running a simulation study first to verify this.

One important limitation of our method is that we condition on the number of tests and use them as a covariate, rather than modelling them jointly with cases. We would expect in practice that the number of tests is also a function of past incidence (with cases rising, more tests will be administered). In mathematical terms, a joint model of cases and tests could be written as

$$P(\mathbf{O}, \mathbf{M} \mid \mathbf{I}, \rho, \kappa) = P(\mathbf{O} \mid \mathbf{M}, \mathbf{I}, \rho, \kappa)P(\mathbf{M} \mid \mathbf{I}, \rho, \kappa).$$

Our method only uses the first term of this product. This leaves our method open to potential bias from model mis-specification. In the simpler context of regression without latent variables, this issue is sometimes called 'feedback', a thorough treatment of the topic is available in Chapter 12 of Diggle et al. (2002). While we think that in practice this will not be a concern in situations where cases and tests increase and decrease together in response to changes in incidence, the possibility does exist. For instance, suppose the testing policy during the peak of an epidemic was that individuals with symptoms could not be tested, as anyone with symptoms should simply assume they have been infected. Tests might still increase in response to increased incidence from the wave, but cases could decline, because no symptomatic people were testing. In such a scenario, we would expect our model to fail. Modelling tests is a non-trivial problem, and implementing a joint model of cases and tests is a promising future direction.

In this paper, we used gamma densities to in order to model changes in latent incidence stochastically. While our choice of a gamma distribution has some desirable benefits, namely that it allows us to use HMC to generate posterior samples, and that it allows for over-dispersion in the variance, there is definitely room for improvement in modelling latent incidence. Recent work by Penn et al. (2022) provides an interesting avenue for improvement, with explicit calculations of the variance of the transition distributions of a time-varying general branching process.

In the real data analysis, we used Rt-estim-gamma with a prior for the over-dispersion parameter derived from a spline fit to the same data as Rt-estim-gamma. This is a workaround we developed to avoid computational problems related to using Hamiltonian Monte Carlo when the prior for the over-dispersion parameter strongly conflicts with the data. Another MCMC method, such as Zig-Zag sampling (Bierkens & Roberts, 2017; Corbella et al., 2022), may not have this issue, and so we could avoid this procedure. While not ideal, we tested our model using this procedure

Table 1. Priors used by the Rt-estim-gamma method in the simulation study

Parameter	Simulation	Prior	Prior median (95% Interval)
ν	All	Log-normal(-2, 0.7)	0.15 (0.03, 0.53)
σ	All	Log-normal(-0.66, 0.6)	0.52 (0.16, 1.68)
λ	All	Exponential(0.3)	2.31 (0.08, 12.26)
$\log R_1$	All	Normal(0, 0.75)	0.01 (-1.49, 1.49)
ρ	Scenario 1	Log-normal(-11.06, 0.3)	1.57E-5 (8.756E-6, 2.85E-5)
ρ	Scenario 2	Log-normal(-11.43, 0.3)	1.09E-5 (5.96E-6, 1.96E-5)
ρ	Scenario 3	Log-normal(-11.56, 0.3)	1.57E-5 (8.81E-6, 2.83E-5)
κ	Scenario 1	Truncated-Normal(59, 60)	72.00 (5.00, 183.15)
κ	Scenario 2	Truncated-Normal(33, 25)	35.65 (3.14, 83.23)
κ	Scenario 3	Truncated-Normal(70, 80)	88.84 (6.00, 235.41)

for choosing the over-dispersion prior on simulated data, and found no discernible loss in performance.

It is clear that more sophisticated representations of the generation time distribution which could change according population dynamics could be incorporated into our model. While this might lead to improved model performance, it is encouraging that in simulations, our model performed well despite using a fixed generation time. It is equally encouraging that our experiments on both simulated and real data showed our model was reasonably robust to different generation time distributions.

One obvious area for improvement in this space is allowing the prior on case detection (ρ) to change over time to better reflect changes in testing policy. For instance, at the start of the SARS-CoV-2 pandemic, only symptomatic individuals could get tested in California, whereas in Fall 2021, anyone was eligible to receive a test. Sherratt et al. (2021) also highlighted how changes in testing eligibility may result in estimating spurious changes in the effective reproduction number. We have found case data alone is insufficient to make a time-varying detection parameter identifiable. Incorporating other sources of data which facilitate real-time estimation, such as data from wastewater treatment facilities, may enable models with time-varying case detection parameters. Enabling effective reproduction number estimation methods to incorporate multiple data streams seems like a fruitful area of future research.

Data availability

All data needed to reproduce the results are available on GitHub at https://github.com/igoldsteinh/improving_rt.

Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series A*.

References

Abbott S., Hellewell J., Thompson R. N., Sherratt K., Gibbs H. P., Bosse N. I., Munday J. D., Meakin S., Doughty E. L., Chun J. Y., Chan Y.-W. D., Finger F., Campbell P., Endo A., Pearson C. A. B., Gimma A., Russell T., CMMID COVID modelling group, Flasche, S., … Funk, S. (2020). Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. *Wellcome Open Research*, 5, 112. https://doi.org/10.12688/wellcomeopenres

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', arXiv, arXiv:2022.01.08.22268920, preprint: not peer reviewed.

Bhatt S., Ferguson N., Flaxman S., Gandy A., Mishra S., & Scott J. A. (2023). Semi-mechanistic Bayesian modeling of COVID-19 with renewal processes. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 186(4), 601–615. https://doi.org/10.1093/jrsssa/qnad030

Bierkens J., & Roberts G. (2017). A piecewise deterministic scaling limit of lifted Metropolis-Hastings in the Curie-Weiss model. *The Annals of Applied Probability*, 27(2), 846–882. https://doi.org/10.1214/16-AAP1217

Bürkner P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80, 1–28. https://doi.org/10.18637/jss.v080.i01

California Open Data Portal. (2023). *California open data portal.* https://data.ca.gov/dataset/covid-19-time-series-metrics-by-county-and-state.

Capistrán M. A., Capella A., & Christen J. A. (2022). Filtering and improved uncertainty quantification in the dynamic estimation of effective reproduction numbers. *Epidemics*, 40, 100624. https://doi.org/10.1016/j.epidem.2022.100624

Champredon D., & Dushoff J. (2015). Intrinsic and realized generation intervals in infectious-disease transmission. *Proceedings of the Royal Society B: Biological Sciences*, 282(1821), 2015–2026. https://doi.org/10.1098/rspb.2015.2026

Champredon D., Dushoff J., & Earn D. J. D. (2018). Equivalence of the Erlang-distributed SEIR epidemic model and the renewal equation. *SIAM Journal on Applied Mathematics*, 78(6), 3258–3278. https://doi.org/10.1137/18M1186411

Corbella A., Spencer S. E., & Roberts G. O. (2022). Automatic Zig-Zag sampling in practice. *Statistics and Computing*, 32(6), 107. https://doi.org/10.1007/s11222-022-10142-x

Cori A., Ferguson N. M., Fraser C., & Cauchemez S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*, 178(9), 1505–1512. https://doi.org/10.1093/aje/kwt133

Diggle P., Diggle P. J., Heagerty P., Liang K.-Y., & Zeger S. (2002). *Analysis of longitudinal data.* Oxford University Press.

Ferretti L., Ledda A., Wymant C., Zhao L., Ledda V., Abeler-Dörner L., Kendall M., Nurtay A., Cheng H.-Y., & Ng T.-C. (2020). The timing of COVID-19 transmission. *medRxiv.* https://doi.org/10.1101/2020.09.04.20188516

Fintzi J., Wakefield J., & Minin V. N. (2022). A linear noise approximation for stochastic epidemic models fit to partially observed incidence counts. *Biometrics*, 78(4), 1530–1541. https://doi.org/10.1111/biom.v78.4

Flaxman S., Mishra S., & Gandy A. e. a. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 584(7820), 257–261. https://doi.org/10.1038/s41586-020-2405-7

Fraser C. (2007). Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS One*, 2(8), 1–12. https://doi.org/10.1371/journal.pone.0000758

Ganyani T., Kremer C., Chen D., Torneri A., Faes C., Wallinga J., & Hens N. (2020). Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Eurosurveillance*, 25(17), 2000257. https://doi.org/10.2807/1560-7917.ES.2020.25.17.2000257

Gostic K. M., McGough L., Baskerville E. B., & Abbott S. e. a. (2020). Practical considerations for measuring the effective reproductive number, Rt. *PLoS Computational Biology*, 16(12), 1–21. https://doi.org/10.1371/journal.pcbi.1009679

Gronau Q. F., Singmann H., & Wagenmakers E.-J. (2020). bridgesampling: An R package for estimating normalizing constants. *Journal of Statistical Software*, 92(10), 1–29. https://doi.org/10.18637/jss.v092.i10

Hart W. S., Miller E., Andrews N. J., Waight P., Maini P. K., Funk S., & Thompson R. N. (2022). Generation time of the alpha and delta SARS-CoV-2 variants: An epidemiological analysis. *The Lancet Infectious Diseases*, 22(5), 603–610. https://doi.org/10.1016/S1473-3099(22)00001-9

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*, e71345. https://doi.org/10.7554/eLife.71345

Meng X.-L., & Wong W. H. (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica*, *6*(4), 831–860. https://www.jstor.org/stable/24306045

Mishra S., Scott J. A., Laydon D. J., Zhu H., Ferguson N. M., Bhatt S., Flaxman S., & Gandy A. (2022). A COVID-19 model for local authorities of the United Kingdom. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185, S86–S95. https://doi.org/10.1111/rssa.12988

Nash R. K., Nouvellet P., & Cori A. (2022). Real-time estimation of the epidemic reproduction number: Scoping review of the applications and challenges. *PLoS Digital Health*, *1*(6), 1–17. https://doi.org/10.1371/journal.pdig.0000052

Pakkanen M. S., Miscouridou X., Penn M. J., Whittaker C., Berah T., Mishra S., Mellan T. A., & Bhatt S. (2023). Unifying incidence and prevalence under a time-varying general branching process. *Journal of Mathematical Biology*, *87*(2), 35. https://doi.org/10.1007/s00285-023-01958-w

Parag K. V. (2021). Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves. *PLoS Computational Biology*, *17*(9), e1009347. https://doi.org/10.1371/journal.pcbi.1009347

Park S. W., Sun K., Champredon D., Li M., Bolker B. M., Earn D. J. D., Weitz J. S., Grenfell B. T., & Dushoff J. (2021). Forward-looking serial intervals correctly link epidemic growth to reproduction numbers. *Proceedings of the National Academy of Sciences*, *118*(2), e2011548118. https://doi.org/10.1073/pnas.2011548118

Penn M. J., Laydon D. J., Penn J., Whittaker C., Morgenstern C., Ratmann O., Mishra S., Pakkanen M. S., Donnelly C. A., & Bhatt S. (2022). 'The uncertainty of infectious disease outbreaks is underestimated', arXiv, arXiv:2210.14221, preprint: not peer reviewed.

R Core Team. (2020). *R: A language and environment for statistical computing*. https://www.R-project.org/.

Scott J. A., Gandy A., Mishra S., Bhatt S., Flaxman S., Unwin H. J. T., & Ish-Horowicz J. (2021). 'Epidemia: An R package for semi-mechanistic bayesian modelling of infectious diseases using point processes', arXiv.

Sender R., Bar-On Y. M., Park S. W., Noor E., Dushoff J., & Milo R. (2021). 'The unmitigated profile of COVID-19 infectiousness', medRxiv. https://www.medrxiv.org/content/early/2021/11/25/2021.11.17.21266051

Sherratt K., Abbott S., Meakin S. R., Hellewell J., Munday J. D., Bosse N., CMMID COVID-19 Working Group Jit, M., & Funk, S. (2021). Exploring surveillance data biases when estimating the reproduction number: With insights into subpopulation transmission of COVID-19 in England. *Philosophical Transactions of the Royal Society B*, *376*(1829), 20200283. https://doi.org/10.1098/rstb.2020.0283

Stan Development Team. (2020). RStan: The R interface to Stan (R package version 2.21.2). http://mc-stan.org/.

Svensson A. (2007). A note on generation times in epidemic models. *Mathematical Biosciences*, *208*(1), 300–311. https://doi.org/10.1016/j.mbs.2006.10.010

Swiss National Covid-19 Science Task Force. (2020). *Situation report: Reproductive number*. https://ncs-tf.ch/en/situation-report

Teh Y. W., Elesedy B., He B., Hutchinson M., Zaidi S., Bhoopchand A., Paquet U., Tomasev N., Read J., & Diggle P. J. (2022). Efficient Bayesian inference of instantaneous reproduction numbers at fine spatial scales, with an application to mapping and nowcasting the COVID-19 epidemic in British local authorities. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *185*(Suppl. 1), S65–S85. https://doi.org/10.1111/rssa.12971

Thompson R. N., Stockwin J. E., van Gaalen R. D., Polonsky J. A., Kamvar Z. N., Demarsh P. A., Dahlqwist E., Li S., Miguel E., Jombart T., & Lessler J. (2019). Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*, *29*, 100356. https://doi.org/10.1016/j.epidem.2019.100356

Wallinga J., & Teunis P. (2004). Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology*, *160*(6), 509–516. https://doi.org/10.1093/aje/kwh255

Wood S. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Chapman and Hall/CRC.

Xin H., Li Y., Wu P., Li Z., Lau E. H., Qin Y., Wang L., Cowling B. J., Tsang T. K., & Li Z. (2022). Estimating the latent period of coronavirus disease 2019 (COVID-19). *Clinical Infectious Diseases*, *74*(9), 1678–1681. https://doi.org/10.1093/cid/ciab746