Infectious Disease Modelling 9 (2024) 680–688

Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm

Deterministic epidemic models overestimate the basic reproduction number of observed outbreaks

Wajid Ali [a], Christopher E. Overton [a], Robert R. Wilkinson [b], Kieran J. Sharkey [a, *]

[a] Department of Mathematical Sciences, University of Liverpool, Peach Street, Liverpool, L69 7ZX, England, United Kingdom
[b] Department of Applied Mathematics, Liverpool John Moores University, Byrom Street, Liverpool, L3 5UX, England, United Kingdom

ARTICLE INFO

Article history:
Received 13 April 2023
Received in revised form 9 February 2024
Accepted 13 February 2024
Handling Editor: Dr. Raluca Eftimie

Keywords:
Estimating R_0
Simple birth-death process
Major outbreak
Conditioned epidemic
Stochastic fade-out

ABSTRACT

The basic reproduction number, R_0, is a well-known quantifier of epidemic spread. However, a class of existing methods for estimating R_0 from incidence data early in the epidemic can lead to an over-estimation of this quantity. In particular, when fitting deterministic models to estimate the rate of spread, we do not account for the stochastic nature of epidemics and that, given the same system, some outbreaks may lead to epidemics and some may not. Typically, an observed epidemic that we wish to control is a major outbreak. This amounts to implicit selection for major outbreaks which leads to the over-estimation problem. We formally characterised the split between major and minor outbreaks by using Otsu's method which provides us with a working definition. We show that by conditioning a 'deterministic' model on major outbreaks, we can more reliably estimate the basic reproduction number from an observed epidemic trajectory.

1. Introduction

A new, emerging infectious disease can potentially spread around the world within days or weeks, as observed during COVID-19 (Carvalho et al., 2021), and swine flu (Coker, 2009). During the early phase of an epidemic, estimation of key epidemiological parameters helps us to estimate its future behaviour including the rate of spread, final size, and the requirements for effective control. In these early stages, epidemics typically exhibit exponential growth.

The basic reproduction number, R_0, is the average number of secondary infections per primary infection in an otherwise susceptible population (Dietz, 1993; Heesterbeek & Dietz, 1996; Heffernan et al., 2005). The basic reproduction number has been shown to have important implications relating to the final epidemic size (Andreasen, 2011) and requirements for control (Lipsitch et al., 2003). It is also directly related to the early growth rate (r) of epidemics (Lipsitch et al., 2003; Ma, 2020) and both of these quantifiers are used for predicting the fate of outbreaks. That is, when R_0 is greater than 1 (or r is positive), then the introduction of an infected individual into a susceptible population may lead to a major epidemic. Given the random nature of infection processes, stochastic models are the natural choice to model epidemics (Bailey, 1975; Britton & Pardoux,

* Corresponding author.
E-mail address: kjs@liverpool.ac.uk (K.J. Sharkey).
Peer review under responsibility of KeAi Communications Co., Ltd.

https://doi.org/10.1016/j.idm.2024.02.007

2019; Whittle, 1955), and methods such as maximum likelihood can be used to estimate R_0, taking into account this inherent randomness (Becker & Britton, 1999; Britton & Pardoux, 2019; Ma et al., 2014).

Although epidemics are stochastic processes, it is sometimes convenient to use a deterministic approach such as the Kermack-Mckendrick SIR model (Kermack & McKendrick, 1927, 1932), SIS model (Lajmanovich & Yorke, 1976), SEIR Model (Anderson & May 1992) or the exponential or logistic growth curves (Chowell et al., 2006) to understand and predict them. Unlike their stochastic counterparts (Bailey, 1975; Britton & Pardoux, 2019; Whittle, 1955), these models guarantee an epidemic when $R_0 > 1$ (Dietz, 1993; Kermack & McKendrick, 1927). Such models have been used to estimate epidemic parameters by fitting them to real epidemic data, for example, influenza (Chowell et al., 2016), cholera (Pourabbas et al., 2001), and COVID-19 (Metelmann et al., 2021). We refer to (Ma, 2020; Ma et al., 2014) for more details on estimating early growth rates and the basic reproduction number from real data.

These classic models can be valuable but may lead to an overestimation of the basic reproduction number (Breban et al., 2007; Chowell, 2017; Green et al., 2006; Keeling & Grenfell, 2000). Generally, uncertainty in the estimation of parameters may arise due to noise in the data and/or the underlying assumptions for building models (Chowell, 2017; Ferrari et al., 2005). However, here we observe that there is also a fundamental bias in the deterministic models which occurs because they do not capture the stochastic effects in the early phases of an outbreak and, in particular, do not distinguish the possibility of stochastic fade-out when R_0 is greater than 1 (Bailey, 1975; Whittle, 1955). Similar issues with deterministic models and stochastic fade-out have been explored in (Overton et al., 2022) in the context of steady-state solutions to the SIS model.

By reducing them to a simple birth-death process, we show that SIR and SIS deterministic models implicitly average over both major and minor outbreaks during their early phases; that is both extinct and extant trajectories are included in the average behaviour. However, an observed epidemic is necessarily a major outbreak and therefore corresponds to an implicit conditioning on major outbreaks. This leads deterministic SIR and SIS models to overestimate the basic reproduction number, R_0, when they are fitted to epidemic data which we illustrate in the next section. This is more pronounced when the probability of minor outbreaks is large; i.e. when we have a small number of initial infections or when R_0 is close 1. In Section 3 we consider a birth-death process conditioned on major outbreaks which we approximate by conditioning on non-extinction (Kot, 2001; Kendall, 1948a, 1948b). This better-describes a typical major outbreak and we show that it performs well in removing the bias from the estimation of R_0.

2. Estimation of R_0 using standard deterministic models

Consider an infectious disease that is spread via contact between susceptible and infected individuals in a well-mixed homogeneous population. Let τ be the rate at which a single individual infects a susceptible individual during its infectious period and let $i(t)$ and $s(t)$ denote the infectious and susceptible populations respectively. We consider an infectious disease with no latent period and suppose that infection occurs according to a Poisson process with rate τsi. Similarly, we assume removal (or recovery) occurs according to a Poisson process with rate γi where γ is the rate of removal/recovery of a single individual. This can be applied to infections that produce no long-term immunity (SIS or SIRS) or permanent immunity (SIR).

In a sufficiently large population, the early phase of the epidemic behaves like a simple birth-death (BD) process. This can be seen from the infection rates τsi; for a total population of size N initiated with a single infected ($i_0 = 1$) in an otherwise susceptible population, the initial infection rates for $i = 1, i = 2, i = 3, \ldots$ infected individuals are $\tau(N-1), 2\tau(N-2), 3\tau(N-3)$,

Table 1
State transitions in the Susceptible-Infected-Susceptible (SIS), Susceptible-Infected-Recovered (SIR) and the simple Birth-Death (BD) processes.

Event	SIS	SIR	BD
Infection	$(s,i) \xrightarrow{\tau si} (s-1, i+1)$	$(s,i) \xrightarrow{\tau si} (s-1, i+1)$	$i \xrightarrow{\beta i} i+1$
Recovery	$(s,i) \xrightarrow{\gamma i} (s+1, i-1)$	$(s,i) \xrightarrow{\gamma i} (s, i-1)$	$i \xrightarrow{\gamma i} i-1$

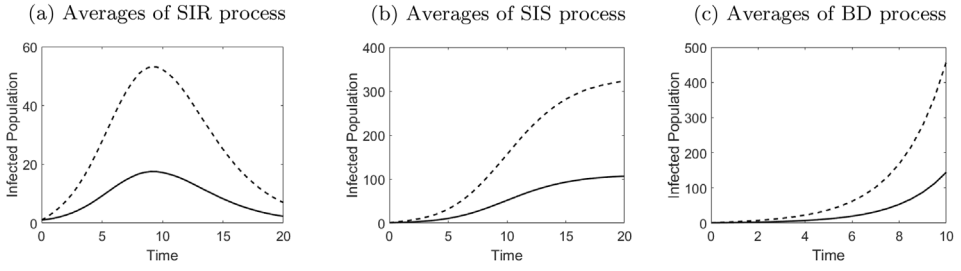

Fig. 1. The average of 10,000 simulations (solid line) and those conditioned on major outbreaks (dashed line) for (a) the SIR process, (b) the SIS process and (c) the simple Birth-Death (BD) process. In each case, $\beta = 1.5$, $\gamma = 1$, $N = 1000$ and the initial number infected is $i_0 = 1$.

… respectively. These are approximately $i\tau N = \beta i$ because the susceptible population is approximately N (Renshaw, 1993). So, the early phases of the infection dynamics are approximated by a simple birth-death process with individual birth rate β and individual death rate γ. For comparison, both SIS and SIR processes and (their approximation) the simple BD process are summarised in Table 1, and the time series curves of the infected populations are illustrated in Fig. 1.

The expected number of infected individuals (denoted by $\langle i \rangle$) in the simple BD process at a given time t can be derived from the master equation for the process. Let the probability that there are i infected individuals at time t be denoted by $p_i(t)$, where $i \in \{0, 1, \dots\}$. Thus, the master equation for the simple birth-death process is given by:

$$\frac{d}{dt}p_i(t) = \beta(i-1)p_{i-1}(t) - (\beta+\gamma)ip_i(t) + \gamma(i+1)p_{i+1}(t). \tag{1}$$

From this, the rate of change of the expected infectious population is (Feller, 1939; Kendall, 1948a):

$$\frac{d}{dt}\langle i \rangle = \sum_i i\frac{d}{dt}p_i(t) = (\beta-\gamma)\langle i \rangle. \tag{2}$$

This has the same form as the deterministic simple BD model given in Table 2, so the deterministic BD model describes the average of all stochastic realisations of the stochastic BD process. This connection is well-known but unusual, although similar connections can be established between the stochastic and deterministic SIS and SIR models under some closure approximations (Kiss et al., 2017; Sharkey et al., 2015) and in limiting cases Kurtz (1970).

Moreover, the deterministic BD model (and equivalently Equation (2)) describes the expected early deterministic dynamics of both SIS and SIR processes when N is large. This is because (see Table 2) the deterministic SIS and SIR (and SIRS) models have the following equation for the infectious population (Kermack & McKendrick, 1927):

$$\frac{dI}{dt} = \tau SI - \gamma I, \tag{3}$$

where we use capital letters I and S for denoting the number of infected and susceptible individuals in deterministic models. This reduces to the form of Equation (2) under the same approximation as we applied to the stochastic models (i.e. $S \approx N$ with $\beta = \tau N$) and so the expected behaviour of the stochastic SIR and SIS models is approximated by the deterministic SIR and SIS models, and by the BD model in the early stages.

The equivalence of the deterministic models to the expected value of the stochastic models and their derivation from the master equation tells us that the deterministic epidemic models approximate an averaging over all epidemic outcomes (Kurtz, 1970; Overton et al., 2022). Crucially this averaging is over both major and minor outbreaks. However, a real epidemic of interest is a major outbreak and this therefore represents conditioning on major outbreaks.

Fig. 2a shows the distribution of final sizes (Andreasen, 2011) of an SIR process when initiated with a single infected individual. The bimodal nature of this distribution tells us that that a group of realisations generate major outbreaks while others go extinct in the early phase. Fig. 2b shows similar bimodal behaviour for SIS dynamics where here the process is run until either extinction or until $2N$ events have occurred. Although the split between major and minor outbreaks is usually obvious, it is not well-defined in finite populations. Throughout this paper we choose to formally characterise the split between major and minor outbreaks by using Otsu's method (Otsu, 1979) which gives a threshold value for clustering bimodal histograms and provides us with a working definition.

To illustrate the issue of over-estimation, we performed least-squares fits of both deterministic SIR and SIS models to simulated major outbreaks of both types to estimate the transmission rate β, assuming that we know γ (here $\gamma = 1$). R_0 is then calculated from $R_0 = \beta/\gamma$ (Keeling & Grenfell, 2000; van den Driessche, 2017).

For all least-squares fits in the paper, epidemic incidence data was constructed by taking the number of infection events between one time step and the next: $j_t = c(t+\delta t) - c(t)$ where $c(t)$ is the cumulative number of infection events up to time t given by $c(t) = N - S(t)$ in the SIR case. We minimise

Table 2
Equations for deterministic SIS, SIR and BD models where S, I, R are the sizes of the classes of susceptible, infected and removed respectively. The parameters τ and γ are the transmission and recovery (or removal) rates. Here, for convenience we have used the same parameter notation as for the stochastic models.

SIS	SIR	BD
$\frac{dS}{dt} = -\tau SI + \gamma I$ $\frac{dS}{dt} = \tau SI - \gamma I$	$\frac{dS}{dt} = -\tau SI$ $\frac{dS}{dt} = \tau SI - \gamma I$ $\frac{dR}{dt} = \gamma I$	$\frac{dI}{dt} = \beta I - \gamma I$

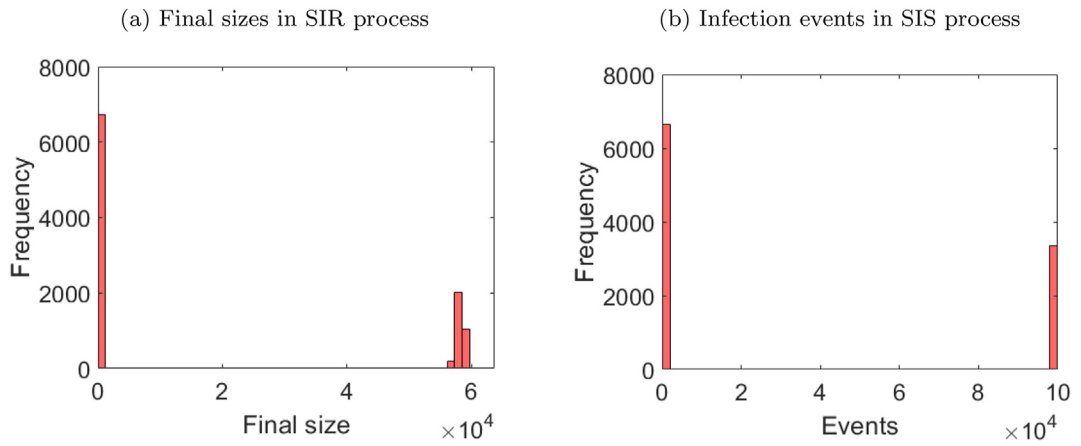

Fig. 2. (a) The final size distribution obtained from 10, 000 stochastic simulations of SIR dynamics and (b) the distribution in the number of (infection and recovery) events for 10,000 SIS simulations capped at $2N$ events. In both sets of simulations, $N = 10,000$, $\beta = 1.5$, $\gamma = 1$, and the initial number infected is $i_0 = 1$.

$$SSE = \sum_{n=0}^{T/\delta t} (J_{n\delta t} - j_{n\delta t})^2$$

where $J_t = C(t+1) - C(t)$ and $C(t)$ is the cumulative number of infections in the corresponding deterministic model (Ma et al., 2014). Here, T is the upper end of the fit window and we use $\delta t = 0.1$. Minimisation was performed using the Nelder-Mead simplex algorithm as implemented by the fminsearch function in MATLAB ver. R2023a.

The resulting distributions in R_0 values are shown for both SIR (Fig. 3a) and SIS (Fig. 3b). Both distributions clearly show an upward bias in the estimates with respect to the "true" value of $R_0 = 1.5$ used to generate the stochastic simulations. Thus, fitting mean-field SIS and SIR models (or the simple BD model) to major real outbreak data will have a tendency to overestimate the transmission rate and overestimate R_0.

To resolve this and obtain a deterministic model that we can fit to the early phases of epidemics to determine R_0, we need to account for the implicit conditioning on major outbreaks. Since the early stages of SIS and SIR dynamics are well-approximated by the simple BD process, we try to obtain a simple BD process which is conditioned on major outbreaks.

3. Estimation of R_0 using a conditioned BD model

We wish to calculate the conditional probability $P(i$ infected at time $t|$ major outbreak) for the simple BD process. Due to the ambiguity in defining a major outbreak and to make analytic progress, we approximate this probability by:

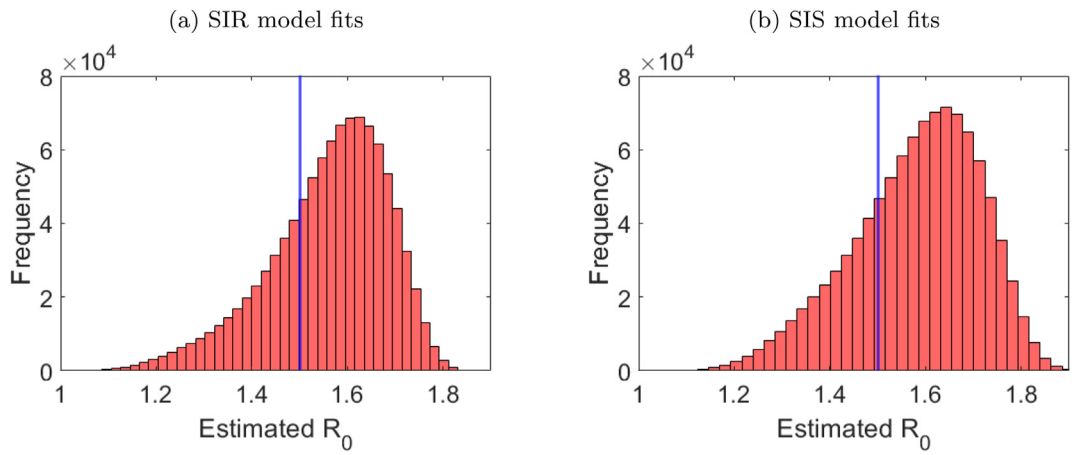

Fig. 3. The distributions of R_0 estimated by performing least squares fits of (a) the SIR model and (b) the SIS model to 1 million major outbreaks of each type generated by taking $\beta = 1.5$, $\gamma = 1$, $N = 100,000$ and the initial number infected $i_0 = 1$. The blue lines in both plots represent the actual value of R_0 used to generate the simulated outbreaks. The fitting window for both histograms is $t \in [0, 10]$.

$$P(i \text{ infected at time } t | \text{ major outbreak}) \approx q_i(t)$$

where

$$q_i(t) = P(i \text{ infected at time } t | i \neq 0 \text{ at time } t)$$

for $i \in \{1, 2, \dots\}$. This is given by (Kot, 2001)

$$q_i(t) = \frac{p_i(t)}{1 - p_0(t)}$$

where, with reference to Equation (1), the probability that the disease has died out at time t, is denoted by $p_0(t)$.

Differentiating gives

$$\frac{d}{dt}q_i(t) = \frac{\frac{d}{dt}p_i(t)}{1 - p_0(t)} + \frac{p_i(t)}{[1 - p_0(t)]^2}\frac{d}{dt}p_0(t), \quad \text{for} \quad i = 1, 2, \dots.$$

Using the expression for $dp_i(t)/dt$ given in Equation (1) when $i \neq 0$, and using $dp_0(t)/dt = \gamma p_1(t)$ and $q_i(t) = p_i(t)/(1 - p_0(t))$, we reach the following model (Kot, 2001):

$$\frac{d}{dt}q_i(t) = \beta(i-1)q_{i-1}(t) - (\beta+\gamma)iq_i(t) + \gamma(i+1)q_{i+1} + \gamma q_1 q_i(t). \tag{4}$$

We now derive an equation for the rate of change of the expected number of infected individuals at time t:

$$\langle i \rangle_c(t) = \sum_{i=1} i q_i(t),$$

where the subscript c indicates that this is the expected value in the conditioned process. Differentiating with respect to time and substituting for $dq_i(t)/dt$ from Equation (4) gives:

$$\frac{d}{dt}\langle i \rangle_c = \beta\sum_{i=1} i(i-1)q_{i-1}(t) - (\beta+\gamma)\sum_{i=1} i^2 q_i(t) + \gamma\sum_{i=1} i(i+1)q_{i+1} + \gamma q_1 \langle i \rangle_c.$$

The summation in the first term on the right-hand-side can be written as:

$$\sum_{i=1} i(i-1)q_{i-1}(t) = \sum_{k=0}(k+1)(k)q_k(t) = \sum_{i=1} i^2 q_i(t) + \sum_{i=1} i q_i(t),$$

and the summation in the third term on the right-hand-side can be written as:

$$\sum_{i=1} i(i+1)q_{i+1}(t) = \sum_{k=2}(k-1)(k)q_k(t) = \sum_{i=2} i^2 q_i(t) - \sum_{i=2} i q_i(t).$$

This leads to:

$$\frac{d}{dt}\langle i \rangle_c = (\beta - \gamma)\langle i \rangle_c + \gamma \langle i \rangle_c q_1 \tag{5}$$

with the cumulative number of infections given by

$$\frac{dC}{dt} = \beta\langle i \rangle_c + \gamma \langle i \rangle_c q_1.$$

This system is not closed because we have one extra variable, q_1, which is the probability of single infection in the conditioned process. However, using the exact solution of Equation (1) (Renshaw, 1993; Kot, 2001, Chapter 3), the exact expression for q_1 is given by

$$q_1(t) = \frac{p_1(t)}{1 - p_0(t)}$$

where

$$p_0(t) = \begin{cases} \left[\dfrac{\beta t}{1+\beta t}\right]^{i_0} & \text{if} \quad \beta = \gamma \\ \left[\dfrac{\gamma(e^{rt}-1)}{\beta e^{et}-\gamma}\right]^{i_0} & \text{if} \quad \beta \neq \gamma \end{cases}$$

and

$$p_1(t) = \begin{cases} i_0\left[\dfrac{\beta t}{1+\beta t}\right]^{i_0-1}\dfrac{1}{(1+\beta t)^2} & \text{if} \quad \beta = \gamma \\ i_0\alpha^{i_0-1}[(1-\alpha)(1-\varphi)] & \text{if} \quad \beta \neq \gamma. \end{cases}$$

Here i_0 is the initial number of infected individuals and

$$\alpha = \frac{\gamma(e^{rt}-1)}{\beta e^{rt}-\gamma}, \quad \varphi = \frac{\beta(e^{rt}-1)}{\beta e^{rt}-\gamma},$$

where $r = \beta - \gamma$.

As discussed in Section 2, the early phases of the SIS and SIR epidemic models are well-approximated by the simple BD process. Similarly the conditional (SIS and SIR) epidemic processes are also well-approximated by the conditional stochastic BD process. Conditioned and non-conditioned processes lead to different average trajectories (Fig. 4a) and it can be seen that the new conditioned BD model (Equation (5)) accurately describes the early part of the SIR, SIS and BD expected behaviour when these processes are conditioned on major outbreaks. This model resolves the issue of upward bias which we saw in Fig. 3a. When it is fitted to major outbreaks of SIR epidemics, we can see in Fig. 4b that the distribution of estimated R_0 values is centred around the true value. Fig. 5a explores the parameter space for small R_0 more fully and shows a significantly improved estimate of R_0 when compared with the standard SIR model (Fig. 5b). The fitting window for Figs. 4b and 5 is the time interval $t \in [0\ T]$ where we determined T to be the time point at which the simple BD approximation gives a 1% error with respect to the infectious time series in the SIR process.

4. Discussion

One method for obtaining the basic reproduction number (R_0) from real data is to fit deterministic models to epidemic data (Breban et al., 2007; Chowell, 2017; van den Driessche, 2017; Green et al., 2006; Keeling & Grenfell, 2000; Ma, 2020). Here we showed that some of these deterministic models are fundamentally biased in their estimation of R_0 (Breban et al., 2007; Chowell, 2017). We showed this explicitly in the context of SIS and SIR epidemic dynamics (Fig. 3a and b) where we can make some analytic progress. This also applies directly to SIRS models as well (Mena-Lorcat & Hethcote, 1992).

The early phases of both the stochastic SIS and SIR epidemic processes and the deterministic SIS and SIR models reduce to a simple birth-death process (see, for example (Renshaw, 1993)), and so the early behaviour of the deterministic models

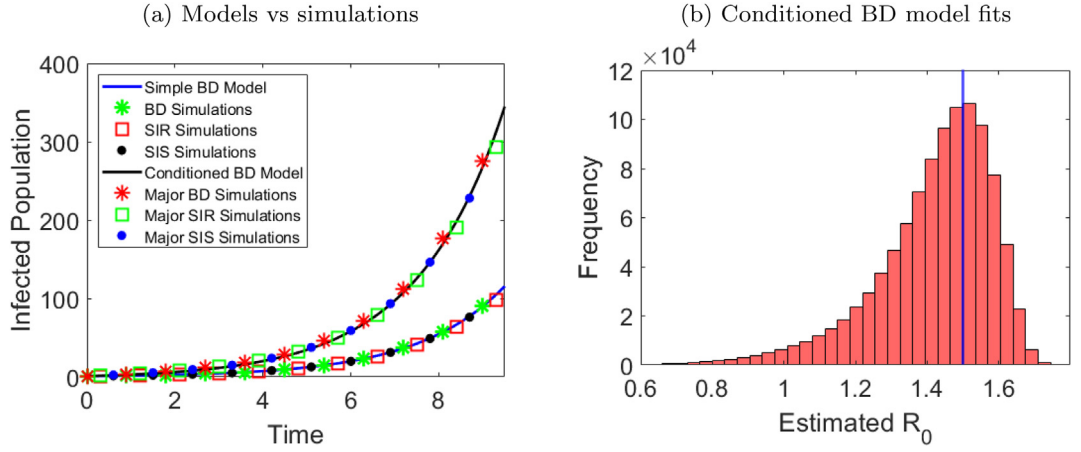

Fig. 4. (a) The deterministic conditioned and unconditioned BD models together with averages of 100,000 conditioned and unconditioned BD, SIS and SIR simulations. (b) Estimates of R_0 by least-square fits of the conditioned BD Model (Equation (5)) to 1 million major outbreaks. The fitting window used is $t \in [0, 9.5]$ which is determined so that the simple BD model is within 1% of the SIR infectious time series. For both subplots, $i_0 = 1$, $\beta = 1.5$, $\gamma = 1$ and $N = 100,000$.

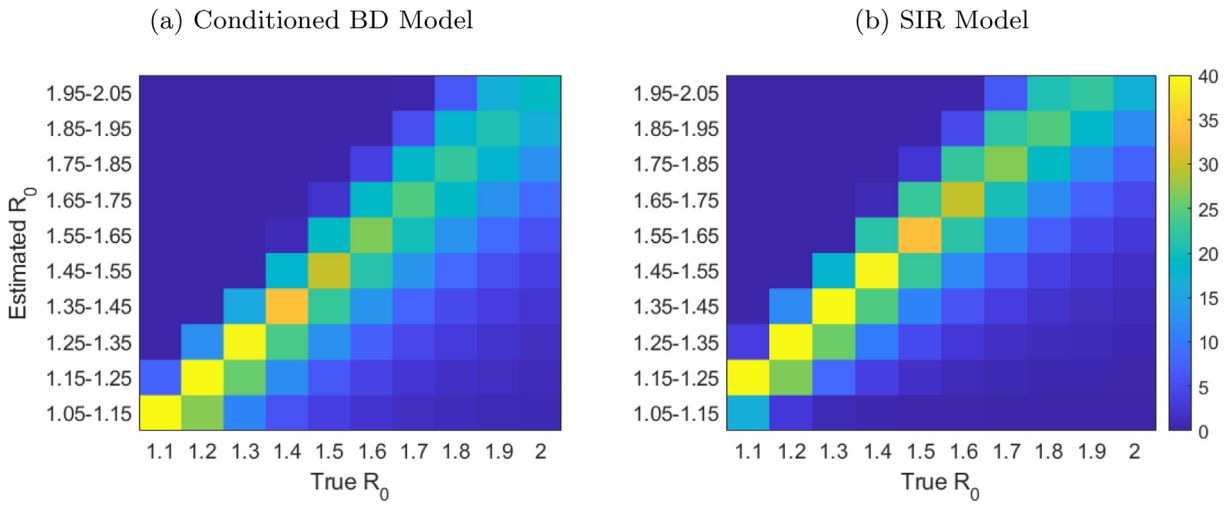

Fig. 5. The distribution of estimated R_0 values by fitting (a) the conditioned BD model and (b) the SIR model. Both these models are fitted to 1 million simulated major SIR outbreaks per R_0 value. The horizontal axis corresponds to the "true" values of R_0 used to generate the simulations. The vertical axis corresponds to the histograms of estimated R_0 (similar to Fig. 4b). The intensity of the color indicates the percentage of estimates in each interval.

describes the expected behaviour of all possible epidemic realisations. The over-estimation of R_0 occurs because a real epidemic of interest is necessarily a major outbreak, representing an implicit conditioning on major outbreaks. This over-estimation arises when deterministic epidemic models are initialised with very few initial infected individuals (i_0) and/or where R_0 is very close to 1 which leads to the probability of a minor outbreak (given by $(1/R_0)^{i_0}$(Whittle, 1955)) to be significant. Once the epidemic is underway and the deterministic models can be initialised with sufficient infected individuals to make the probability of minor outbreaks negligible, then the overestimation problem does not arise.

We resolved the issue of overestimation by developing a simple birth-death model with conditioning on major outbreaks. To make analytic progress we approximated conditioning on major outbreaks by conditioning against extinction (Equation (5)) (Kot, 2001; Kendall, 1948a, 1948b). We made use of the analytic solution of the simple birth-death master equation and showed that fitting this model to the early stages of epidemic outbreaks resolves the issue of overestimation of R_0 (Figs. 4b and 5a). Similar models were used for approximating the average phylogenic lineages (Harvey et al., 1994) and for correcting a similar bias in deterministic coalescent models (Stadler et al., 2015).

It is expected that similar issues would apply to other epidemic models such as SEIR type models (Anderson & May 1992) and models that do not rely on Poisson processes (Kenah, 2011; KhudaBukhsh et al., 2020). It would be of interest to determine the extent of this error in these scenarios. For this, we currently lack the ingredients that we made use of which are the approximation of SIS and SIR dynamics by the simple birth-death (BD) process, the analytic solution of the BD process and the equivalence of the deterministic BD and the average stochastic BD processes. Nevertheless, this implicit conditioning on major outbreaks seems likely to cause problems with these models as well. The approximation to the conditioned SIS model proposed in (Overton et al., 2022) can also be extended to cover transient dynamics and to model more complex epidemiological structures. While the validity of using a deterministic description early in an epidemic is questionable due to the stochastic fluctuations of the infected population, the main fluctuation is the one between major and minor outbreaks. Our work emphasises the importance of not defining the initial conditions of standard deterministic models too early in the epidemic unless the models have conditioning against minor outbreaks.

CRediT authorship contribution statement

Wajid Ali: Writing – review & editing, Writing – original draft, Methodology, Investigation. **Christopher E. Overton:** Writing – review & editing, Methodology, Conceptualization. **Robert R. Wilkinson:** Writing – review & editing, Methodology, Conceptualization. **Kieran J. Sharkey:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 955708.

References

Anderson, R., & May, R. (1992). *Infectious diseases of humans: Dynamics and control*.
Andreasen, V. (2011). The final size of an epidemic and its relation to the basic reproduction number. *Bulletin of Mathematical Biology, 73*, 2305–2321. https://doi.org/10.1007/s11538-010-9623-3
Bailey, N. (1975). *The elements of stochastic processes with applications to the natural sciences*.
Becker, N. G., & Britton, T. (1999). Statistical studies of infectious disease incidence. *Journal of the Royal Statistical Society: Series B, 61*, 287–307. https://doi.org/10.1111/1467-9868.00177
Breban, R., Vardavas, R., & Blower, S. (2007). Theory versus data: How to calculate R_0? *PLoS One, 2*. https://doi.org/10.1371/JOURNAL.PONE.0000282
Britton, T., & Pardoux, E. (2019). *Lecture notes in mathematics 2255 stochastic epidemic models with inference*. https://doi.org/10.1111/1467-9868.00177
Carvalho, T., Krammer, F., & Iwasaki, A. (2021). The first 12 months of covid-19: A timeline of immunological insights. *Nature Reviews Immunology, 21*, 245–256. https://doi.org/10.1038/s41577-021-00522-1
Chowell, G. (2017). Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling, 2*, 379–398. https://doi.org/10.1016/J.IDM.2017.08.001
Chowell, G., Ammon, C., Hengartner, N., & Hyman, J. (2006). Transmission dynamics of the great influenza pandemic of 1918 in geneva, Switzerland: Assessing the effects of hypothetical interventions. *Journal of Theoretical Biology, 241*, 193–204. https://doi.org/10.1016/j.jtbi.2005.11.026
Chowell, G., Sattenspiel, L., Bansal, S., & Viboud, C. (2016). Mathematical models to characterize early epidemic growth: A review. *Physics of Life Reviews, 18*, 66–97. https://doi.org/10.1016/j.plrev.2016.07.005
Coker, R. (2009). Swine flu. *BMJ, 338*. https://doi.org/10.1136/bmj.b1791. b1791–b1791.
Dietz, K. (1993). The estimation of the basic reproduction number for infectious diseases. *Statistical Methods in Medical Research, 2*, 23–41. https://doi.org/10.1177/096228029300200103
Feller, W. (1939). Die grundlagen der volterraschen theorie des kampfes ums dasein in wahrscheinlichkeitstheoretischer behandlung. *Acta Biotheoretica, 5*, 11–40. https://doi.org/10.1007/BF01602932/METRICS
Ferrari, M. J., Bjørnstad, O. N., & Dobson, A. P. (2005). Estimation and inference of R_0 of an infectious pathogen by a removal method. *Mathematical Biosciences, 198*, 14–26. https://doi.org/10.1016/J.MBS.2005.08.002
Green, D. M., Kiss, I. Z., & Kao, R. R. (2006). Parameterization of individual-based models: Comparisons with deterministic mean-field models. *Journal of Theoretical Biology, 239*, 289–297. https://doi.org/10.1016/J.JTBI.2005.07.018
Harvey, P. H., May, R. M., & Nee, S. (1994). Phylogenies without fossils. *Evolution, 48*, 523. https://doi.org/10.2307/2410466
Heesterbeek, J. A. P., & Dietz, K. (1996). The concept of R_0 in epidemic theory. *Statistica Neerlandica, 50*, 89–110. https://doi.org/10.1111/j.1467-9574.1996.tb01482.x
Heffernan, J., Smith, R., & Wahl, L. (2005). Perspectives on the basic reproductive ratio. *Journal of The Royal Society Interface, 2*, 281–293. https://doi.org/10.1098/rsif.2005.0042
Keeling, M. J., & Grenfell, B. T. (2000). Individual-based perspectives on R_0. *Journal of Theoretical Biology, 203*, 51–61. https://doi.org/10.1006/JTBI.1999.1064
Kenah, E. (2011). Contact intervals, survival analysis of epidemic data, and estimation of r0. *Biostatistics, 12*, 548–566. https://doi.org/10.1093/BIOSTATISTICS/KXQ068
Kendall, D. G. (1948a). On the generalized "birth-and-death" process. *The Annals of Mathematical Statistics, 19*, 1–15. https://doi.org/10.1214/AOMS/1177730285
Kendall, D. G. (1948b). On some modes of population growth leading to r. a. Fisher's logarithmic series distribution. *Biometrika, 35*, 6–15. https://doi.org/10.1093/biomet/35.1-2.6
Kermack, W. O., & McKendrick, W. O. A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London - Series A: Containing Papers of a Mathematical and Physical Character, 115*, 700–721. https://doi.org/10.1098/rspa.1927.0118
Kermack, W. O., & McKendrick, W. O. A. G. (1932). Contributions to the mathematical theory of epidemics. ii. —the problem of endemicity. *Proceedings of the Royal Society of London - Series A: Containing Papers of a Mathematical and Physical Character, 138*, 55–83. https://doi.org/10.1098/rspa.1932.0171
KhudaBukhsh, W. R., Choi, B., Kenah, E., & Rempała, G. A. (2020). Survival dynamical systems: Individual-level survival analysis from population-level epidemic models. *Interface Focus, 10*. https://doi.org/10.1098/RSFS.2019.0048
Kiss, I. Z., Miller, J. C., & Simon, P. L. (2017). *Mathematics of epidemics on networks: From exact to approximate models, ume 46*. Springer International Publishing. https://doi.org/10.1007/978-3-319-50806-1
Kot, M. (2001). *Elements of mathematical ecology*. Cambridge University Press.
Kurtz, T. G. (1970). Solutions of ordinary differential equations as limits of pure jump markov processes. *Journal of Applied Probbaility*, 49–58. https://doi.org/10.2307/3212147
Lajmanovich, A., & Yorke, J. A. (1976). A deterministic model for gonorrhea in a nonhomogeneous population. *Mathematical Biosciences, 28*, 221–236. https://doi.org/10.1016/0025-5564(76)90125-5
Lipsitch, M., Cohen, T., Cooper, B., Robins, J. M., Ma, S., James, L., Gopalakrishna, G., Chew, S. K., Tan, C. C., Samore, M. H., Fisman, D., & Murray, M. (2003). Transmission dynamics and control of severe acute respiratory syndrome. *Science, 300*, 1966–1970. https://doi.org/10.1126/science.1086616
Ma, J. (2020). Estimating epidemic exponential growth rate and basic reproduction number. *Infectious Disease Modelling, 5*, 129–141. https://doi.org/10.1016/j.idm.2019.12.009
Ma, J., Dushoff, J., Bolker, B. M., & Earn, D. J. (2014). Estimating initial epidemic growth rates. *Bulletin of Mathematical Biology, 76*, 245–260. https://doi.org/10.1007/S11538-013-9918-2/FIGURES/6
Mena-Lorcat, J., & Hethcote, H. W. (1992). Dynamic models of infectious diseases as regulators of population sizes. *Journal of Mathematical Biology, 30*, 693–716. https://doi.org/10.1007/BF00173264
Metelmann, S., Pattni, K., Brierley, L., Cavalerie, L., Caminade, C., Blagrove, M. S., Turner, J., Sharkey, K. J., & Baylis, M. (2021). Impact of climatic, demographic and disease control factors on the transmission dynamics of covid-19 in large cities worldwide. *One Health, 12*, Article 100221. https://doi.org/10.1016/j.onehlt.2021.100221
Otsu, N. (1979). Threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern SMC-, 9*, 62–66. https://doi.org/10.1109/TSMC.1979.4310076
Overton, C. E., Wilkinson, R. R., Loyinmi, A., Miller, J. C., & Sharkey, K. J. (2022). Approximating quasi-stationary behaviour in network-based sis dynamics. *Bulletin of Mathematical Biology, 84*, 1–32. https://doi.org/10.1007/S11538-021-00964-7/FIGURES/4
Pourabbas, E., d'Onofrio, A., & Rafanelli, M. (2001). A method to estimate the incidence of communicable diseases under seasonal fluctuations with application to cholera. *Applied Mathematics and Computation, 118*, 161–174. https://doi.org/10.1016/S0096-3003(99)00212-X
Renshaw, E. (1993). *Modelling biological populations in space and time*. Cambridge University Press.
Sharkey, K. J., Kiss, I. Z., Wilkinson, R. R., & Simon, P. L. (2015). Exact equations for sir epidemics on tree graphs. *Bulletin of Mathematical Biology, 77*, 614–645. https://doi.org/10.1007/s11538-013-9923-5

Stadler, T., Vaughan, T. G., Gavryushkin, A., Guindon, S., Kühnert, D., Leventhal, G. E., & Drummond, A. J. (2015). How well can the exponential-growth coalescent approximate constant-rate birth–death population dynamics? *Proceedings of the Royal Society B: Biological Sciences, 282*. https://doi.org/10.1098/RSPB.2015.0420

van den Driessche, P. (2017). Reproduction numbers of infectious disease models. *Infectious Disease Modelling, 2*, 288–303. https://doi.org/10.1016/J.IDM.2017.06.002

Whittle, P. (1955). The outcome of a stochastic epidemic-a note on Bailey's paper. *Biometrika, 42*, 116–122. https://doi.org/10.1093/BIOMET/42.1-2.116