

RESEARCH ARTICLE

Estimation of the Basic Reproductive Number and Mean Serial Interval of a Novel Pathogen in a Small, Well-Observed Discrete Population

Kendra M. Wu^{1,2*}, Steven Riley²

1 School of Public Health, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, **2** Medical Research Council Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, Imperial College London, London, United Kingdom

* kendrawu@connect.hku.hk



OPEN ACCESS

Citation: Wu KM, Riley S (2016) Estimation of the Basic Reproductive Number and Mean Serial Interval of a Novel Pathogen in a Small, Well-Observed Discrete Population. PLoS ONE 11(2): e0148061. doi:10.1371/journal.pone.0148061

Editor: Edward Goldstein, Harvard School of Public Health, UNITED STATES

Received: August 1, 2014

Accepted: January 12, 2016

Published: February 5, 2016

Copyright: © 2016 Wu, Riley. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All results in the paper can be reproduced using R code available in the github repository <https://github.com/c97sr/EpilInference>.

Funding: KMW has received scholarships from Swire Company and The University of Hong Kong as well as funding from the LKS Faculty of Medicine of The University of Hong Kong. SR has received research funding from the Medical Research Council (UK, Project MR/J008761/1), the Wellcome Trust (UK, Project 093488/Z/10/Z), the Fogarty International Centre (USA, R01 TW008246-01), Fogarty International Centre with the Science &

Abstract

Background

Accurately assessing the transmissibility and serial interval of a novel human pathogen is public health priority so that the timing and required strength of interventions may be determined. Recent theoretical work has focused on making best use of data from the initial exponential phase of growth of incidence in large populations.

Methods

We measured generational transmissibility by the basic reproductive number R_0 and the serial interval by its mean T_g . First, we constructed a simulation algorithm for case data arising from a small population of known size with R_0 and T_g also known. We then developed an inferential model for the likelihood of these case data as a function of R_0 and T_g . The model was designed to capture a) any signal of the serial interval distribution in the initial stochastic phase b) the growth rate of the exponential phase and c) the unique combination of R_0 and T_g that generates a specific shape of peak incidence when the susceptible portion of a small population is depleted.

Findings

Extensive repeat simulation and parameter estimation revealed no bias in univariate estimates of either R_0 and T_g . We were also able to simultaneously estimate both R_0 and T_g . However, accurate final estimates could be obtained only much later in the outbreak. In particular, estimates of T_g were considerably less accurate in the bivariate case until the peak of incidence had passed.

Technology Directorate, Department of Homeland Security (USA, RAPIDD program), and the National Institute for Health Research (UK, for Health Protection Research Unit funding). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Conclusions

The basic reproductive number and mean serial interval can be estimated simultaneously in real time during an outbreak of an emerging pathogen. Repeated application of these methods to small scale outbreaks at the start of an epidemic would permit accurate estimates of key parameters.

Introduction

Uncertainty dominates important early policy decisions for emerging respiratory pathogens such as Severe Acute Respiratory Syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV), and pandemic influenza A/H1N1. This uncertainty is reduced greatly once the basic reproductive number and the serial interval of infection are known. The basic reproductive number quantifies the transmissibility of a pathogen and is defined as the average number of secondary infections generated by one typically infectious individual in an otherwise susceptible population. The serial interval in each infector-infectee pair is the time between disease onset in the infector and the onset in the infectee.

In this study, we present a temporal likelihood model that allows real-time simultaneous estimation of R_0 and the average of serial interval T_g in a small well-observed population, so that it provides timely information for informed public health responses. We build on previous inferential studies [1–4] by focusing on capturing unbiased parameter estimates as the epidemic leaves the exponential phase. Our two illustrative scenarios were designed to have similar exponential growth phases and are based loosely on pandemic influenza A/H1N1 and SARS-CoV infections.

Methods

We first define the underlying stochastic process that simulates an epidemic, before describing our calculation of the likelihood of data arising from this simulated process and the parameter estimation based on this likelihood. The required inputs of the likelihood estimation are; the number of infected people from outside the community who triggered the epidemic, the size of susceptible in the community, and the case incidence time-series. Conversely, the primary outcomes are estimates of the basic reproductive number, R_0 , and the mean serial interval, T_g .

Simulate epidemic process

In the analysis, we defined our simulation process to produce the random time series

$$\mathbf{X} = \{X_1, \dots, X_j, \dots, X_T\}$$

of new cases with onsets at time 1 to T. Specific realisations of \mathbf{X} were defined to be $\mathbf{x} = \{x_1, \dots, x_j, \dots, x_T\}$.

We defined serial interval by assuming the probability that an individual with symptom onset on day i generates a new infection on day $i + \tau$ is

$$f_s(\tau, Tg) \propto \frac{Tg^\tau \exp(-Tg)}{\tau!} \tag{1}$$

where the right-hand side is the probability mass τ for a Poisson distribution with mean Tg . Implicitly, this assumption about the serial interval implies that infectiousness was not

constant for individuals in the days following their infection. The value of k , where $k \gg T_g$, was always sufficiently high that the probability of a secondary case was negligible after k relative to the infector's onset time i . The offspring distribution [5] also assumed to be Poisson distributed with mean R_0 .

To initiate the simulation we assumed that there were n initially infectious individuals in a population of size N . Those initially infectious were assumed to have onsets at time t_1 . The simulation algorithm kept track of the onsets of the total number of infected individuals at each time unit $\mathbf{t} = \{t_1, \dots, t_j, \dots, t_T\}$ during the observed period. Symbolically, the total number of infected individuals having onsets prior to and at time j was denoted as I_j , where $I_j = \sum_{t=1}^j X_t$. For example $I_1 = n$.

For an arbitrary infector of onset time i , we first simulated the number of secondary cases it had infected from the offspring distribution of mean $R_0, f_O(R_0)$, without considering the depletion of the susceptible population. For each of these secondary cases, we then drew the delay from the onset of the infector i to the onset of each secondary case from the serial interval distribution. The onsets fall between $\{i + 1, \dots, i + k\}$ with mean $i + T_g$. This process yielded the secondary case time-series \mathbf{x} for a specific infector that had an onset time i .

When the number of infectees becomes significant relative to the susceptible population size, saturation effects must be accounted for. The effects are two-fold [6, 7]: first, through susceptible depletion, the number of secondary cases per primary case is reduced until there are no longer any susceptible people in the homogeneous population, which is when the number of secondary cases per primary case is reduced. Second, when there are many infectious cases, competition to infect the remaining susceptible shortens the serial interval. In this manuscript, we took the serial interval distribution to mean the distribution in the absence of competing sources of infection. The mean T_g was the average over of this distribution in the absence of competition.

In order to represent the finite pool of susceptible individuals, we defined the actual number of new cases with onsets at time j that accounted for the depletion of the susceptible population as \tilde{z}_j , and they were drawn from the as-yet uninfected pool using a binomial distribution, with probability of infection equal to the ratio of z_j and N :

$$\tilde{z}_j \sim \mathcal{B}\left(N - I_{j-1}, \frac{z_j}{N}\right). \tag{2}$$

We repeated this process for each infector at time i and added up the secondary cases between $\{i + 1, \dots, i + k\}$. It should be noted that each infector produced different numbers of secondary cases and each of these infector-infectee pairs had different serial intervals. Each of these secondary cases were assumed to be infectious and were considered subsequent infectors if $i > j$. We repeated this process for all infectors with onset time between 1 and T . The overall case time-series \mathbf{X} was the sum of each of the realization of secondary case time-series \mathbf{x} of each infector.

Likelihood

We defined Y_j to be the expected number of secondary cases of infectors X_i from $\{j - k, \dots, j - 1\}$ with onsets at time j that could be expressed as,

$$Y_j = \sum_{i=j-k}^{j-1} Y_i R_0 f_s(j - i, T_g). \tag{3}$$

where f_s was the Poisson density of serial interval.

We note that the formulation used above may not be correct for offspring and serial interval distributions that are not Poisson.

Therefore, the likelihood of a whole times series of observations could then be expressed in terms of the expected numbers of secondary cases in each time unit and the parameters of the model, when accounting to the finite pool of susceptible, as such:

$$L(\tilde{\mathbf{Y}}) = \prod_{j=2}^T p_B \left(\tilde{Y}_j | N - I_{j-1}, \frac{Y_j}{N} \right) \quad (4)$$

where p_B was the probability mass function of a binomial distribution, and Y_j is defined in [Eq \(3\)](#).

Standard maximum likelihood were used to obtain point estimates and confidence intervals. We used the optim function in R [8] to find point estimates and the optimise function to find univariate confidence regions.

To initiate an epidemic and ensure that it would form a chain of transmission to establish an exponential phase of growth without early extinction, we decided to seed fifty infectious individuals on day 1 in each of our simulated epidemics. The subsequent number of incidences were then generated by the stochastic process described in the Simulate epidemic process.

All results in the paper can be reproduced using R code available in the github repository [c97sr/EpiInference](#).

Illustrative parameter regimes

We constructed two illustrative scenarios to investigate the properties of our model and likelihood function. Scenario 1: $R_0 = 1.8$, $T_g = 2.5$ when using Poisson-distributed serial interval distribution; and Scenario 2: $R_0 = 3.0$, $T_g = 6.25$ when using Poisson-distributed serial interval distribution. We chose these scenarios because they cover the range of SARS in hospitals ($R_0 = 2.7$ [9]), pandemic A/H1N1 influenza in the community ($R_0 = 1.5$ [10, 11]) and pandemic A/H1N1 in schools ($R_0 = 2.4$ [12]). Then we compute T_g so that they had growth rate during the exponential phase was 0.32 according to $r = \frac{R_0 - 1}{T_g}$ [13]. Unless stated otherwise, we seeded day 1 with fifty individuals to avoid the epidemic prematurely died out. In each scenario, 1,000 people were assumed to be living in this community, where all but fifty were assumed to be susceptible initially. This population size was chosen to be consistent with a small town or large village in many social settings.

When using Poisson-distributed serial interval, we set $k = 10$ in Scenario 1 to simulate influenza infections [14] and $k = 25$ in Scenario 2 to loosely mimic SARS infections [15, 16]. They were chosen so that these values were at least four times greater than T_g and the probability of a secondary case after k , relative to the onset time i of the infector, was negligible. The effect of the truncation of the distribution was subsequently pushed to the right-hand tail of the estimate time-series when $(i + k) > T$.

Results

The different stages of an epidemic pass quickly in an outbreak of respiratory infections in a small population ([Fig 1](#)). In our simulated population of 1,000 people for an influenza-like pathogen (Scenario 1) there was an average delay of only 16.5 days (median, 95% prediction interval: 8.4, 26.2) from the introduction of 10 infectious individuals to peak incidence. The more transmissible, but slower Scenario 2 still achieved a peak of incidence after 23 days (18.1, 29.2). In addition to an early peak of incidence, partial saturation occurred very quickly in this population. An initial exponential increase in incidence can be seen on the log y-axes in [Fig 1](#) (as a straight line). However, this exponential phase ends quickly in such a small population. In

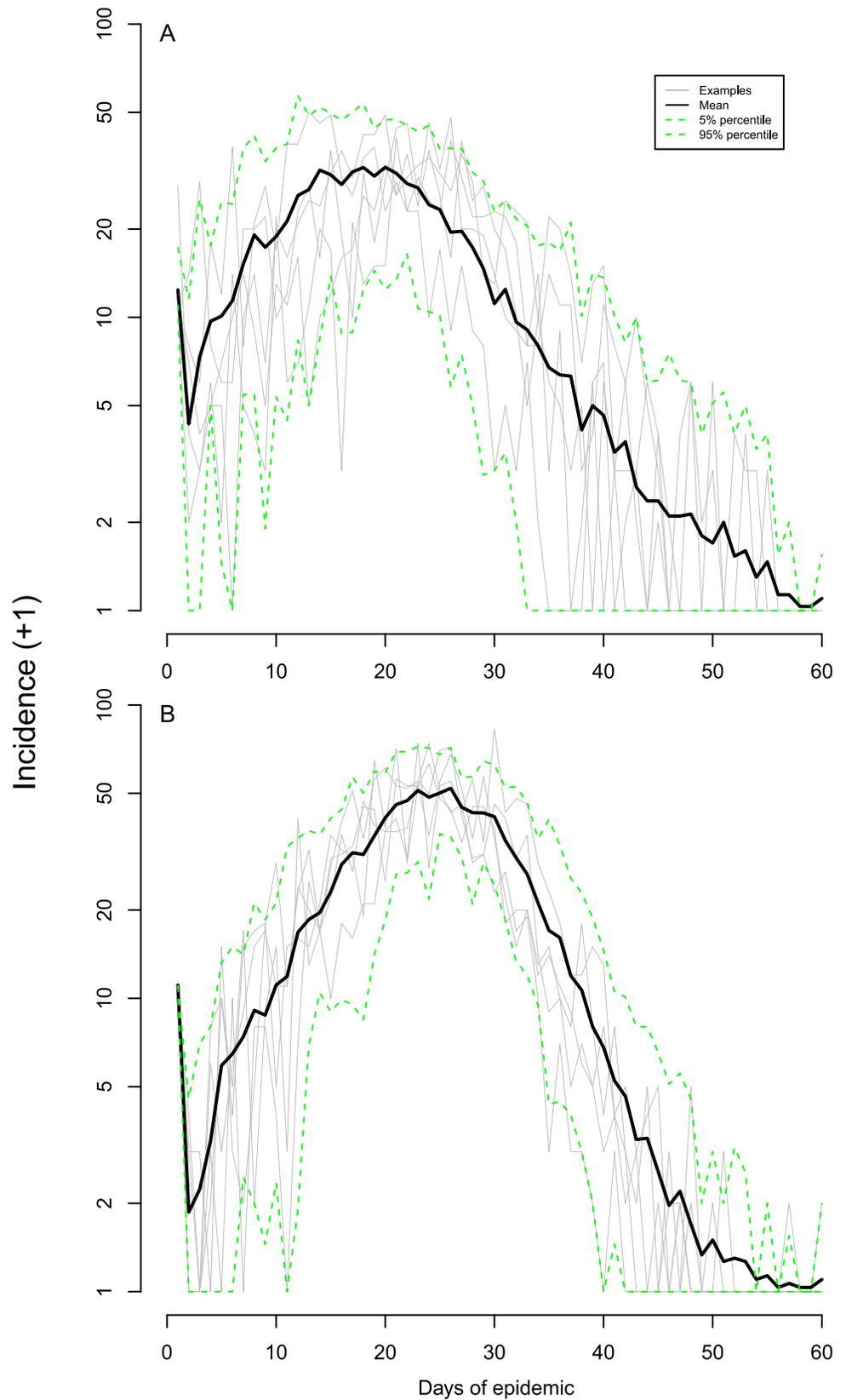


Fig 1. Simulated epidemics. Part **A** shows Scenario 1 with $R_0 = 1.8$, $T_g = 2.5$. Part **B** shows Scenario 2 with $R_0 = 3.0$, $T_g = 6.25$. Offspring and the serial interval assumed to follow Poisson distribution. In both scenarios, 1,000 individuals were initially susceptible and 10 people were infectious on day 1. Grey lines trace 5 single example realisations. Black lines show average case incidence, and the green dotted lines are the 5% and 95% prediction intervals from 30 realisations.

doi:10.1371/journal.pone.0148061.g001

Scenario 1, incidence appears to be occurring sub-exponentially as early as 10 days after the initial seed.

The speedy transition from exponential growth to partial depletion of susceptible individuals and then peak incidence permitted rapid and accurate univariate parameter estimation. We attempted to recover either R_0 or T_g (but not both at this stage) from simulated data similar to those described above. As would be expected, for both Scenario 1 (Fig 2a and 2b showed 8 arbitrary epidemics and prediction intervals of 100 epidemics) and Scenario 2 (Fig 3a and 3b showed 8 arbitrary epidemics and prediction intervals of 100 epidemics), we were able to obtain accurate values of R_0 prior to the peak of the epidemics when T_g was known, even during the earliest exponential phase of incidence. Estimates of R_0 did not become any more accurate once the peak of incidence had passed. Although the pattern of univariate inference for T_g was similar to that for R_0 , the final confidence intervals were wider (in relative terms) than for R_0 .

We extended our analysis of the univariate parameter estimation routines to test the model's performance (Fig 4 for Scenario 1). For a day 50 in the simulated epidemic of 50 realizations, we held R_0 and T_g constant in the simulated epidemics while varying the values of a) R_0 estimates, denoted as \hat{R}_0 , or b) T_g estimates, denoted as \hat{T}_g , by ± 1.5 in the likelihood function. Next, we repeated the process by c) varying R_0 from 1.0 to 3.0 and d) varying T_g from 1.1 to 10.0 in the simulated epidemics, then computed the residuals of c) \hat{R}_0 and d) \hat{T}_g from the likelihood function respectively as previously done.

Nevertheless, it was possible to jointly estimate both R_0 and T_g in this small population. We assumed that neither T_g nor R_0 were known for an emerging pathogen and attempted to estimate them jointly for Scenario 1 (Fig 2c and 2d showed 8 arbitrary epidemics and prediction intervals of 100 epidemics) and Scenario 2 (Fig 3c and 3d showed 8 arbitrary epidemics and prediction intervals of 100 epidemics). Compared to the univariate cases, for both scenarios, reliable information was obtained on both R_0 and T_g . However, prediction intervals were wider. Once the peak had been reached, the convergence of estimates of T_g in bivariate case were achieved, but in relative terms, the accuracy was considerably lower.

The estimates of different values of k were tabulated on Table 1 for the model using Poisson-distributed serial interval for the estimates on the last day of observation using different numbers of iterations: (i) 10, (ii) 20, (iii) 100, along with different k values.

Discussion

We have described a quantitative method that allows the rapid joint estimation of the basic reproductive number, and mean and variance of serial interval for an emerging pathogen based only on case data from an outbreak in a small population. These estimates show no signs of bias and could be available quickly for small populations. In essence, this approach is able to tease apart the two parameters from case data because it extracts key information from the transmission dynamics of different phases of a respiratory epidemic: the initial stochastic period of growth, the deterministic period of exponential growth, the sub-exponential period of growth prior to peak incidence, and the peak itself. Previous similar work [1, 2, 4] has incorporated either the initial stochastic phase or one or more of the deterministic phases, but not all phases together.

The serial interval distribution plays a crucial role in the estimation accuracy. It is because when the variance was large, the estimator had more difficulty to tease apart the two parameters from the observed incidence data, especially during phase II and beyond. In this model, the serial intervals are assumed to follow a Poisson distribution. When it has large variance and thus less distinguishable between generations of infection comparing to a shape Gamma

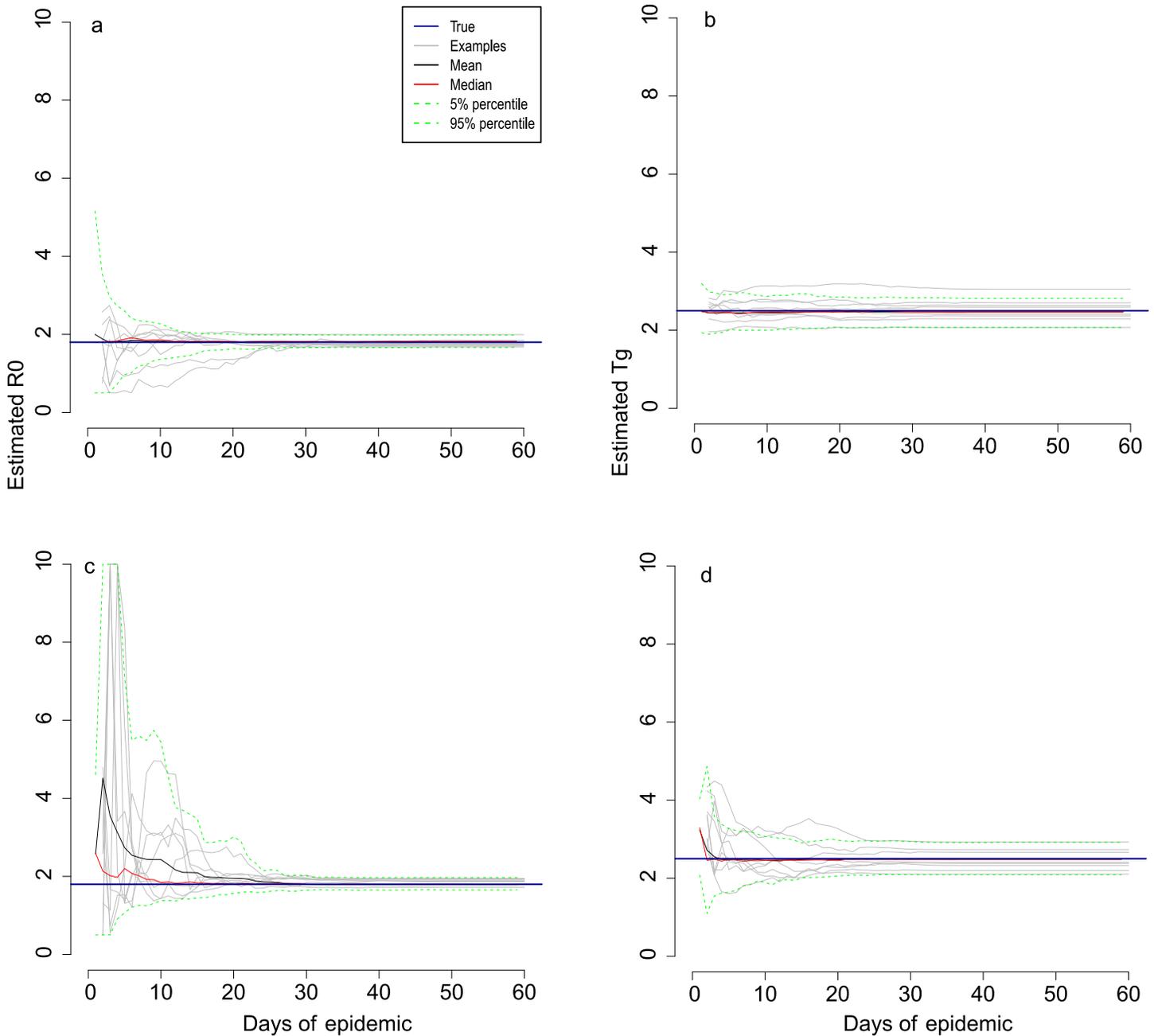


Fig 2. Univariate and bivariate parameter estimation from simulated daily data for Scenario 1. Part **a** shows real-time univariate estimates of R_0 in which the true value of T_g was assumed known. Part **b** shows real-time univariate estimates of T_g with R_0 assumed known. The results of jointly estimating R_0 and T_g are shown in parts **c** (R_0) and **d** (T_g). Scenario 1 is as per Fig 1A with $R_0 = 1.8$, $T_g = 2.5$. The epidemic was simulated in a population of 1,000 people all of whom were initially susceptible, other than 50 who were infectious on day 1. Grey lines trace the corresponding estimates of 8 arbitrary epidemics. Black lines represent the average of 100 estimates at that time point, red lines represent the median, and the green dotted lines are the 5% and 95% prediction intervals for 100 epidemics. The dark blue solid horizontal lines show the true parameter values.

doi:10.1371/journal.pone.0148061.g002

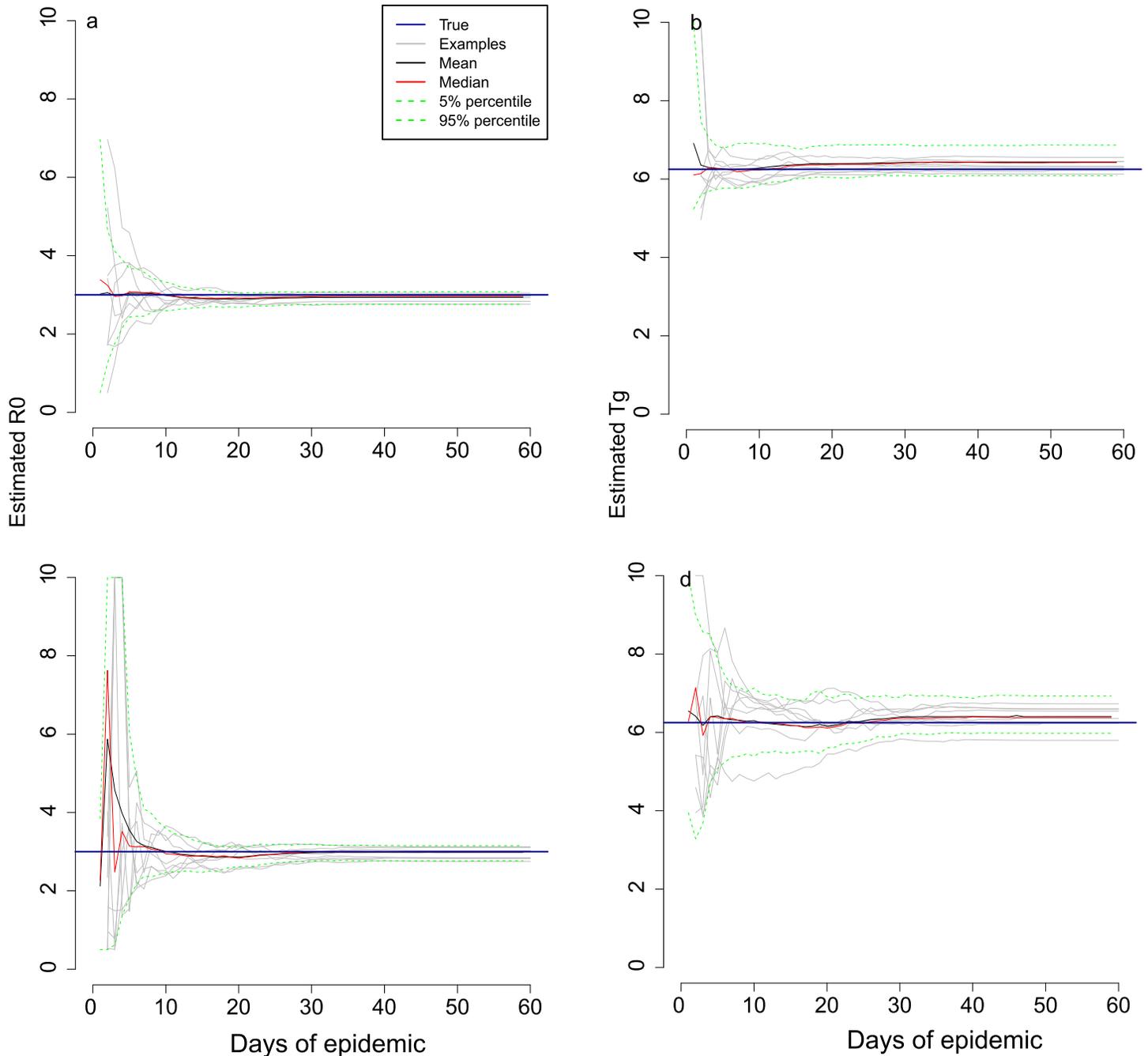


Fig 3. Univariate and bivariate parameter estimation from simulated daily data for Scenario 2. Part a shows real-time univariate estimates of R_0 in which the true value of T_g was assumed known. Part b shows real-time univariate estimates of T_g with R_0 assumed known. The results of jointly estimating R_0 and T_g are shown in parts c (R_0) and d (T_g). Scenario 2 is as per Fig 1B with $R_0 = 3.0$, $T_g = 6.25$. The epidemic was simulated in a population of 1,000 people all of whom were initially susceptible, other than 100 who were infectious on day 1. Grey lines trace the corresponding estimates of 8 arbitrary epidemics. Black lines represent the average of 100 estimates at that time point, red lines represent the median, and the green dotted lines are the 5% and 95% prediction intervals for 100 epidemics. The dark blue solid horizontal lines show the true parameter values.

doi:10.1371/journal.pone.0148061.g003

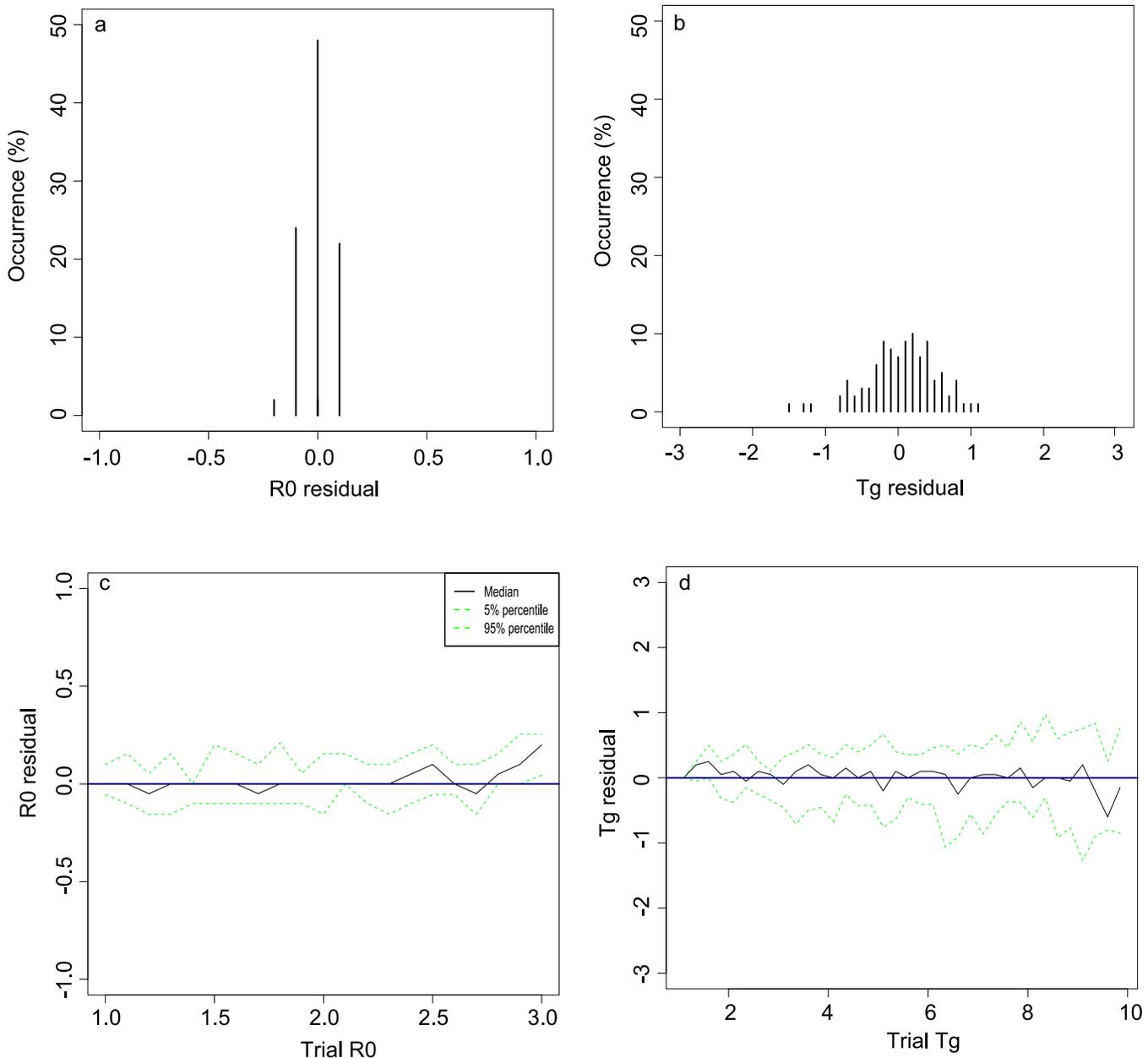


Fig 4. Extended assessment of univariate inference for Scenario 1. Sensitivity analysis using univariate estimations at single time point and assumed poisson-distributed serial interval. **a** and **b** show the distribution of the marginals for estimation of R_0 and T_g of 50 realisations respectively. All estimates were made at day 50. **c** shows the mean and confidence intervals for the marginals for estimates for R_0 across a wide range of known values for T_g . Similarly, **d** shows the mean and confidence intervals for T_g across a wide range of value for R_0 . Scenario 1 is as per Fig 1A with $R_0 = 1.8$, $T_g = 2.5$, and $k = 10$.

doi:10.1371/journal.pone.0148061.g004

distribution, the joint estimations have poorer performance, although they remain reasonably accurate in infinite population [3]. In finite population, the estimations of \hat{T}_g showed no bias (Table 1) if it had sufficiently large seed size.

By seeding fifty and a hundred infectious people on day 1 in each epidemic, the number of incidences at the early stage of the outbreaks would likely be higher and plausibly reached the exponential phase of growth earlier than if there were smaller number of infectious individuals

Table 1. Univariate and bivariate estimates of the model using Poisson distributed serial interval, different k values, and number of simulation iterations. The following parameters were used: (a) $R_0 = 1.8, T_g = 2.5, k = 10$; (b) $R_0 = 1.8, T_g = 2.5, k = 25$; (c) $R_0 = 3.0, T_g = 6.25, k = 10$; (d) $R_0 = 3.0, T_g = 6.25, k = 25$. PI: prediction interval.

Parameters	Iterations	Univariate estimations					
		R0 estimates			Tg estimates		
		Mean	Median	(5%, 95%)	Mean	Median	(5%, 95%)
a	10	1.74	1.74	(1.58, 1.90)	2.55	2.50	(2.15, 2.97)
	20	1.79	1.81	(1.66, 1.90)	2.50	2.44	(2.06, 2.91)
	100	1.79	1.79	(1.63, 1.93)	2.48	2.48	(2.01, 2.96)
b	10	1.80	1.77	(1.73, 1.90)	2.50	2.53	(2.12, 2.79)
	20	1.81	1.80	(1.66, 2.00)	2.48	2.41	(2.17, 3.04)
	100	1.82	1.81	(1.70, 1.96)	2.46	2.47	(2.07, 2.82)
c	10	2.99	3.01	(2.76, 3.16)	6.27	6.22	(6.06, 6.65)
	20	2.99	3.02	(2.82, 3.14)	6.31	6.28	(5.89, 6.63)
	100	2.99	2.99	(2.82, 3.18)	6.31	6.31	(5.95, 6.68)
d	10	2.93	2.94	(2.75, 3.06)	6.33	6.38	(6.01, 6.57)
	20	2.92	2.91	(2.80, 3.07)	6.40	6.45	(6.02, 6.68)
	100	2.99	2.99	(2.77, 3.22)	6.28	6.24	(5.91, 6.72)
Bivariate estimations							
		Mean	Median	(5%, 95%)	Mean	Median	(5%, 95%)
a	10	1.74	1.73	(1.58, 1.91)	2.50	2.41	(2.18, 2.89)
	20	1.78	1.80	(1.64, 1.89)	2.48	2.47	(2.01, 2.97)
	100	1.79	1.78	(1.64, 1.91)	2.47	2.46	(1.99, 2.98)
b	10	1.80	1.80	(1.71, 1.94)	2.50	2.47	(2.09, 2.90)
	20	1.81	1.79	(1.69, 1.98)	2.48	2.48	(2.19, 2.92)
	100	1.81	1.81	(1.71, 1.95)	2.48	2.48	(2.09, 2.92)
c	10	2.99	3.00	(2.75, 3.18)	6.26	6.26	(5.97, 6.64)
	20	3.01	2.99	(2.79, 3.18)	6.32	6.32	(6.04, 6.68)
	100	3.01	2.99	(2.81, 3.24)	6.32	6.29	(5.99, 6.78)
d	10	2.92	2.94	(2.71, 3.09)	6.23	6.18	(5.89, 6.61)
	20	2.96	2.92	(2.75, 3.18)	6.31	6.31	(5.80, 6.74)
	100	3.00	2.97	(2.77, 3.27)	6.27	6.24	(5.88, 6.70)

doi:10.1371/journal.pone.0148061.t001

seeded. However, with the same transmissibility and speed of transmission, higher number of incidences and large seed size appear to have no impact to the accuracy of estimates aside from avoiding outbreaks prematurely died out.

Our study is intended as an initial theoretical illustration of the potential for this approach to be applied more widely. As such, it suffers from a number of limitations caused directly by the restrictive assumptions we have made. For example, we have assumed that all cases in this small population were observed. However, despite the possibility of intense surveillance in a small population, it is unlikely that even all symptomatic cases will be detected. Also, most pathogens generate some mild or entirely asymptomatic cases that would evade even the most intense surveillance efforts. Although we doubt, based on related previous study [17] that the reporting of only a proportion of cases would bias results (if that proportion is constant), it would most likely increase the delay from the start of the epidemic to the acquisition of robust parameter estimates.

We have also assumed throughout that offspring distribution and serial interval distribution follow a Poisson density. These are convenient assumptions that allow very rapid likelihood

calculations for it is a single-parameter defining function. We anticipate few theoretical issues in the extension of these approaches to include distributions with more than one parameter (other than careful modification of Eq (3) as indicated above). However, the ability of these methods to independently estimate both means and variances of the offspring distribution and serial interval distribution remains to be described. We would anticipate that the estimation of variance for the serial interval distribution would be challenging [18] for cases other than where there are very high volumes of data for the first few generations (which could arise from a large seeding event) or when the true distribution has a low variance.

There are likely substantial opportunities to gain valuable insights into outbreak dynamics if we relax our assumption that transmissibility is constant over time and thus consider an inferential framework that accounts for both changes in transmissibility and depletion of susceptible individuals. The reproduction number would vary over time because of both effects. This approach might be especially useful for severe infections that affect small populations, such as Ebola.

Acknowledgments

KMW thanks MRC Centre for Outbreak Analysis and Modelling of Imperial College London for hosting her visiting research fellowship and thanks LKS Faculty of Medicine of The University of Hong Kong for funding her stay in London, where the foundation of the theoretical work on this paper was partially completed. KMW also acknowledges scholarship support from John Swire & Sons and The University of Hong Kong, as well as funding from the LKS Faculty of Medicine of The University of Hong Kong.

Author Contributions

Conceived and designed the experiments: SR. Performed the experiments: KMW. Analyzed the data: KMW SR. Wrote the paper: KMW SR.

References

1. White LF, Pagano M. A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic. *Stat Med*. 2008; 27(16):2999–3016. doi: [10.1002/sim.3136](https://doi.org/10.1002/sim.3136) PMID: [18058829](https://pubmed.ncbi.nlm.nih.gov/18058829/)
2. White LF, Wallinga J, Finelli L, Lyn C, Riley S, Lipsitch M, et al. Estimation of the reproductive number and the serial interval in early phase of the 2009 influenza A/H1N1 pandemic in the USA. *Influenza Other Respir Viruses*. 2009; 3(6):267–276. doi: [10.1111/j.1750-2659.2009.00106.x](https://doi.org/10.1111/j.1750-2659.2009.00106.x) PMID: [19903209](https://pubmed.ncbi.nlm.nih.gov/19903209/)
3. Griffin JT, Garske T, Ghani AC, Clarke PS. Joint estimation of the basic reproduction number and generation time parameters for infectious disease outbreaks. *Biostatistics*. 2011; 12(2):303–312. doi: [10.1093/biostatistics/kxq058](https://doi.org/10.1093/biostatistics/kxq058) PMID: [20858771](https://pubmed.ncbi.nlm.nih.gov/20858771/)
4. Wallinga J, Teunis P. Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures. *Am J Epidemiol*. 2004; 160(6):506–516. doi: [10.1093/aje/kwh255](https://doi.org/10.1093/aje/kwh255)
5. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. *Nature*. 2005; 438:355–359. doi: [10.1038/nature04153](https://doi.org/10.1038/nature04153) PMID: [16292310](https://pubmed.ncbi.nlm.nih.gov/16292310/)
6. Svensson A. A note on generation times in epidemic models. *Math Biosci*. 2006; 208(2007):300–311. PMID: [17174352](https://pubmed.ncbi.nlm.nih.gov/17174352/)
7. Kenah E, Lipsitch M, Robins JM. Generation interval contraction and epidemic data analysis. *Math Biosci*. 2008; 213(1):71–79. Available from: <http://www.sciencedirect.com/science/article/pii/S0025556408000394> doi: [10.1016/j.mbs.2008.02.007](https://doi.org/10.1016/j.mbs.2008.02.007) PMID: [18394654](https://pubmed.ncbi.nlm.nih.gov/18394654/)
8. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2013. ISBN 3-900051-07-0. Available from: <http://www.R-project.org/>
9. Riley S, Fraser C, Donnelly CA, Ghani AC, Abu-Raddad LJ, Hedley AJ, et al. Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions. *Science*. 2003; 300(5627):1961–1966. doi: [10.1126/science.1086478](https://doi.org/10.1126/science.1086478) PMID: [12766206](https://pubmed.ncbi.nlm.nih.gov/12766206/)

10. Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Kerkhove MDV, Hollingsworth TD, et al. Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings. *Science*. 2009; 324(5934):1557–1561. doi: [10.1126/science.1176062](https://doi.org/10.1126/science.1176062) PMID: [19433588](https://pubmed.ncbi.nlm.nih.gov/19433588/)
11. Boëlle PY, Ansart S, Cori A, Valleron AJ. Transmission parameters of the A/H1N1 (2009) influenza virus pandemic: a review. *Influenza Other Respi Viruses*. 2011 September; 5(5):306–316. doi: [10.1111/j.1750-2659.2011.00234.x](https://doi.org/10.1111/j.1750-2659.2011.00234.x)
12. Yang Y, Sugimoto JD, Halloran ME, Basta NE, Chao DL, Matrajt L, et al. The Transmissibility and Control of Pandemic Influenza A (H1N1) Virus. *Science*. 2009; 326(5953):729–733. doi: [10.1126/science.1177373](https://doi.org/10.1126/science.1177373) PMID: [19745114](https://pubmed.ncbi.nlm.nih.gov/19745114/)
13. Anderson RM, May RM. A framework for discussing the population biology of infectious diseases. In: *Infectious disease of humans: dynamics and control*. Oxford Univ Press; 1992. p. 17–19.
14. Lau LLH, Cowling BJ, Fang VJ, Chan KH, Lau EHY, Lipsitch M, et al. Viral Shedding and Clinical Illness in Naturally Acquired Influenza Virus Infections. *J Infect Dis*. 2010; 201(10):1509–1516. doi: [10.1086/652241](https://doi.org/10.1086/652241) PMID: [20377412](https://pubmed.ncbi.nlm.nih.gov/20377412/)
15. Anderson RM, Fraser C, Ghani AC, Donnelly CA, Riley S, Ferguson NM, et al. Epidemiology, transmission dynamics and control of SARS: the 2002–2003 epidemic. *Philos Trans R Soc Lond B Biol Sci*. 2004; 359(1447):1091–1105. doi: [10.1098/rstb.2004.1490](https://doi.org/10.1098/rstb.2004.1490) PMID: [15306395](https://pubmed.ncbi.nlm.nih.gov/15306395/)
16. Donnelly CA, Ghani AC, Leung GM, Hedley AJ, Fraser C, Riley S, et al. Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet*. 2003; 361:1761–1766. doi: [10.1016/S0140-6736\(03\)13410-1](https://doi.org/10.1016/S0140-6736(03)13410-1) PMID: [12781533](https://pubmed.ncbi.nlm.nih.gov/12781533/)
17. Kucharski A, Mills H, Pinsent A, Fraser C, Van Kerkhove M, Donnelly CA, et al. Distinguishing Between Reservoir Exposure and Human-to-Human Transmission for Emerging Pathogens Using Case Onset Data. *PLoS currents*. 2013; 6.
18. Kucharski AJ, Mills HL, Donnelly CA, Riley S. Transmission potential of Influenza A(H7N9) virus, China, 2013–2014. *Emerg Infect Dis*. 2015; 21(5):852–855. doi: [10.3201/eid2105.141137](https://doi.org/10.3201/eid2105.141137) PMID: [25897624](https://pubmed.ncbi.nlm.nih.gov/25897624/)