Available online at www.sciencedirect.com
SCIENCE DIRECT®
Mathematical Biosciences 198 (2005) 14–26
Mathematical Biosciences
www.elsevier.com/locate/mbs

Estimation and inference of R_0 of an infectious pathogen by a removal method

Matthew J. Ferrari [a,*], Ottar N. Bjørnstad [b], Andrew P. Dobson [c]

[a] *IGDP in Ecology, 501 ASI Building, The Pennsylvania State University, University Park, PA 16802, USA*
[b] *Department of Entomology and Biology, The Pennsylvania State University, University Park, PA 16801, USA*
[c] *Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA*

Received 9 August 2004; received in revised form 20 April 2005; accepted 11 August 2005
Available online 7 October 2005

Abstract

The basic reproductive ratio, R_0, is a central quantity in the investigation and management of infectious pathogens. The standard model for describing stochastic epidemics is the continuous time epidemic birth-and-death process. The incidence data used to fit this model tend to be collected in discrete units (days, weeks, etc.), which makes model fitting, and estimation of R_0 difficult. Discrete time epidemic models better match the time scale of data collection but make simplistic assumptions about the stochastic epidemic process. By investigating the nature of the assumptions of a discrete time epidemic model, we derive a bias corrected maximum likelihood estimate of R_0 based on the chain binomial model. The resulting 'removal' estimators provide estimates of R_0 and the initial susceptible population size from time series of infectious case counts. We illustrate the performance of the estimators on both simulated data and real epidemics. Lastly, we discuss methods to address data collected with observation error.

Keywords: Basic reproductive ratio; Birth-and-death model; Estimation; Chain binomial; R_0

* Corresponding author. Tel.: +1 814 865 1815; fax: +1 814 865 3048.
E-mail addresses: mferrari@psu.edu (M.J. Ferrari), onb1@psu.edu (O.N. Bjørnstad), dobber@princeton.edu (A.P. Dobson).

0025-5564/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.mbs.2005.08.002

1. Introduction

The basic reproductive ratio, R_0, is a critical parameter in the dynamics of infectious diseases [1,2]. This parameter is defined as the number of secondary cases resulting from a single infectious individual in a population of susceptible individuals [3]. A pathogen with R_0 greater than unity will, on average, initiate an epidemic, whereas a pathogen with $R_0 < 1$ will not [4]. Further, R_0 is a determinant of the duration of a closed epidemic within a naïve population and the total number of susceptible individuals that will be infected [3]. The proportion of the population that needs to be removed (either through vaccination or culling) from the pool of susceptibles in order to prevent or preempt an epidemic, is given by $1 - 1/R_0$ [3]. Estimating R_0 is therefore important towards management and intervention of epidemic pathogens.

Despite its central theoretical and practical importance, estimation of R_0 has presented a challenge in epidemiology and disease ecology [2]. One obstacle is the difficulty of fitting the continuous time epidemic birth-and-death model [3] to incidence data [5]; infections are necessarily reported in discrete time units (days, weeks, months; [6]). Discrete time models with non-overlapping infectious generations match the discrete nature of the data but introduce additional approximations as they assume that infections and state transitions occur instantaneously and independently [6,7].

An obstacle to many R_0 estimators is the requirement of detailed knowledge of host demography [1,8–10] and/or retrospective studies of epidemics that have run their course [11,12]. Furthermore, reliance on estimates from previous and completed outbreaks assumes that epidemiological parameters remain constant in space and time (see [13] for a detailed illustration that counters this assumption). The possibility of temporal changes in pathogen virulence, host population size, and geometries of social and/or spatial transmission networks among hosts may result in deviations from historic parameters. In the face of an emergent epidemic, one may wish for a complementary method that can be implemented in real-time with minimal reliance on historical estimates.

In this paper, we develop a 'removal method' to estimate R_0. Removal methods for population estimation had an early history in wildlife and fisheries applications [14–16]. The method presented here bears conceptual similarities in that estimation is done on the basis of individuals removed by infection. We derive a likelihood for the time series of removals due to infection and calculate a correction for the bias introduced by the temporal (weekly, monthly, etc.) discretization of mortality and morbidity reports. We evaluate the performance of the resultant maximum likelihood estimator on simulated epidemic data and apply the method to data from some recent epidemics. Finally, as underreporting is common in time series of disease incidence we discuss the estimation of R_0 in the presence of measurement error.

2. Methods

Our question is a simple one: how can we use time series of case reports, such as those from weekly mortality and morbidity reports, to estimate R_0. Before proposing a method, however, we need to consider the relationship between temporally binned incidence reports and the underlying continuous time process.

2.1. Discrete observation of a continuous epidemic process

Consider a simple epidemic birth-and-death model with initial transmission rate $\lambda = \beta S_0$ and recovery rate γ; here S_0 represents the initial number of susceptibles. For such a process, the classic epidemic birth-and-death model assumes that the infectious period follows an exponential distribution with a mean length of $1/\gamma$ (though Keeling and Grenfell discuss more realistic distributions [17]). The expected number of new infections generated from each infectious individual at the onset of an epidemic (R_0) is given by the integral $R_0 = \int_0^\infty \lambda e^{-\gamma\tau}\, d\tau$; i.e. the transmission rate integrated over the probability of remaining infectious [2,4] (the area under the dashed curve in Fig. 1).

The data from which we must estimate the characteristics of epidemics, however, are generally observed in discrete time units (days, weeks, months, etc.). As a result, discrete time epidemic models have been proposed (e.g. chain-binomial [18], TSIR [7]) which better match the timescale on which data are gathered. A natural time step for these discrete models – assuming no latent period – is the mean infectious period, $1/\gamma$. We note that all the results presented below are easily generalized to models in which the time step is equal to the epidemic generation time (the sum of the latent and infectious periods). When discretizing one implicitly assumes that 'late' infections in the interval $(t + 1/\gamma, \infty)$ caused by index cases are balanced by infections from subsequent chains of infection in the interval $(t, t + 1/\gamma)$; i.e. the shaded regions in Fig. 1 are equal. By late infections we mean any that would occur if an individual remained infectious beyond the generation break at time $t + 1/\gamma$; by secondary, tertiary, etc. infections we mean any subsequent new infections due

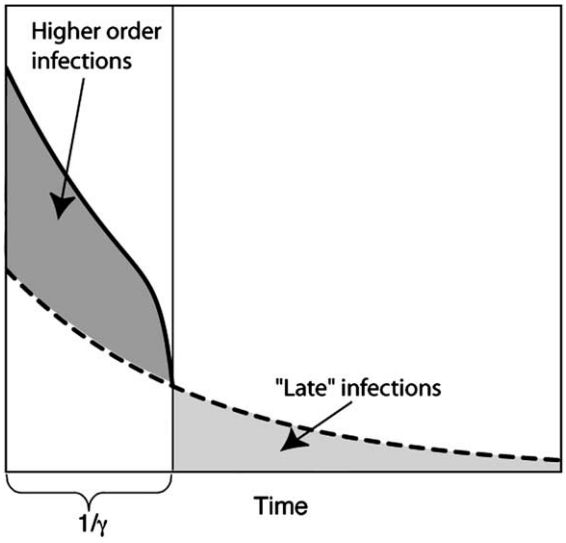

Fig. 1. Expected number of new infections under the epidemic birth-and-death model. Dashed line indicates rate of infection due to a single infectious host for the standard epidemic birth-and-death model. Dark shading indicates infections due to secondary (and higher) chains of infection within the mean infectious period, $1/\gamma$. Light shading indicates the expected number of primary infections that occur after the mean infectious period.

to early infections in the interval $(t, t + 1/\gamma)$. By balancing the late and secondary infections, we can calculate biases from using discretized mortality and morbidity data.

Given a primary infection at time τ, the probability of secondary (or greater) infection in the interval $(\tau, 1/\gamma)$ is $(1 - e^{-\lambda(1/\gamma-\tau)})$; that is, the complement of the null probability of a time invariant infectious process over the interval $(\tau, 1/\gamma)$. The approximate total number of new infections due to a single infectious individual in one infectious period is the expected number of primary infections plus any secondary infections, given a primary infection, etc. that falls within the time step (the area under the solid curve in Fig. 1):

$$I_{t+1/\gamma} \approx \int_0^{1/\gamma} [\lambda e^{-\gamma\tau} + \lambda e^{-\gamma\tau}(1 - e^{-\lambda(1/\gamma-\tau)})]\, d\tau.$$

Solving the integral gives,

$$\begin{aligned} I_{t+1/\gamma} &\approx \lambda \left[-\frac{2}{\gamma} e^{-\gamma\tau} \Big|_0^{1/\gamma} - \frac{1}{\lambda - \gamma} e^{-\lambda/\gamma} (e^{\tau(\lambda-\gamma)}) \Big|_0^{1/\gamma} \right] \\ &\approx \lambda \left[\frac{2}{\gamma}(1 - e^{-1}) \right] + \frac{\lambda}{\lambda - \gamma} e^{-\lambda/\gamma}(1 - e^{(\lambda-\gamma)/\gamma}). \end{aligned} \tag{2.1}$$

The second term in the sum is small, relative to the first and goes to e^{-1} as $\lambda \to \infty$. Thus, the expected number of new infections is approximately $\frac{2(1-e^{-1})\lambda}{\gamma} - e^{-1}$, rather than $\frac{\lambda}{\gamma}$ as implicitly assumed in many applications [2]. In this way, the number of infections that occur in the common time step, $1/\gamma$, for discrete models underestimates the true number of infections by a factor of about $2(1 - e^{-1}) - e^{-1}$ relative to the epidemic birth-and-death model.

2.2. Chain binomial likelihood

The classic stochastic model for discrete time epidemics is the so-called chain binomial [18]. Let S_t and I_t be the number of susceptible and infectious individuals at the end of time period t, and let the time step, as above, be $1/\gamma$. Let β be the rate of contacts adequate to transfer infection between any two members of the population in interval Δt. Then $e^{-\beta I_t}$ is the probability that any given susceptible individual will have no contacts with infectious individuals during Δt. The complement is then $1 - e^{-\beta I_t}$. From this, we can write the infectious time series as a chain of binomial random variables,

$$I_{t+1} \sim \text{binomial}(S_t, 1 - e^{-\beta I_t}) \tag{2.2}$$

and the balance equation for the number of susceptible individuals is given by

$$S_{t+1} = S_t - I_{t+1}.$$

Noting that S_{t+1} can be re written as $S_t = S_0 - \sum_{i=1}^{t} I_i$, where S_0 is the size of the initial susceptible population, we write the conditional probability of I_{t+1} as

$$P(I_{t+1}|I_t, \ldots, I_1, S_0, \beta) = \binom{S_0 - \sum_{i=1}^{t} I_i}{I_{t+1}} (1 - e^{-\beta I_t})^{I_{t+1}} (e^{-\beta I_t})^{S_0 - \sum_{i=1}^{t} I_i - I_{t+1}}, \tag{2.3}$$

which is a function of the time series of infected individuals to time t, β, and the initial susceptible population size, S_0. By the law of total probability we can write the likelihood for both β and S_0 given the entire time series of infectives, $\mathbf{I} = [I_t,\ t = 0, \ldots, T]$, as

$$L(S_0, \beta|\mathbf{I}) = \prod_{t=1}^{T} P(I_t|I_{t-1}, \ldots, I_1, S_0, \beta), \tag{2.4}$$

with the initial condition that $I_0 = 1$. Maximizing (2.4) over S_0 and β provides a maximum likelihood estimator (MLE), $\widehat{S}_0$ and $\hat{\beta}$. Barring the biases discussed above, and noting that (2.2) gives an approximate value for R_0 of $S_0\beta$ for any reasonable values of β, we can calculate an approximate removal MLE of R_0 as $\widehat{R}_0 = \widehat{S}_0\hat{\beta}$. This, however, will be biased because of discretization. We bias correct according to

$$\widehat{R}_0^{\text{corrected}} \approx \widehat{R}_0 2(1 - \mathrm{e}^{-1}) - \mathrm{e}^{-1}. \tag{2.5}$$

Standard estimates of the standard errors for S_0 and β based on Fisher's information are not justified because the support of the chain binomial probability mass function depends on S_0 [19]. Therefore, we propose that inference for the individual parameters, and R_0 should be based on profile likelihoods [20] which rely on likelihood ratio statistics and do not make the same restriction on the support of the mass function [19]. To construct a profile likelihood for the composite parameter R_0, we re-parameterize the joint likelihood in terms of R_0 and S_0. For a fixed value of R_0, we maximize the likelihood for S_0. The resulting maxima, for a range of R_0 values, gives a profile likelihood for R_0 (Ben Bolker pers. comm.). We then construct confidence intervals using the approximate χ^2 distribution of the likelihood ratio statistic (e.g. [19]).

2.3. Measurement error

A simplifying assumption of the method we presented is that the time series of infected individuals is observed without error. A state space modeling approach [21,22], incorporating a binomial observation model coupled with the chain binomial likelihood may provide a more rigorous method in the presence of measurement error. Such an approach presents computational hurdles that may be solved using a Markov chain Monte Carlo algorithm; we are currently pursuing this line of enquiry (Ferrari, in preparation). Meanwhile, motivated by the historical development of removal estimators, we propose that a less rigorous solution to the problem of measurement error. The first removal estimator in ecology [14] proposed that abundance could be estimated by fitting a regression line to the decline in the number of removed animals as a function of cumulative number removed. A parallel argument can be made for estimation of R_0 if we consider the expectation of the chain binomial (2.2). Recall that $E[I_{t+1}|I_t, S_t, \beta] = S_t[1 - \exp(-\beta I_t)] \approx \beta S_t I_t$ and $S_t = S_0 - \sum I_t$. Substituting the latter equation into the former we get $E[I_{t+1}] \approx \beta(S_0 - \sum_{i=1}^{t} I_i)I_t$, which we can rewrite as $E[I_{t+1}]/I_t = \beta S_0 - \beta\sum_{i=1}^{t} I_i$. Denoting the epidemic ratio I_{t+1}/I_t by R_t and recalling that $R_0 = \beta S_0$, we obtain an approximate relationship

$$R_t = R_0 - \beta \sum_{i=1}^{t} I_i. \tag{2.6}$$

Note that in this relation we substitute the observed case count for its expectation. From this heuristic we suggest that R_0 can be estimated by the linear regression of the epidemic ratios on the cumulative removals due to infection. If reporting is incomplete (e.g. only fatalities are reported) such that the observed cases, O, are a proportion of the true number of cases, $O_t = pI_t$, then (2.6) becomes

$$\frac{pO_{t+1}}{pO_t} = R_0 - \beta \sum_{i=1}^{t} \frac{1}{p} O_i.$$

Ignoring the usually stochastic nature of under reporting, this is

$$\frac{O_{t+1}}{O_t} = R_0 - \frac{1}{p}\beta \sum_{i=1}^{t} O_i.$$

Thus, even in the presence of underreporting R_0 may be estimated using a regression method of incidence ratios on cumulative incidence. Because the stochastic variance inherent in a chain binomial infection process will change over the course of the epidemic, a weighted regression with weights proportional to the number observed infectious cases is appropriate (Appendix A). Again, the bias due to discretization may be corrected according to (2.5).

2.4. Evaluation and application

We evaluated the performance of the removal MLE on both simulated and real epidemics. We simulated epidemics using a stochastic epidemic birth-and-death model assuming exponentially distributed infectious periods [3,18]. For R_0 between 1.5 and 6 we simulated 5000 epidemics in populations of 1000 individuals. We assumed the pathogen to have a mean infectious period of 7-days. For each simulation we estimated R_0 at successive time steps to assess the value of the estimator at the completion of the epidemic and as a 'real-time' method. We additionally simulated observed time series with binomial reporting (probability of reporting = 1.0, 0.9, 0.7, and 0.5) and estimated R_0 using our approximate regression method.

We applied the removal estimator and the approximate regression method to data from four historical epidemics: the foot-and-mouth outbreak in Britain [23], swine fever in the Netherlands [24], and Ebola in Democratic Republic of Congo [25] and Uganda [26]. These data were chosen because all are directly transmitted, acute infections in relatively naïve populations that result in long term immunity or death. Further, each epidemic progressed quickly with respect to the host life history, so the birth and death dynamics of the host can be assumed inconsequential, which is consistent with the simple life history assumed in the model. The foot-and-mouth and swine fever data are weekly reports of the number of newly infected herds; R_0 thus reflects the inter-farm transmission rate. The Ebola data are weekly reports of the onset of symptoms. We also estimated R_0 for the recently emerged SARS epidemic in Hong Kong and Singapore (March 17–June 1, 2003) [27]. These data were daily reports of the onset of symptoms. The infectious periods for each pathogen were taken from literature reports and the case counts aggregated accordingly (Table 1). We compared the estimates with published values.

Table 1
Comparison of estimates of R_0 using the removal MLE with estimates from published literature for several diseases

Disease	Location	Year	Infectious period[a]	Removal MLE (95% CI)	Regression estimate	Estimates from literature
Ebola	DRC[b]	1995	14 days	3.65 (3.05 – 4.33)	3.07	1.8 (SD = 0.06)[c]
	Uganda[b]	2000	14 days	1.79 (1.52 – 2.30)	2.13	–
SARS	Hong Kong[b]	2003	5 days	2.27 (0.81 – 2.33)	1.50	2.2–3.6[d]; 2.2–3.7[e]
	Singapore[b]	2003	5 days	1.23 (0.95 – 1.49)	1.43	2.2–3.6[d]
Swine fever	Netherlands[b]	1997–1998	7 days	1.16 (0.83 – 1.54)	0.93	1.33[f]; 1.02–1.30[g]
FMD	Britain[b]	2000–2001	21 days	3.57 (3.04 – 3.93)	4.04	~3.9[h], 3.3[i]

Confidence intervals were calculated from the profile likelihood as described in the text. Literature estimates are given as point estimates or 95% confidence limits.
[a] Assumed.
[b] See references in text for data sources.
[c] [29].
[d] [28].
[e] [30].
[f] [24].
[g] [10].
[h] [31, Fig. 2c].
[i] [32].

3. Results

The bias corrected estimator is virtually unbiased for R_0 (Fig. 2). Further, it is unbiased for the size of the initial susceptible population (Fig. 2). When applied in real-time the removal MLE produced reasonable estimates of R_0 within ~4 infectious generations (Fig. 3). The estimation error was significantly improved over these 4 time steps, with marginal reduction thereafter (Fig. 3). Inspection of the likelihood profiles for β and S_0 show that these parameters are negatively correlated (Fig. 4). Nominal 95% profile confidence intervals for R_0 had true coverage rates of 67–80%. In application to data from real epidemics the removal MLE produced estimates of R_0 in consistent with those from the published literature (Table 1).

The approximate regression estimator for simulated data with measurement error had a slight negative bias (Fig. 5). Despite the small bias the mean value of the estimator was not appreciably affected, even under significant measurement error. In the presence of measurement error, estimates tended to be more variable than under perfect reporting. Estimates of R_0 from the regression method were generally similar to the MLE (Table 1).

4. Discussion

Discrete time models such as the chain binomial are appealing because they can match the timescale on which data are gathered. However, they necessarily oversimplify the inherent biology. It

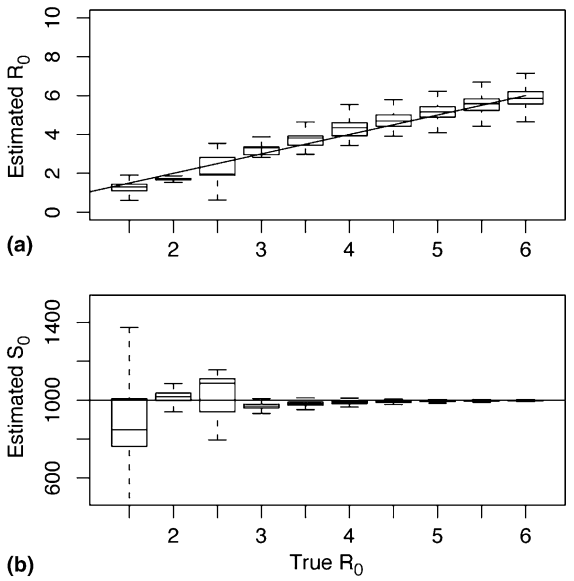

Fig. 2. Performance of the removal MLE's for simulated epidemics as a function of true R_0; (a) point estimates for R_0, (b) point estimates for S_0. Each box plot in (a) and (b) represents a summary of estimates from 5000 simulations. The central bar indicates the median estimate, the box represents inter-quartile range, and the dashed lines indicate the most extreme observation that is no more than 1.5 times the inter-quartile range from the box. The solid line indicates the true value of each parameter.

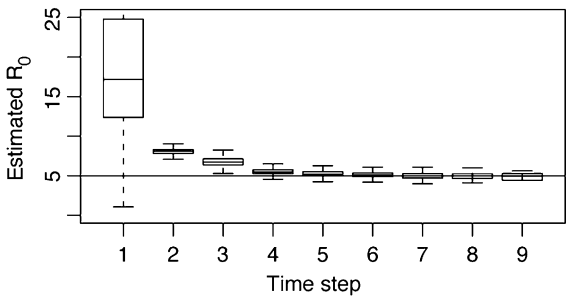

Fig. 3. 'Real time' estimates of R_0 using the removal MLE for simulation with true $R_0 = 5$. Each box plot represents a summary of 5000 simulations and is constructed as in Fig. 1.

is therefore important to consider the utility of using such discrete time models for conducting inference on continuous time processes [6]. Using a discrete time epidemic model with infectious generations that approximate the epidemic birth-and-death model results in bias because it ignores rapid secondary chains of transmission and late primary infections from individuals that remain infectious longer than the average infectious period. By analyzing the nature of these sources of bias we derived a correction that allows implementation of a chain binomial likelihood to derive estimates of R_0 and the initial susceptible population size from data on case counts. The

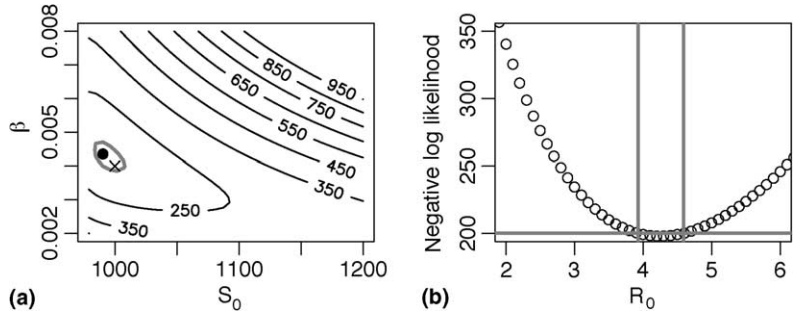

Fig. 4. Likelihoods for β, S_0, and R_0. (a) Likelihood contours for β and S_0 for one simulated epidemic with $S_0 = 1000$ and $\beta = 0.004$. The heavy grey contours the 95% confidence region based on the likelihood ratio test. The X indicates the true parameters and the dot indicates the MLE's. (b) The profile likelihood for R_0 for the same simulated epidemic. The horizontal line gives the critical value of the likelihood ratio test and the vertical lines give the corresponding lower and upper 95% confidence limits.

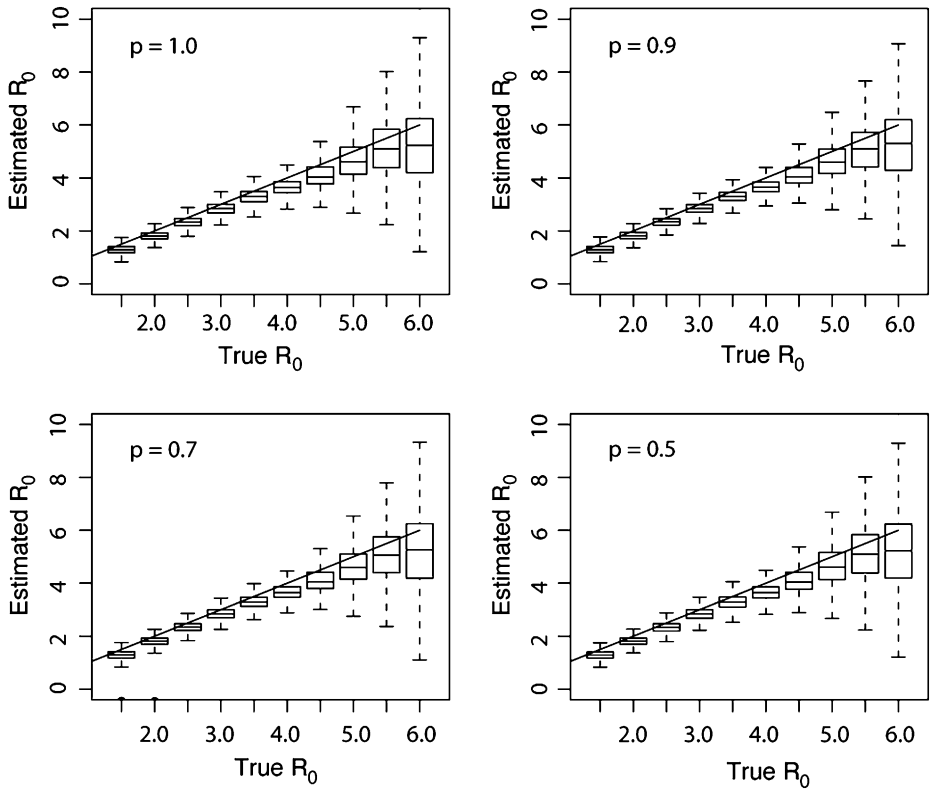

Fig. 5. Performance of the regression estimator for simulated epidemics with binomial observation probability of 1.0, 0.9, 0.7, and 0.5. Each box plot represents a summary of 5000 simulations and is constructed as in Fig. 1. The solid line indicates the true value of R_0.

benefit of the method is that it is computationally simple (an *R*-library for calculation is available from the authors upon request), require minimal data and assumptions of historical mixing rates, and can provide continual upgrading of the estimates of both R_0 and the estimated initial susceptible population size during the course of an outbreak.

The bias correction can be interpreted in two ways. First, it may be seen as correcting the time step in the discrete model to balance the underestimation of transmission due to late infection and the overestimation due to higher order transmission (the shaded regions in Fig. 1). In practice, however, the correction simply adjusts the transmission rate within the conventional time step of $1/\gamma$. It is interesting to note that binning the observations according to a time step of length $2(1 - e^{-1})/\gamma$ also corrects the negative bias in the removal estimator, but results in a positive bias of approximately e^{-1} as predicted from (2.1); we have confirmed through with simulations, but have not presented results here.

The corrected R_0 estimator is not strictly unbiased; however, the magnitude of the bias is so small (<0.5) as to be of little practical relevance. The removal MLE is remarkably consistent when applied to simulated data (Fig. 2), but full statistical inference should be conducted with caution. The confidence limits do reflect the uncertainty in R_0, given the observed number of epidemic generations. However, coverage rates perform below the nominal rate and confidence limits do not reflect the additional uncertainty due to the random duration of epidemics.

The explicit form of the bias correction is specific to the assumption that the infectious period is exponentially distributed. Keeling and Grenfell [17] have proposed that a model with gamma distributed infectious periods may better reproduce epidemic dynamics. Discretization of a model with gamma distributed infectious periods will still result in biases due to the assumption of generation separation. An analogous bias correction may be constructed for the gamma distributed model following arguments similar to those proposed here.

The removal estimator is sensitive to the time period in which the data are binned. As the infectious period for a given pathogen may change from historical values, it will be important to verify the length of the infectious period from clinical records for novel epidemics. In such cases where the infectious period is unknown, it may be wise to first use estimators based on the exponential approximation to the initial rate of increase of infectious cases (e.g. [28]). However, as the epidemic is limited by the number of susceptible hosts, the exponential approximation performs poorly when applied to the full epidemic trajectory.

In general, our estimates of R_0 from observed data are consistent with reported values. The estimates for SARS were on the low end of reported confidence limits for Hong Kong, and considerably lower for Singapore. In light of the negative bias due to binomial reporting error in Fig. 5, it seems possible that the results for the SARS outbreak were affected by the imperfect reporting. In contrast, the Swine fever and FMD estimates, for which the case count data have lower errors, were very close to previously reported estimates.

The many methods available to estimate R_0 are a testament to its central importance in epidemiology. Through the development of the removal estimator we have addressed two recurrent issues in the problem of estimating R_0: uncertainty in the true susceptible population size, and the bias introduced by fitting an inherently continuous time process with discretely sampled data. The latter is an issue in applications well beyond epidemiology [6]. By considering the expectation of the continuous time epidemic process over discrete intervals, we can make use of a discrete time observational model to conduct inference on R_0.

Acknowledgments

We would like to thank Ben Bolker for his help with the profile likelihoods and three anonymous reviewers for constructive comments and bringing to our attention an algebraic error that greatly improved the performance of the estimator. This research was supported in part by the John E. Fogarty Center of the National Institutes of Health (ONB).

Appendix A. Regression weights

By examining the random process that generates the epidemic it is possible to derive a relationship between the weights and the independent variable. We will consider the continuous SI model because it allows an appeal to results for general Poisson processes. From the delta method, a first-order approximation to the variance of R_t conditional on $I_t, \ldots, I_1$ is:

$$\mathrm{Var}[R_t] = \mathrm{Var}\left[\frac{I_{t+1}}{I_t}\right] \approx \frac{\mathrm{Var}[I_{t+1}]}{E[I_t]^2} + \frac{E[I_{t+1}]^2}{E[I_t]^4}\mathrm{Var}[I_t]. \tag{A.1}$$

Note that because of the exponential distribution of waiting times to infection and death, the expected number of infected individuals in interval t, given the course of the epidemic prior to time t, $I_t|\underline{I}_{t-1}, \ldots, I_1$ is approximately a Poisson random variable. Because there may be multiple infections and/or deaths over the interval, we write the Poisson parameter as $\beta\overline{S}_t\overline{I}_t = \lambda_t$, where the bar indicates the average number of susceptible and infectious individuals over the interval. Inserting λ_t into (A.1), we get

$$\mathrm{Var}[R_t] \approx \frac{\lambda_{t+1}}{\lambda_t^2} + \frac{\lambda_{t+1}^2}{\lambda_t^4}\lambda_t \approx \frac{\lambda_{t+1}}{\lambda_t}\left(\frac{1}{\lambda_t}\right) + \left(\frac{\lambda_{t+1}}{\lambda_t}\right)^2\left(\frac{1}{\lambda_t}\right). \tag{A.2}$$

Noting that λ_{t+1}/λ_t is the first-order Taylor series approximation to $E[R_t]$, (A.2) should behave qualitatively similar to

$$\frac{E[R_t] + E[R_t]^2}{E[I_t]}. \tag{A.3}$$

Thus, from (A.3) the variance in R_t conditional on the cumulative number of cases should be proportional to a polynomial of the expected value and the inverse of the number of cases in interval t. If we note that the $E[I_t]$ should be large relative to $E[R_t]$, we see that the behavior of (A.3) should be dominated by the denominator. Thus, weights proportional to the number of cases in each observation interval should help stabilize the variance in the residuals.

References

[1] K. Dietz, The estimation of the basic reproduction number for infectious diseases, Stat. Meth. Med. Res. 2 (1993) 23.
[2] J.A.P. Heesterbeek, A brief history of R-0 and a recipe for its calculation, Acta Biotheor. 50 (2002) 189.
[3] R.M. Anderson, R.M. May, Infectious Diseases of Humans: Dynamics and Control, Oxford University, Oxford, 1991.
[4] W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics, Proc. Roy. Soc. Lond. A 115 (1927) 700.

[5] G.J. Gibson, E. Renshaw, Estimating parameters in stochastic compartmental models using Markov chain methods, IMA J. Math. Appl. Med. 15 (1998) 19.
[6] K. Glass, Y. Xia, B.T. Grenfell, Interpreting time-series analyses for continuous-time biological models-measles as a case study, J. Theor. Biol. 223 (2003) 19.
[7] O.N. Bjørnstad, B.F. Finkenstadt, B.T. Grenfell, Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model, Ecol. Monogr. 72 (2002) 169.
[8] M. Begon, S.M. Hazel, D. Baxby, K. Bown, R. Cavanagh, J. Chantrey, T. Jones, M. Bennett, Transmission dynamics of a zoonotic pathogen within and between wildlife host species, P. Roy. Soc. Lond. B Bio. 266 (1999) 1939.
[9] O.N. Bjørnstad, B. Finkenstädt, B.T. Grenfell, Endemic and epidemic dynamics of measles. I. Estimating transmission rates and their scaling using a time series SIR model, Ecol. Monogr. 72 (2002) 185.
[10] R. Meester, J. de Koning, M.C.M. de Jong, O. Diekmann, Modeling and real-time prediction of classical swine fever epidemics, Biometrics 58 (2002) 178.
[11] J. Swinton, J. Harwood, B.T. Grenfell, C.A. Gilligan, Persistence thresholds for phocine distemper virus infection in harbour seal Phoca vitulina metapopulations, J. Anim. Ecol. 67 (1998) 54.
[12] R. Gani, S. Leach, Transmission potential of smallpox in contemporary populations, Nature 414 (2001) 748.
[13] D.J.D. Earn, P. Rohani, B.M. Bolker, B.T. Grenfell, A simple model for complex dynamical transitions in epidemics, Science 287 (2000) 667.
[14] P.H. Leslie, D.H.S. Davis, An attempt to determine the absolute number of rats on a given area, J. Anim. Ecol. 8 (1939) 94.
[15] P.A.P. Moran, A mathematical theory of animal trapping, Biometrika 38 (1951) 307.
[16] C.M. Harding, A.W. Heathwood, R.G. Hunt, L.Q. Read, The estimation of animal population size by the removal method, Appl. Stat. 33 (1984) 196.
[17] M.J. Keeling, B.T. Grenfell, Understanding the persistence of measles: reconciling theory, simulation and observation, P. Roy. Soc. Lond. B 269 (2002) 335.
[18] N.T.J. Bailey, The Mathematical Theory of Epidemics, Griffin, London, 1957.
[19] P. McCullagh, J.A. Nelder, Generalized Linear Models, 2nd ed., in: D.R. Cox et al. (Eds.), Monographs on Statistics and Applied Probability, Vol. 37, Chapman and Hall, London, 1989.
[20] S.A. Murphy, A.W. Van der Vaart, On profile likelihood, J. Am. Stat. Assoc. 95 (2000) 449.
[21] P. de Valpine, A. Hastings, Fitting population models with process noise and observation error, Ecology 72 (2002) 57.
[22] C. Calder, M. Lavine, P. Muller, J.S. Clark, Incorporating multiple sources of stochasticity into dynamic population models, Ecology 84 (2003) 1395.
[23] M.J. Keeling, M.E.J. Woolhouse, D.J. Shaw, L. Matthews, M. Chase-Topping, D.T. Haydon, S.J. Cornell, J. Kappey, J. Wilesmith, B.T. Grenfell, Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape, Science 294 (2001) 813.
[24] A. Stegeman, A.R.W. Elbers, J. Smak, M.C.M. de Jong, Quantification of the transmission of classical swine fever virus between herds during the 1997–1998 epidemic in The Netherlands, Prev. Vet. Med. 42 (1999) 219.
[25] A.S. Khan, F.K. Tshioko, D.L. Heymann, B. Le Guenno, P. Nabeth, B. Kerstiens, Y. Fleerackers, P.H. Kilmarx, G.R. Rodier, O. Nkuku, P.E. Rollin, A. Sanchez, S.R. Zaki, R. Swanepoel, O. Tomori, S.T. Nichol, C.J. Peters, J.J. Muyembe-Tamfum, T.G. Ksiazek, The reemergence of Ebola hemorrhagic fever, Democratic Republic of the Congo, 1995, J. Infect. Dis. 179 (1999) S76.
[26] Outbreak of Ebola haemorrhagic fever, Uganda, August 2000–January 2001, in MMWR, 2001, p. 73.
[27] WHO, Disease Outbreak News. 2003, http://www.who.int/csr/don/archive/disease/severe_acute_respiratory_syndrome/en/.
[28] M. Lipsitch, T. Cohen, J.M. Robins, S. Ma, L. James, G. Gopalakrishna, S.K. Chew, C.C. Tan, M.H. Samore, D. Fisman, M. Murray, Transmission dynamics and control of severe acute respiratory syndrome, Sciencexpress 300 (2003) 1966.
[29] G. Chowell, N.W. Hengartner, C. Castillo-Chavez, P.W. Fenimore, J.M. Hyman, The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda, J. Theor. Biol. 299 (2004) 119.

[30] S. Riley, C. Fraser, C.A. Donnelly, A.C. Ghani, L.J. Abu-Raddad, A.J. Hedley, G.M. Leung, L. Ho, T. Lam, T.Q. Thach, P. Chau, K. Chan, S. Lo, P. Leung, T. Tsang, W. Ho, K. KLee, E.M. Lau, N.M. Ferguson, R.M. Anderson, Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions, Science 300 (2003) 1961.

[31] N.M. Ferguson, C.A. Donnelly, R.M. Anderson, Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain (vol. 413, p. 542, 2001), Nature 414 (2001) 329.

[32] M. Woolhouse, M. Chase-Topping, D. Haydon, J. Friar, L. Matthews, G. Hughes, D. Shaw, J. Wilesmith, A. Donaldson, S. Cornell, M. Keeling, B. Grenfell, Epidemiology-Foot-and-mouth disease under control in the UK, Nature 411 (2001) 258.