Sensitivity of Model-Based Epidemiological Parameter Estimation to Model Assumptions

A.L. Lloyd

Abstract Estimation of epidemiological parameters from disease outbreak data often proceeds by fitting a mathematical model to the data set. The resulting parameter estimates are subject to uncertainty that arises from errors (noise) in the data; standard statistical techniques can be used to estimate the magnitude of this uncertainty. The estimates are also dependent on the structure of the model used in the fitting process and so any uncertainty regarding this structure leads to additional uncertainty in the parameter estimates. We argue that if we lack detailed knowledge of the biology of the transmission process, parameter estimation should be accompanied by a structural sensitivity analysis, in addition to the standard statistical uncertainty analysis. Here we focus on the estimation of the basic reproductive number from the initial growth rate of an outbreak as this is a setting in which parameter estimation can be surprisingly sensitive to details of the time course of infection.

1 Introduction

Estimation of epidemiological parameters, such as the average duration of infectiousness or the basic reproductive number of an infection, is often an important task when examining disease outbreak data [see, for example, 12, 19, 26]. In many instances, one or more parameters of interest cannot be estimated directly from the available data, so an indirect approach is adopted in which a mathematical model of the transmission process is formulated and is fitted to the data. The resulting parameter estimates will have uncertainty due to noise in the data but they will also depend on the form chosen for the model. Any uncertainties in our knowledge of the biology underlying the transmission process lead to uncertainties in the parameter estimates over and above those that arise from noise in the data.

Standard statistical approaches (see Chapters 1, 5, 7, 10 and 11 of this book) can be used to quantify the uncertainty in parameter estimates that arises from noise in

A.L. Lloyd (✉)
Biomathematics Graduate Program and Department of Mathematics, North Carolina State University, Raleigh, NC 27695, USA
e-mail: alun_lloyd@ncsu.edu

G. Chowell et al. (eds.), *Mathematical and Statistical Estimation Approaches in Epidemiology*, DOI 10.1007/978-90-481-2313-1_6,

the data, but these are not designed to provide insight into the sensitivity of the estimates to the *structure* of the model. In this chapter, we demonstrate that uncertainty due to model structure can, in some instances, dwarf noise-related uncertainty by discussing an estimation problem in which details of the description of the biology of the transmission process can have an important impact. This argues that when there is incomplete knowledge of the biology of the infection, structural sensitivity analysis should accompany statistical uncertainty analysis when model-based approaches are used to interpret epidemiological data.

In this chapter, we illustrate the potential importance of model assumptions by examining the model-based estimation of the basic reproductive number using data obtained from the initial stages of a disease outbreak. We review studies [21, 25, 27, 31, 33, 34] that illustrate that such estimates can be highly sensitive to the assumptions made concerning the natural history of the infection, particularly regarding the timing of secondary transmission events. These results are of major significance in the setting of emerging infectious disease outbreaks, when a rapid quantification of the basic reproductive number is highly desirable to guide control efforts, but when information on the transmission cycle may be scarce. Importantly, the work shows that the use of simple models can greatly underestimate the value of the basic reproductive number, providing overly optimistic predictions for how effective control measures have to be in order to curtail the spread of the disease.

2 The Basic Reproductive Number and Its Estimation Using the Simple SIR Model

The basic reproductive number, R_0, is defined as the average number of secondary infections caused by a typical infective individual in an otherwise entirely susceptible population [see, for example, 11]. In the simplest settings, its value can be calculated as the product of the rate at which such an individual gives rise to infections and the duration of their infectious period. In turn, the infection rate is a product of the rate at which an infective meets susceptible individuals, *i.e.* the contact rate, and the per-contact probability of transmission.

Direct estimation of the basic reproductive number could be undertaken if secondary infections of individual infectives could be quantified. Unfortunately, the most commonly available type of data—aggregated incidence data—does not reveal transmission chains in sufficient detail to identify the source of secondary cases. More detailed data, such as contact tracing data, can elucidate chains of transmission, but is rarely complete enough to allow direct calculation of R_0. In the absence of complete contact tracing data, statistical techniques have been suggested for the estimation of R_0 via reconstruction of transmission chains [13, 32].

The basic reproductive number could also be directly estimated if both the contact rate and transmission probability were known. Again, direct estimation of these quantities is typically difficult. Transmission probabilities can be estimated using certain types of epidemiological data, obtained, for instance, from observation

of transmission within families, or other transmission experiments. Such data, however, are often unavailable during the early stages of a disease outbreak.

An alternative approach involves fitting a mathematical model to outbreak data, obtaining estimates for the parameters of the model, allowing R_0 to be calculated. The simplest model that can be used for this purpose is the standard deterministic compartmental SIR model [see, for example, 11]. Individuals are assumed to either be susceptible, infectious or removed, with the numbers of each being written as S, I, and R, respectively. Susceptible individuals acquire infection through contacts with infectious individuals, and the simplest form of the model assumes that new infections arise at rate $\beta SI/N$. Here N is the population size and β is the transmission parameter, which is given by the product of the contact rate and the transmission probability. Recovery of infectives is assumed to occur at a constant rate γ, corresponding to an average duration of infection of $1/\gamma$, and leads to permanent immunity. Throughout this chapter we shall denote the average duration of infectiousness by D_{I} and assume permanent immunity following infection. We shall also ignore demographic processes (births and deaths), which is a good approximation if the disease outbreak is short-lived and the infection is non-fatal. Ignoring demography leads to the population size N being constant. The model can be written as the following set of differential equations

$$dS/dt = -\beta SI/N \tag{1}$$

$$dI/dt = \beta SI/N - \gamma I \tag{2}$$

$$dR/dt = \gamma I. \tag{3}$$

During the early stages of an outbreak with a novel pathogen, almost the entire population will be susceptible, and, since $S \approx N$, the transmission rate equals βI. The transmission parameter β is the rate at which each infective gives rise to secondary infections and so the basic reproductive number can be written as $R_0 = \beta D_{\mathrm{I}} = \beta/\gamma$. During this initial period, the changing prevalence of infection can, to a very good approximation, be described by the single linear equation $dI/dt = \gamma(R_0 - 1)I$. (We remark that the $S = N$ assumption corresponds to linearizing the model about its infection free equilibrium.) In other words, provided that R_0 is greater than one, which we shall assume to be the case throughout this chapter, prevalence initially increases exponentially with growth rate

$$r = \gamma(R_0 - 1). \tag{4}$$

The incidence of infection is given by $\beta SI/N$ and so, during the early stages of an outbreak, prevalence and incidence are proportional in the SIR setting, so this equation also describes the rate at which incidence grows.

Equation (4) provides a relationship, $R_0 = 1 + rD_{\mathrm{I}}$, between R_0 and quantities that can typically be measured (the initial growth rate of the epidemic and the average duration of infection), and as a result has provided one of the most straightforward ways to estimate R_0.

3 More Complex Compartmental Models

The SIR model of Section 2 employs a very simple, but quite unrealistic, description of the time course of infection. The infectious period is assumed to start immediately upon infection, and the constant recovery rate corresponds to infectious periods being exponentially distributed across the population. In reality, there is a delay—the latent period—between acquisition of infection and the start of infectiousness: an individual typically receives a small dose of an infectious agent and several rounds of replication have to occur within the infected person before they become infectious. The exponential distribution has a much larger variance than infectious period distributions observed in the real world: it predicts that a large number of individuals recover very soon after infection and that a sizeable number of individuals have infectious periods that are much longer than the average. In reality, infectious periods are much more closely centered about their mean [2, 4].

3.1 Inclusion of Latency

A latent period can easily be incorporated within the compartmental framework with the addition of an exposed class (E) of infected but not yet infectious individuals. Assuming that movement between the E and I classes occurs at a constant per-capita rate of σ, we get the standard SEIR model

$$dS/dt = -\beta SI/N \tag{5}$$

$$dE/dt = \beta SI/N - \sigma E \tag{6}$$

$$dI/dt = \sigma E - \gamma I. \tag{7}$$

The latent period here is exponentially distributed with average duration $1/\sigma$. Throughout this chapter, we shall refer to the average duration of latency as D_{E}. The inclusion of the exposed class does not affect the algebraic expression for the basic reproductive number: we again have $R_0 = \beta D_{\mathrm{I}} = \beta/\gamma$.

The initial behavior of an outbreak can be well described by a linear model, consisting of Equations (6) and (7) with the transmission term being replaced by βI. Provided that R_0 is greater than one, and following an initial transient, prevalence increases exponentially at rate r given by the dominant eigenvalue, the value of which is the larger of the roots of the quadratic

$$r^2 + (\sigma + \gamma)r - \sigma\gamma(R_0 - 1) = 0. \tag{8}$$

Provided that both the average durations of latency and infectiousness are known, Equation (8) can be rearranged to give R_0 in terms of the initial growth rate, giving $R_0 = (1 + rD_{\mathrm{E}})(1 + rD_{\mathrm{I}})$ [19, 25]. As for the SIR model, the incidence of infection will also grow at this rate.

Intuitively, it is clear that latency will decrease the initial growth rate of an outbreak: latency delays the start of an individual's infectious period, making their secondary infections occur later than they would if infectiousness were to begin immediately. This can be confirmed mathematically by comparing the roots of Equations (4) and (8). The constant coefficient of the quadratic in Equation (8) is equal to the product of its roots, and, because R_0 is greater than one, its value is negative. The quadratic therefore has one negative and one positive root. The value of the quadratic is negative when $r = 0$ and positive when $r = \gamma(R_0 - 1)$ and so its positive root lies in the interval $(0, \gamma(R_0 - 1))$. The growth rate for the SEIR model is lower than it was for the SIR model.

This effect is illustrated in Fig. 1, where the prevalence of infection seen in an SIR model outbreak (solid curve) is compared to that seen in the corresponding SEIR model (dotted curve). In both cases, the average infectious period is 5 days and R_0 is 5, and for the SEIR model there is a two day average duration of latency. At the initial time the entire population of one million people is taken to be susceptible except for a single infective individual. The latent period has a dramatic effect on the initial growth, and indeed on the entire timecourse, of the outbreak. (We remark that the non-exponential change in prevalence seen at the start of the outbreak in the SEIR model is the transient behavior mentioned above and arises from the second, negative, value for r in Equation 8).

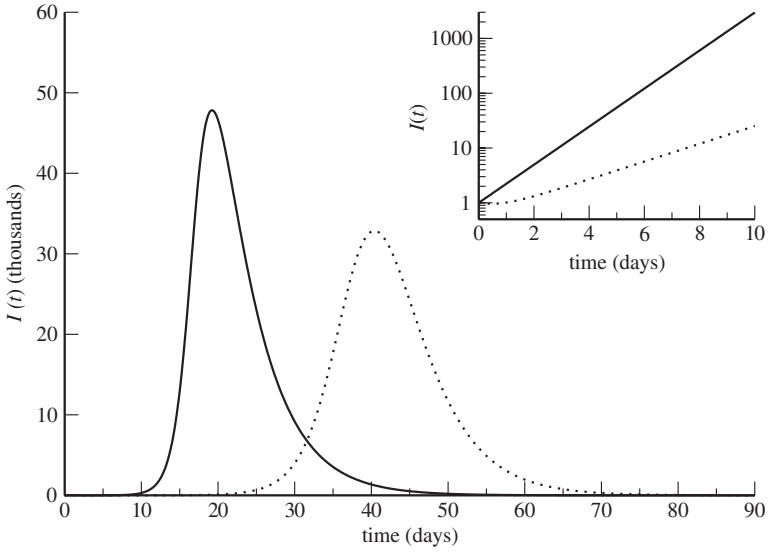

Fig. 1 Impact of latency on a disease outbreak, comparing SIR and SEIR models. *Solid curve*: no latent period (SIR model). *Dotted curve*: exponentially-distributed latent period (SEIR model). The inset (plotted on log-linear axes) focuses on the initial behavior of the two outbreaks, when the epidemics are well-described by linear models, and shows the slower initial growth rate of the SEIR outbreak. The average infectious period is taken to be $D_I = 5$ days, R_0 is 5, and, for the SEIR model, the latent period has an average duration of $D_E = 2$ days. At the initial time, the entire population of $N = 10^6$ is susceptible to the infection, except for one individual who is taken to have just become infectious

3.2 More General Compartmental Models: Gamma Distributed Latent and Infectious Periods

An individual's chance of recovery is not constant over time: typically, the recovery rate increases over time. In terms of a mathematical model, this leads to the complication that the times at which different individuals became infected must be tracked. In contrast, the constant rate assumptions of the SIR and SEIR models are mathematically convenient as their rates of recovery and loss of latency can be written just in terms of the current numbers of infectives and exposeds.

A mathematical trick [1, 10, 17] allows the inclusion of non-exponential distributions within the compartmental framework. The infective class can be subdivided into n stages, arranged in series. Newly infected individuals enter the first infective stage, pass through each in turn, and recover upon leaving the nth stage. It is assumed that progression between stages occurs at constant per-capita rate, leading to an exponential waiting time in each stage and allowing movement between stages to be described by a linear system of differential equations. The stage approach allows the modeler to retain the convenience of the differential equation approach, albeit at the cost of an increased number of state variables and hence dimensionality of the model.

In the simplest setting, the average waiting time (or equivalently the departure rate) in each stage is assumed to be equal: the overall infectious period is then described by the sum of n independent exponential distributions, *i.e.* infectious periods are gamma distributed [10, 17] with shape parameter n, as illustrated in Fig. 2. To allow comparison between models with different numbers of stages, the average duration of infectiousness is often held fixed, meaning that the departure rate is equal to $n\gamma$ for each stage. In a similar way, a non-exponential latent period can be described by the use of m exposed stages. A general form of the SEIR model, which we dub the $\mathrm{SE}_m\mathrm{I}_n\mathrm{R}$ model, is then given by

$$
\begin{aligned}
dS/dt &= -\beta SI/N && (9)\\
dE_1/dt &= \beta SI/N - m\sigma E_1 && (10)\\
dE_2/dt &= m\sigma E_1 - m\sigma E_2 && (11)\\
&\vdots\\
dE_m/dt &= m\sigma E_{m-1} - m\sigma E_m && (12)\\
dI_1/dt &= m\sigma E_m - n\gamma I_1 && (13)\\
dI_2/dt &= n\gamma I_1 - n\gamma I_2 && (14)\\
&\vdots\\
dI_n/dt &= n\gamma I_{n-1} - n\gamma I_n. && (15)
\end{aligned}
$$

Here $I = I_1 + I_2 + \cdots + I_n$ is the total number of infectives. We remark that the $\mathrm{SE}_m\mathrm{I}_n\mathrm{R}$ model has just two extra parameters compared to the SEIR model, and

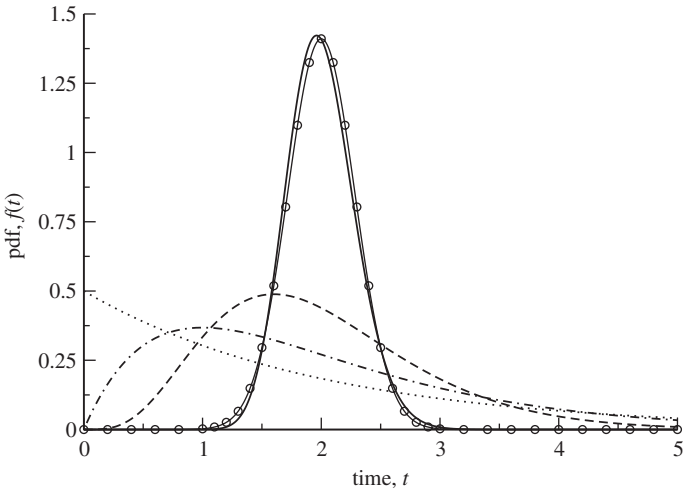

Fig. 2 Gamma distributed infectious periods. The graph illustrates the probability density function (pdf) of gamma distributions with $n = 1$ (*dotted curve*), $n = 2$ (*dot-dashed curve*), $n = 5$ (*dashed curve*) or $n = 50$ (*solid curve*) stages. In each case, the average duration of infection D_I is two days. The variances of the gamma distributions are given by ${D_I}^2/n$. As discussed in the text, for large n, the gamma distribution approaches a normal distribution: for comparison, the curve with circles depicts a normal distribution with mean and variance equal to those of the $n = 50$ gamma distribution

that if $n = m = 1$, the model reduces to the standard SEIR model. If either m or n is large, then, by the Central Limit Theorem, the relevant gamma distribution becomes approximately normal (see Fig. 2). In the limit $m \to \infty$ or $n \to \infty$, either the exposed or infectious period distribution becomes of fixed duration.

More general distributions can be described using variations of the stage device, for instance by having unequal movement rates or more complicated arrangements of stages, such as stages in parallel as well as in series. Furthermore, the infectiousness of different stages can be allowed to vary, giving a transmission term of the form $\Sigma_i \beta_i S I_i / N$. In some instances, the stages are identified with biologically-defined different stages of an infection, as, for example, in the case of a number of models for HIV [23]. But we emphasize that, in general, the stages are a mathematical device and need not have any biological interpretation.

Linearization of the model (9)–(15) gives the growth rate (of both prevalence and incidence) as the dominant root of the equation

$$\gamma R_0 \left\{ 1 - \left(1 + \frac{r D_I}{n} \right)^{-n} \right\} = r \left(1 + \frac{r D_E}{m} \right)^{m} . \tag{16}$$

This equation is equivalent to Equation (9) of Anderson and Watson [1], but Lloyd [21] employed this version in which R_0 appears explicitly. Here R_0 is

again equal to $\beta D_I = \beta/\gamma$. We remark that in the limit of m approaching infinity, the term $(1 + r D_E/m)^m$ approaches $\exp(r D_E)$, and as n approaches infinity, the term $(1 + r D_I/n)^{-n}$ approaches $\exp(-r D_I)$. Fixed duration latent and/or infectious periods lead to the appearance of exponential terms and a transcendental equation for r in terms of R_0.

Decreasing the variance of the latent period distribution, *i.e.*, increasing m while keeping σ constant, reduces the initial growth rate. This effect can be seen in Fig. 3a, comparing the initial growth of the SEIR model (solid curve) to those seen in the corresponding SE_5IR and $SE_{50}IR$ models (dotted and dashed curves, respectively). The impact of adding extra stages decreases as the value of m increases: a larger change is seen in the growth rate when m is changed from 1 to 5 than is seen when m is increased from 5 to 50.

Reduction in the variance of the infectious period distribution, *i.e.* increasing n while keeping γ fixed increases the initial growth rate of an outbreak (Fig. 3b).

4 A General Formulation

The stage approach provides a simple way to incorporate gamma-distributed waiting times within the compartmental framework. More general descriptions of the time-course of infection can be accounted for using a number of different approaches, including partial differential equations, delay differential equations, integral equations and integro-differential equations [5–7, 11, 14–16, 18]. As an example, the following integro-differential equation can be used to describe the number of susceptibles

$$\frac{dS}{dt} = -\delta(t) - \frac{S(t)}{N}\int_0^\infty \left(-\frac{dS}{dt}\right)\Big|_{(t-\tau)} \mathcal{A}(\tau)\, d\tau. \tag{17}$$

Here $\mathcal{A}(\tau)$ is the infectivity kernel, *i.e.*, the expected infectiousness of an individual τ time units after infection. (In an entirely susceptible population, this would be the rate at which such an individual gives rise to secondary infections.) The delta function depicts the infection of a single individual at the initial time. The integral that appears in this equation depicts the force of infection experienced by susceptibles at time t, while the incidence of infection, which we shall write as $X(t)$, is equal to $-dS/dt$. Notice that, as with all the models we consider in this chapter, we ignore replenishment of the susceptible population.

A number of variants of this formulation appear in the literature. In some instances, the contact rate, c, appears explicitly in Equation (17), with the infectivity kernel being written as $c\mathcal{A}(\tau)$. Several authors write the infectivity kernel as the product $\mathcal{A}(\tau) = A(\tau)\beta(\tau)$, where $A(\tau)$ is the probability that an individual is infectious at time τ and $\beta(\tau)$ is the expected infectiousness of an individual who is infectious at that time. In this formulation, if the duration of the latent period is given by the random variable T_E and the duration of the infectious period by the random variable T_I, then $A(\tau) = \Pr(T_E \le \tau < T_E + T_I)$.

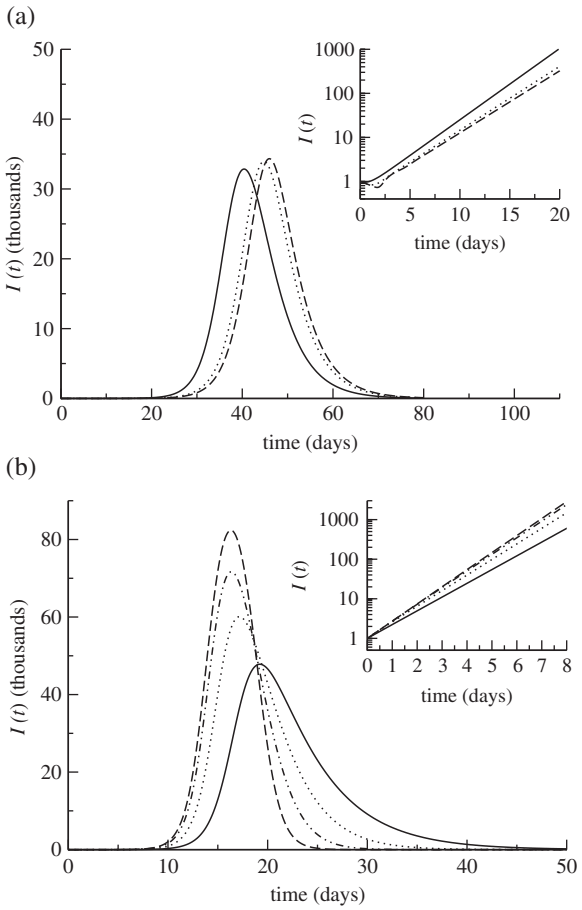

Fig. 3 Impact of distributional assumptions on epidemic behavior seen in SIR and SEIR-type models. Panel (**a**) shows the impact of the latent period distribution in SEIR-type models. In each case the infectious period is exponentially distributed. *Solid curve*: exponentially-distributed latent period (SEIR model). *Dotted and dashed curves*: gamma-distributed latent period, with $m = 5$ and $m = 50$ exposed stages (SE_5IR and $SE_{50}IR$ models), respectively. Panel (**b**) depicts the effect of various descriptions of the infectious period in SIR-type models (no latent period). *Solid curve*: exponentially distributed infectious period (SIR model). *Dotted curve*: gamma-distributed infectious period with $n = 2$ stages. *Dot-dashed curve*: gamma-distributed infectious period with $n = 5$ stages. *Dashed curve*: gamma-distributed infectious period with $n = 50$ stages. For both panels (**a**) and (**b**), the average infectious period D_I is taken to be 5 days, R_0 is 5, and, where relevant, the latent period has an average duration of $D_E = 2$ days. At the initial time, the entire population of $N = 10^6$ is susceptible to the infection, except for one individual who is taken to have just become infectious. The insets focus on the early behavior, including the phase when the behavior can be well approximated by a linear model

The compartmental models described in previous sections can be recast in terms of an infectivity kernel. For the SIR model, the constant level of infectivity over an exponentially distributed infectious period of average duration $1/\gamma$ gives

$$\mathcal{A}(\tau) = \beta e^{-\gamma\tau}, \tag{18}$$

and for the corresponding SEIR model, with average duration of latency D_{E} equal to $1/\sigma$, we have

$$\mathcal{A}(\tau) = \begin{cases} \beta \dfrac{\sigma}{\gamma - \sigma} \left(e^{-\sigma\tau} - e^{-\gamma\tau} \right) & \text{if } \sigma \neq \gamma \\ \beta\gamma\tau e^{-\gamma\tau} & \text{if } \sigma = \gamma. \end{cases} \tag{19}$$

The basic reproductive number for this model is given by

$$R_0 = \int_0^\infty \mathcal{A}(\tau)\, d\tau. \tag{20}$$

During the early stages of an outbreak, $S(t) \approx N$, and use of the approximation $S(t) = N$ gives the following linear integral equation for the incidence

$$X(t) = \delta(t) + \int_0^\infty X(t-\tau)\mathcal{A}(\tau)\, d\tau. \tag{21}$$

Substitution of an exponentially growing form for the incidence, $X(t) = X(0)e^{rt}$, for $t \geq 0$ gives the equation

$$1 = \int_0^\infty e^{-r\tau}\mathcal{A}(\tau)\, d\tau, \tag{22}$$

which can be solved for the rate r at which incidence grows. This equation is the familiar Euler-Lotka formula from demographic theory [see, for example, 30].

The integral that appears in Equation (22) is the Laplace transform of the infectivity kernel. Yan [34] derived a relationship between R_0 and r for a general class of infectivity kernels for which the random variables describing the latent and infectious periods, T_E and T_I, are independent and under the assumption that secondary infections arise at constant rate β over the duration of the infectious period. Assuming that the Laplace transforms of the distributions of both T_E and T_I exist, and writing them as $\mathcal{L}_E(r)$ and $\mathcal{L}_I(r)$, Yan obtained the following general result

$$R_0 = \frac{D_{\mathrm{I}}}{\mathcal{L}_E(r)\mathcal{L}_I^*(r)}. \tag{23}$$

Here, $\mathcal{L}_I^*(r) = (1 - \mathcal{L}_I(r))/r$.

All of the relationships between R_0 and r obtained from compartmental models in the earlier sections of this chapter can be obtained as special cases of this result. In particular, the earlier Equation (16) can be seen as a special case of Yan's general result and holds for general gamma distributed latent and infectious periods (*i.e.,* with any positive shape parameters—not just integers).

5 Comparing R_0 Estimates Obtained Using Different Models

The relationships between R_0 and r described in previous sections, obtained from a number of SIR and SEIR-type models, are collected together in Table 1. It is immediately clear that use of the SIR-based formula provides a lower estimate of R_0 than would be obtained using the SEIR-based formula [21, 25, 33]. Ignoring an infection's latent period leads to underestimates of R_0, with the underestimate being more serious for faster growth rates or longer durations of latency (*e.g.*, compare Tables 2, 3 and 4).

The origin of the underestimate is clear from the analysis and simulations presented earlier: given the same values for the transmission parameter and the average

Table 1 Relationships between the initial growth rate r and the basic reproductive number R_0 obtained from various models

Model	Formula
SIR	$R_0 = 1 + r D_\mathrm{I}$
$\mathrm{SI}_n\mathrm{R}$	$R_0 = \dfrac{r D_\mathrm{I}}{1 - (1 + r D_\mathrm{I}/n)^{-n}}$
$\mathrm{SI}_\infty\mathrm{R}$	$R_0 = \dfrac{r D_\mathrm{I}}{1 - e^{-r D_\mathrm{I}}}$
SEIR	$R_0 = (1 + r D_\mathrm{I})(1 + r D_\mathrm{E})$
$\mathrm{SEI}_\infty\mathrm{R}$	$R_0 = \dfrac{r D_\mathrm{I}(1 + r D_\mathrm{E})}{1 - e^{-r D_\mathrm{I}}}$
$\mathrm{SE}_m\mathrm{IR}$	$R_0 = (1 + r D_\mathrm{I})(1 + r D_\mathrm{E}/m)^m$
$\mathrm{SE}_m\mathrm{I}_n\mathrm{R}$	$R_0 = \dfrac{r D_\mathrm{I}(1 + r D_\mathrm{E}/m)^m}{1 - (1 + r D_\mathrm{I}/n)^{-n}}$
$\mathrm{SE}_\infty\mathrm{IR}$	$R_0 = (1 + r D_\mathrm{I}) e^{r D_\mathrm{E}}$
$\mathrm{SE}_\infty\mathrm{I}_\infty\mathrm{R}$	$R_0 = \dfrac{r D_\mathrm{I} e^{r D_\mathrm{E}}}{1 - e^{-r D_\mathrm{I}}}$

Table 2 R_0 and p_c estimates obtained using various models when $r = 0.04\ \mathrm{day}^{-1}$, $D_\mathrm{E} = 3$ days and $D_\mathrm{I} = 8$ days. These parameters were chosen to be similar to those employed in [8] to describe SARS

Model	R_0 estimate	Control fraction p_c
SIR	1.32	0.242
$\mathrm{SI}_5\mathrm{R}$	1.20	0.167
$\mathrm{SI}_\infty\mathrm{R}$	1.17	0.144
SEIR	1.48	0.324
$\mathrm{SEI}_\infty\mathrm{R}$	1.31	0.236
$\mathrm{SE}_5\mathrm{IR}$	1.49	0.327
$\mathrm{SE}_5\mathrm{I}_5\mathrm{R}$	1.35	0.260
$\mathrm{SE}_\infty\mathrm{IR}$	1.49	0.328
$\mathrm{SE}_\infty\mathrm{I}_\infty\mathrm{R}$	1.32	0.241

Table 3 Impact of faster growth rate on R_0 estimates. Here, $r = 0.12$ day^{-1}, while $D_E = 3$ days and $D_I = 8$ days take the same values as in the previous table

Model	R_0 estimate	Control fraction p_c
SIR	1.96	0.490
SI_5R	1.64	0.391
$SI_\infty R$	1.56	0.357
SEIR	2.67	0.625
$SEI_\infty R$	2.12	0.527
SE_5IR	2.77	0.640
SE_5I_5R	2.33	0.570
$SE_\infty IR$	2.81	0.644
$SE_\infty I_\infty R$	2.23	0.552

duration of infectiousness, the SEIR model would predict a lower growth rate than the corresponding SIR model. Consequently, in order to achieve the same growth rate in this forward problem setting, a higher transmission parameter, and hence R_0, must be used in the SEIR model than in the corresponding SIR model.

Larger estimates of the basic reproductive number are obtained when non-exponential descriptions of the latent period distribution are used. Because the function $(1 + a/x)^x$ is monotonic increasing for $a > 0$, increasing the number of latent stages m (*i.e.,* reducing the variance of the latent period distribution) increases the estimate [21, 27, 33, 34], with the estimate that employs a fixed duration of latency (*i.e.,* $m \to \infty$) providing an upper bound. On the other hand, lower estimates of the basic reproductive number are obtained when the number of infectious stages n is increased [21, 27, 33, 34], with the estimate obtained using a fixed duration of infectiousness (*i.e.,* $n \to \infty$) being a lower bound.

Because estimates of R_0 are often used to determine the severity of measures needed to bring an outbreak under control, assumptions made about the timecourse of infection can have important public health consequences [21, 33]. If the aim of control is to bring the basic reproductive number below one, the transmissibility of the infection must be reduced by a factor of $p_c = 1 - 1/R_0$. Here, we call p_c the control fraction. (In the context of mass vaccination, p_c is called the critical

Table 4 Impact of longer average duration of latency on R_0 estimates. Here, $D_E = 5$ days, while $r = 0.12$ day^{-1} and $D_I = 8$ days, as in the previous table

Model	R_0 estimate	Control fraction p_c
SIR	1.96	0.490
SI_nR	1.64	0.391
$SI_\infty R$	1.56	0.357
SEIR	3.14	0.681
$SEI_\infty R$	2.49	0.598
SE_mIR	3.45	0.710
SE_mI_nR	2.89	0.655
$SE_\infty IR$	3.57	0.720
$SE_\infty I_\infty R$	2.83	0.647

vaccination fraction.) As we have seen, use of SIR models underestimates the basic reproductive number and hence leads to lower estimates of p_c compared to those obtained using SEIR models. This could be a serious problem as it leads to an overly optimistic prediction of the strength of control needed to curtail an outbreak: a control measure that, on the basis of the incorrect model, is predicted to succeed could, instead, be doomed to failure [21].

We first illustrate these results by providing a few examples using a growth rate and durations of latency and infectiousness based roughly on a SARS modeling study of Chowell et al. [8]. (The study of Chowell et al. accounted for treatment and isolation, using a more complex model than those employed here, so direct comparisons cannot be made.) Table 2 shows estimates obtained for an observed growth rate of 0.04 per day, assuming a three day average duration of latency and an eight day average duration of infectiousness. We imagine that details of how latent and infectious periods are distributed about their means are unknown, and so present estimates based on a number of models. This provides an indication of the degree of uncertainty that arises from incomplete knowledge of these distributions. (Here we ignore the additional complication that the estimate of r would also have some uncertainty.) For this example, comparing the SIR and SEIR-based estimates, we see that ignoring the latent period leads to R_0 being underestimated by about 10%. This translates into a 25% underestimate of the control fraction.

Interestingly, for this set of parameters, the distribution of the latent period (provided that one is used in the first place) has little impact on the estimates, while the infectious period distribution has a more noticeable effect. In this case the latter effect is sufficiently large to offset the differences introduced by ignoring a latent period: the estimates obtained using the SIR and the $SE_\infty I_\infty R$ models are almost identical.

For a more rapidly growing outbreak, in which $r = 0.12$ day^{-1} is three times larger than its previous value, the SIR model underestimates R_0 by a larger amount, roughly 25%, compared to the SEIR estimate (Table 3). This corresponds to a 22% underestimate of the control fraction. We remark that while this underestimate is slightly smaller in percentage terms than that seen under the previous set of parameters, it is larger in absolute terms. Also, given that the required level of control is higher, the increase in effectiveness needed to go from the SIR-based estimate of p_c to the SEIR-based estimate may be much more difficult to achieve. For this set of parameters, the form of the latent period distribution has a more noticeable impact.

If the infection is both more rapidly growing and has a longer duration of latency the underestimate of R_0 is more severe. In the example of Table 4, in which the average duration of latency has been raised from 3 to 5 days, use of the SIR model underestimates R_0 by roughly 38% of the SEIR-based estimate. The two estimates of the control fraction are 0.490 (SIR) and 0.681 (SEIR).

Figure 4 shows how estimates of R_0 obtained using the SE_mI_nR model depend in turn on each of the quantities D_E, D_I, m, n and r for a situation corresponding to Table 4. We take $D_E = 5$ days, $D_I = 8$ days, $r = 0.12$ day $^{-1}$, $m = 5$ and $n = 5$ as a baseline, and vary just one of these at a time. As discussed above, the estimate of R_0 increases with m, D_E, D_I and r, but decreases with n. We also see that the

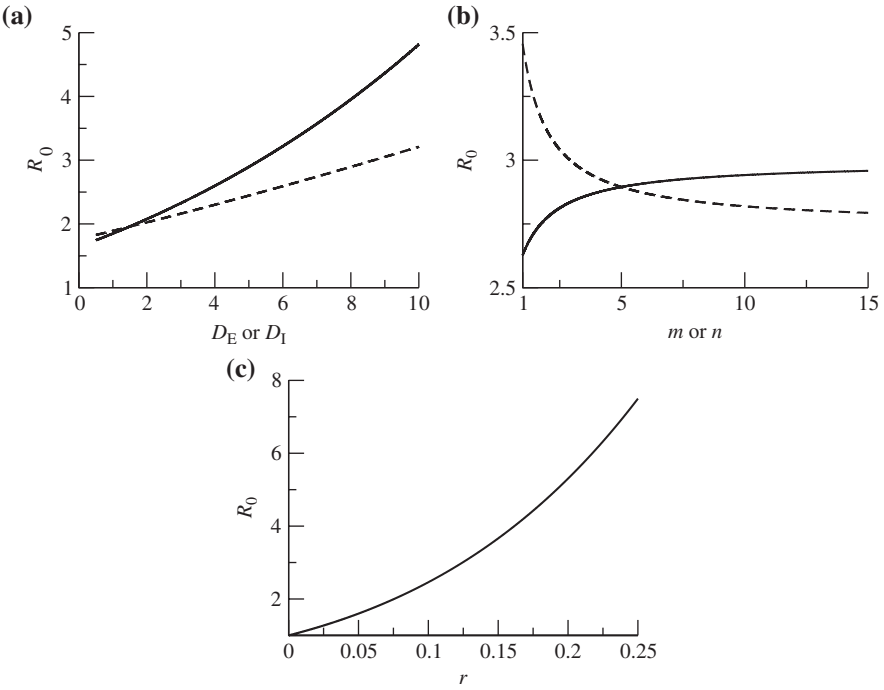

Fig. 4 Sensitivity of the R_0 estimate to variations in single parameter values or the initial growth rate. One of five quantities is varied in turn: Panel (**a**) D_E (*solid curve*) or D_I (*dashed curve*); Panel (**b**) m (*solid curve*) or n (*dashed curve*); Panel (**c**) r. The four other values are taken from the baseline set of $D_E = 5$ days, $D_I = 8$ days, $m = 5$, $n = 5$, and $r = 0.12$ day $^{-1}$

sensitivity of the estimate varies with these parameters, for instance, the R_0 estimate is less sensitive to m for larger values of m.

A dramatic example of the potential for the underestimation of R_0 was provided by Nowak et al. [25] in a within-host setting that can be modeled using virus dynamics models that are directly analogous to the epidemiological models considered here. The initial growth rate of simian immunodeficiency virus (SIV) in one particular animal in an experimental infection study was found to be 2.2 day^{-1} and the average duration of infectiousness (of SIV infected cells) was 1.35 days. Use of the SIR model gave an estimate of $R_0 = 4.0$, the SEIR model, assuming a one day latent period (*i.e.,* the delay between a cell becoming infected and becoming infectious), gave $R_0 = 13$, while the $SE_\infty IR$ model gave $R_0 = 36$. We remark on the large impact of the distribution of the latent period in this instance. In terms of control fractions, the three models predict values of 0.75 (SIR), 0.92 (SEIR) and 0.97 ($SE_\infty IR$). While there may be hope in achieving a 75% reduction in transmissibility, a 92 or 97% reduction would be much harder to achieve. In this case, use of the SIR model gives a wildly optimistic picture of the effectiveness required of a control measure.

If the latent period in this within-host example was instead assumed to be 0.5 days, the effect would be reduced, with the SEIR and $SE_\infty IR$-based estimates of R_0 falling to 8.3 and 12, respectively. These values are still considerably larger than the SIR-based estimate of 4.0, and the estimate is still highly sensitive to the distribution of the latent period. The corresponding control fractions are 0.88 and 0.92, respectively.

6 Sensitivity Analysis

The numerical examples presented above give an idea of the dependency of R_0 estimates on parameter values in particular settings, but a more systematic exploration can be achieved using sensitivity analysis. It is straightforward to calculate the partial derivatives of the estimated value of R_0 as provided by the SE_mI_nR model (*i.e.*, using Equation 16) with respect to the parameters D_E, D_I, m, n, and the initial growth rate r. The elasticity E_x, which approximates the fractional change in the R_0 estimate that results from a unit fractional change in parameter x (while keeping all other parameters constant), is given by $E_x = (x/R_0) \cdot \partial R_0/\partial x$. The elasticities for the quantities of interest are

$$E_{D_E} = \frac{r D_E}{1 + r D_E/m} \tag{24}$$

$$E_{D_I} = 1 - \frac{r D_I}{(1 + r D_I/n)\left((1 + r D_I/n)^n - 1\right)} \tag{25}$$

$$E_m = m \ln\left(1 + \frac{r D_E}{m}\right) - \frac{r D_E}{1 + r D_E/m} \tag{26}$$

$$E_n = \frac{1}{(1 + r D_I/n)^n - 1}\left(-n \ln(1 + r D_I/n) + \frac{r D_I}{1 + r D_I/n}\right) \tag{27}$$

$$E_r = 1 + \frac{r D_E}{1 + r D_E/m} - \frac{r D_I}{(1 + r D_I/n)\left((1 + r D_I/n)^n - 1\right)}. \tag{28}$$

We remark that if the curves that appear in Fig. 4 were replotted on log-log axes, these elasticities would describe the slopes of these new graphs.

The signs of the elasticities confirm the earlier discussion of how the estimate of R_0 varies as parameter values are changed. Clearly E_{D_E} is positive, meaning that increases in D_E lead to larger estimates of R_0. E_{D_I} is also seen to be positive, since the second term in Equation (25) is smaller than one for positive values of the parameters r, D_I and n: increases in D_I again lead to larger estimates of R_0. E_m is seen to be positive when m, r and D_E are positive because the function $m \ln(1 + x/m) - x/(1 + x/m)$ is monotonic increasing in x and takes the value 0 when x equals zero. A similar argument shows that E_n is negative. Finally, E_r equals the sum of E_{D_E} and E_{D_I} and so is positive, and is greater than either E_{D_E} or E_{D_I}.

A little algebra shows that E_{D_E} is an increasing function of r, D_E or m, *i.e.*, the elasticity of the R_0 estimate with respect to D_E increases with these parameters. E_{D_I} is an increasing function of r or D_I, but parameter sets can be found for which it is non-monotonic as n changes. E_m increases with r or D_E, but can be non-monotonic as m changes. E_n can be non-monotonic as r or D_I changes. As before, E_r inherits the properties of E_{D_E} and E_{D_I}, and so increases with D_E, D_I, m or r, but need not be a monotonic function of n.

For the parameters of Table 2 and when $m = n = 1$ we find that the elasticities are given by $E_{D_\mathrm{E}} = 0.107$, $E_{D_\mathrm{I}} = 0.242$, $E_m = 0.006$, $E_n = -0.110$, and $E_r = 0.350$. If, instead, we take $m = n = 5$ the elasticities are $E_{D_\mathrm{E}} = 0.117$, $E_{D_\mathrm{I}} = 0.173$, $E_m = 0.001$, $E_n = -0.026$, and $E_r = 0.290$. In both cases, the R_0 estimate is more sensitive to changes in D_I than to changes in D_E, and here we see that sensitivities to m and n are of smaller magnitude for larger values of m and n, while the sensitivity to D_E increases with increasing m and the sensitivity to D_I decreases with increasing n.

Whether the estimate is more sensitive to changes in D_E or D_I (or to m or n) depends on the values of the parameters. For example, if the parameters of Table 4 are taken, and $m = n = 1$ is assumed, the elasticities are $E_{D_\mathrm{E}} = 0.375$, $E_{D_\mathrm{I}} = 0.490$, $E_m = 0.095$, $E_n = -0.191$, and $E_r = 0.865$. If, instead, we assumed $m = n = 5$, the estimate of R_0 would be more sensitive to D_E than to D_I ($E_{D_\mathrm{E}} = 0.536$ and $E_{D_\mathrm{I}} = 0.427$), and if we took $m = 1$ and $n = 5$, the estimate would be more sensitive to m than to n ($E_m = 0.095$ and $E_n = -0.052$).

One important question that the elasticities discussed in this section do not address is the impact of neglecting the latent period entirely. Having said this, they are useful in understanding how uncertainties in the average duration of the latent or infectious period, the dispersions of these distributions, as described by m or n, or the initial growth rate impact the estimation of R_0 in the $\mathrm{SE}_m\mathrm{I}_n\mathrm{R}$ model framework.

7 Discussion

The importance of non-exponential infectious periods and time-varying infectiousness has long been appreciated for chronic infections, such as HIV, for which a constant recovery rate assumption is clearly untenable [5–7, 16, 23, 24]. Even in the setting of models of acute infections, there is a surprisingly long history of the use of more complex models: Kermack and McKendrick's groundbreaking paper of 1927 [18] contains an integral equation formulation along the lines of Equation (17), and Bailey [3] used the stage approach and the resulting $\mathrm{SE}_m\mathrm{I}_n\mathrm{R}$ model. The importance of distributional assumptions has typically been viewed in terms of their impact on the behavior of a model for a given set of parameters (*i.e.*, the forward problem): effects such as the slower growth of epidemics for infections with latency have long been appreciated.

The impact of distributional assumptions on the inverse problem, (*i.e.*, the estimation of parameters given the observed behavior), however, appears to have only recently become fully appreciated. Nowak et al. [25] showed that the SIR-based

estimates of the within-host basic reproductive number of SIV (simian immunodeficiency virus) severely underestimated R_0 when compared to estimates obtained using more realistic SEIR models. Little et al. [20] carried out a similar analysis in the setting of HIV infection. Much of the theory and results discussed in this chapter were laid out by Lloyd [21], in the setting of within-host infections, although, because of the obvious correspondence between within-host and between-host models, the application to estimation in the epidemiological setting was highlighted [see also the discussion of 22]. Wearing et al. [33] further illustrated these results in an epidemiological setting, and broadened consideration to include estimation of R_0 based on data from the entire outbreak, as discussed below. A complementary approach was taken by Wallinga and Lipsitch [31] and Roberts and Heesterbeek [27], who examined the relationship between R_0 and r in terms of the generation interval of the infection (*i.e.*, the time between an individual becoming infected and the secondary infections that they cause). Both of these studies considered gamma distributed latent and infectious periods, including the exponential and fixed duration cases, giving equivalent results to those discussed here. Additional families of distributions were also considered, including trapezoidal infectivity kernels [27] and normally distributed generation intervals [31]. Yan [34] provided a comprehensive analysis that encompassed and unified most of these earlier studies, deriving general results in terms of Laplace transforms of the latent and infectious period distributions.

The results presented here demonstrate that estimates of the basic reproductive number obtained from the initial growth rate of a disease outbreak can be sensitive to the details of the timing of secondary infection events (*i.e.*, to the distribution of infectious and latent periods). Such details, while clearly important, are often difficult to obtain. Data that identifies when an individual was exposed to infection and when their secondary transmissions occurred, such as family-based transmission studies or contact tracing data—even if incomplete—can be highly informative in this regard [2, 4, 12, 13]. It is important to realize that models are often framed in terms of transmission status, *e.g.* whether an individual is infectious, while data may reflect disease status, *e.g.* whether an individual is symptomatic or not. This distinction is important in the interpretation of the most commonly available distributional data, namely the incubation period distribution [28], because the incubation period of an infection may not, and often does not, correspond to its latent period [see, for example, 29].

In this chapter we only considered the estimation of R_0 from initial growth data, but similar results are obtained if models are instead fitted to data obtained over the entirety of an outbreak [33]. This makes sense given the observation that distributional assumptions affect not only the initial growth rate but the whole time course of an outbreak in the forward problem (see Figs. 1 and 3). Whole-outbreak data is considerably more informative than initial growth data, for instance Wearing et al. [33] used a least-squares approach to estimate β, D_E and D_I as well as the shape parameters m and n of the gamma distributions describing latency and infectiousness. Initial growth rate data, on the other hand, does not even allow β and γ to be independently estimated. Capaldi et al. (manuscript in preparation) examine the types of data that allow for the estimation of different parameters in more detail.

Wearing et al. [33] also make the important observation that different estimates of R_0 can, in some instances, be obtained if initial data is used rather than data from an entire outbreak [see also 9].

If detailed information on the distribution of latent and infectious periods is absent, caution should be taken in basing an estimate of R_0 by fitting a single model. The use of a number of models can provide bounds on the estimate, giving an indication of the uncertainty arising from our incomplete knowledge of the transmission process. Any model-based uncertainty is in addition to that which arises from noise in the data—an issue that we have not discussed in this chapter—and so the most informative uncertainty estimate would account for both sources of error. (Sensitivity calculations, such as those discussed above, can be informative in this regard.) In some instances, however, model-based uncertainty may place a much greater limit on our ability to estimate parameters. As in the within-host example of Nowak et al. [25], this uncertainty can be so large as to render the estimates almost uninformative—but at least the deficiency is exposed by the approach advocated here.

References

1. Anderson DA, Watson RK (1980) On the spread of disease with gamma distributed latent and infectious periods. *Biometrika* 67:191–198.
2. Bailey NTJ (1954) A statistical method for estimating the periods of incubation and infection of an infectious disease. *Nature* 174:139–140.
3. Bailey NTJ (1964) Some stochastic models for small epidemics in large populations. *Appl. Stat.* 13:9–19.
4. Bailey NTJ (1975) *The Mathematical Theory of Infectious Diseases*. Griffin, London.
5. Blythe SP, Anderson RM (1988) Variable infectiousness in HIV transmission models. *IMA J. Math. Appl. Med. Biol.* 5:181–200.
6. Blythe SP, Anderson RM (1988) Distributed incubation and infectious periods in models of the transmission dynamics of the human immunodeficiency virus (HIV). *IMA J. Math. Appl. Med. Biol.* 5:1–19.
7. Castillo-Chavez C, Cooke K, Huang W, Levin SA (1989) On the role of long incubation periods in the dynamics of HIV/AIDS, Part 1: Single population models. *J. Math. Biol.* 27: 373–398.
8. Chowell G, Fenimore PW, Castillo-Garsow MA, Castillo-Chavez C (2003) SARS outbreaks in Ontario, Hong Kong and Singapore: the role of diagnosis and isolation as a control mechanism. *J. Theor. Biol.* 224:1–8.
9. Chowell G, Nishiura H, Bettencourt LMA (2007) Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *J. R. Soc. Interface* 4: 155–166.
10. Cox DR, Miller HD (1965) *The Theory of Stochastic Processes*. Methuen, London.
11. Diekmann O, Heesterbeek JAP (2000) *Mathematical Epidemiology of Infectious Diseases*. John Wiley & Son, Chichester.
12. Donnelly CA, Ghani AC, Leung GM, Hedley AJ, Fraser C, Riley S, Abu-Raddad LJ, Ho L-M, Thach T-Q, Chau P, Chan K-P, Lam T-H, Tse L-Y, Tsang T, Liu S-H, Kong JHB, Lau EMC, Ferguson NM, Anderson RM (2003) Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet* 361:1761–1766.

13. Haydon DT, Chase-Topping M, Shaw DJ, Matthews L, Friar JK, Wilesmith J, Woolhouse MEJ (2003) The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. *Proc. R. Soc. Lond. B* 269:121–127.
14. Hethcote HW, Tudor DW (1980) Integral equation models for endemic infectious diseases. *J. Math. Biol.* 9:37–47.
15. Hoppensteadt F (1974) An age dependent epidemic model. *J. Franklin Inst.* 297:325–333.
16. Hyman JM, Stanley EA (1988) Using mathematical models to understand the AIDS epidemic. *Math. Biosci.* 90:415–473.
17. Jensen A (1948) An elucidation of Erlang's statistical works through the theory of stochastic processes. In E. Brockmeyer, H. L. Halstrøm, and A. Jensen, editors, *The Life and Works of A. K. Erlang*, pages 23–100. The Copenhagen Telephone Company, Copenhagen.
18. Kermack WO, McKendrick AG (1927) A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A* 115:700–721.
19. Lipsitch M, Cohen T, Cooper B, Robins JM, Ma S, James L, Gopalakrishna G, Chew SK, Tan CC, Samore MH, Fisman D, Murray M (2003) Transmission dynamics and control of severe acute respiratory syndrome. *Science* 300:1966–1970.
20. Little SJ, McLean AR, Spina CA, Richman DD, Havlir DV (1999) Viral dynamics of acute HIV-1 infection. *J. Exp. Med.* 190:841–850.
21. Lloyd AL (2001) The dependence of viral parameter estimates on the assumed viral life cycle: Limitations of studies of viral load data. *Proc. R. Soc. Lond. B* 268:847–854.
22. Lloyd AL (2001) Destabilization of epidemic models with the inclusion of realistic distributions of infections periods. *Proc. R. Soc. Lond. B* 268:985–993.
23. Longini IM, Clark WS, Byers RH, Ward JW, Darrow WW, Lemp GF, Hethcote HW (1989) Statistical analysis of the stages of HIV infection using a Markov model. *Stat. Med.* 8: 831–843.
24. Malice M-P, Kryscio RJ (1989) On the role of variable incubation periods in simple epidemic models. *IMA J. Math. Appl. Med. Biol.* 6:233–242.
25. Nowak MA, Lloyd AL, Vasquez GM, Wiltrout TA, Wahl LM, Bischofberger N, Williams J, Kinter A, Fauci AS, Hirsch VM, Lifson JD (1997) Viral dynamics of primary viremia and antiretroviral therapy in simian immunodeficiency virus infection. *J. Virol.* 71:7518–25.
26. Riley S, Fraser C, Donnelly CA, Ghani AC, Abu-Raddad LJ, Hedley AJ, Leung GM, Ho L-M, Lam T-H, Thach T-Q, Chau P, Chan K-P, Lo S-V, Leung P-Y,Tsang T, Ho W, Lee K-H, Lau EMC, Ferguson NM, Anderson RM (2003) Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions. *Science* 300:1961–1966.
27. Roberts MG, Heesterbeek JAP (2007) Model-consistent estimation of the basic reproduction number from the incidence of an emerging infection. *J. Math. Biol.* 55:803–816.
28. Sartwell PE (1950) The distribution of incubation periods of infectious disease. *Am. J. Hygiene* 51:310–318.
29. Sartwell PE (1966) The incubation period and the dynamics of infectious disease. *Am. J. Epidemiol.* 83:204–216.
30. Smith D, Keyfitz N (1977) *Mathematical Demography. Selected Papers*. Biomathematics, Volume 6. Springer-Verlag, Berlin.
31. Wallinga J, Lipsitch M (2007) How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. R. Soc. Lond. B* 274:599–604.
32. Wallinga J, Teunis P (2004) Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am. J. Epidemiol.* 160:509–516.
33. Wearing HJ, Rohani P, Keeling M (2005) Appropriate models for the management of infectious diseases. *PLoS Medicine* 2(7):e174.
34. Yan P (2008) Separate roles of the latent and infectious periods in shaping the relation between the basic reproduction number and the intrinsic growth rate of infectious disease outbreaks. *J. Theor. Biol.* 251:238–252.