Model-consistent estimation of the basic reproduction number from the incidence of an emerging infection

M. G. Roberts · J. A. P. Heesterbeek

Abstract We investigate the merit of deriving an estimate of the basic reproduction number $\mathcal{R}_0$ early in an outbreak of an (emerging) infection from estimates of the incidence and generation interval only. We compare such estimates of $\mathcal{R}_0$ with estimates incorporating additional model assumptions, and determine the circumstances under which the different estimates are consistent. We show that one has to be careful when using observed exponential growth rates to derive an estimate of $\mathcal{R}_0$, and we quantify the discrepancies that arise.

Keywords Basic reproduction number · Emerging infectious disease · Epidemics

Mathematics Subject Classification (2000) 92B05

1 Introduction

The basic reproduction number $\mathcal{R}_0$ of an infectious agent is defined as the expected number of secondary cases caused by one typical infected individual in a population consisting of susceptibles only [3,6,7]. When an outbreak has started and the approximation that the population is fully susceptible no longer holds, one generally

M. G. Roberts
Centre for Mathematical Biology, Institute of Information and Mathematical Sciences,
Massey University, Private Bag 102 904, North Shore Mail Centre,
Auckland, New Zealand
e-mail: m.g.roberts@massey.ac.nz

J. A. P. Heesterbeek (✉)
Faculty of Veterinary Medicine, University of Utrecht, Yalelaan 7,
3584 CL, Utrecht, The Netherlands
e-mail: j.a.p.heesterbeek@vet.uu.nl

refers to the effective reproduction number $\mathcal{R}$. The value of $\mathcal{R}_0$ is, as a rule, different for different infectious agents and depends among other things on the characteristics of the population that the agent invades. Given this, it is not immediate that one can adopt previously determined values or size ranges for a new outbreak, unless many of the complicated characteristics of, for example, population composition and contact structure are comparable. For various reasons one can be interested in the value of $\mathcal{R}_0$ or $\mathcal{R}$ early in an outbreak and during the outbreak. Notably, under a homogeneous mixing assumption, the values give insight into the extent of the control problem and a means of calculating how much control effort is needed.

In recent years several new methods for estimating $\mathcal{R}_0$ from outbreak data have been published, either as a general tool or for specific applications [4,9,11,17,21,24]. Some, like Wallinga and Teunis [24] do not need much data, but require knowledge of the generation time distribution. Most of these methods, however, are data-hungry: they either need contact information, use the whole outbreak time series (so are effectively retrospective measures), or increase in accuracy as the time series becomes longer. While several methods are promising many problems remain, and for various reasons. For example, we do not observe infections we observe detections, i.e. individuals (people, animals, farms, plants) exhibiting symptoms. There is then a possibly unknown incubation period distribution that convolutes the infection process into the observed process. Detections are not necessarily in the order of infection. Moreover, $\mathcal{R}_0$ is a generation-based concept [12], but generations are not observed—a daily number of new detections is observed from possibly mixed generations. Early in the outbreak stochastic influences play a large role. Also, heterogeneity between individual infectivity and susceptibility and in contact pattern may cause the distribution of which $\mathcal{R}_0$ is the mean to be highly skewed (e.g. [16]). To make matters worse, we almost never have data from an uncontrolled situation - some control measures, effective or not, often operate from the moment of detection of the index case.

Despite advances, but in light of the problems encountered, many publications in which $\mathcal{R}_0$ is estimated from outbreak data still depend on cumulative incidence and generation interval only (see for example [5,8,14,18,26]). As a rule, the cumulative incidence in outbreaks of an infectious disease is observed to initially grow approximately exponentially with time (and hence the incidence grows exponentially too). A frequently used approach is to fit an exponential function to the (cumulative) incidence and to use the approximate relationship $\mathcal{R}_0 \approx e^{rT_G}$ to estimate $\mathcal{R}_0$, where r is the exponential growth rate and T_G is the observed mean generation interval of the epidemic. For rT_G small the further approximation $\mathcal{R}_0 \approx 1 + rT_G$ is sometimes used. Many of the problems mentioned above apply to these estimates. For example, the 'real' value of r is not observed, not only because of control measures in operation but also due to the stochasticity in the early phase. In addition, the definition of the generation interval is not always consistently used and the method presupposes that the population is homogeneously mixing. Still, the method is easy and intuitive and one can wonder in which circumstances it would be a 'good' approximation, and how large discrepancies can be when these circumstances are not met. In this paper we investigate these questions.

2 Model-consistent estimation of $\mathcal{R}_0$

We now derive estimates of $\mathcal{R}_0$ based on specific models and compare these with the previously mentioned approximations, which we denote $\mathcal{R}_0^+ = e^{rT_G}$ and $\mathcal{R}_0^- = 1 + rT_G$. We need to emphasise that $\mathcal{R}_0$ is independent of timescale, whereas T_G has dimension time and r has dimension time^{-1}. We also need to emphasise that we do not know r or T_G, we assume that they have been estimated from data in some way, for example by estimating the doubling time of the incidence, D, and writing $r = \log(2)/D$. We do not write $\hat{r}$ or $\hat{T}_G$ for these estimates, as this would result in far too many hats in one paper.

Assume for simplicity that the population is mixing homogeneously. The incidence of an emerging infection may be calculated from

$$i(t) = \delta(t) + \frac{S(t)}{N} \int_0^\infty A(\tau) i(t - \tau) \, d\tau \tag{1}$$

where $\delta(t)$, a unit spike, is the incidence of infection at time zero, the kernel $A(\tau)$ is the expected infectivity of an infected as a function of τ, the time since exposure to infection [1,6,19]. The number in the population susceptible at time t is

$$S(t) = N - \int_0^t i(u) \, du$$

For an emerging infection we assume the entire population to be susceptible at time zero. If this is not the case, we take N to be the size of the susceptible population prior to infection. As a first step towards developing a model, we specify the general form of the kernel $A(\tau)$. We write $A(\tau) = \mathcal{R}_0 f(\tau)$, where $\mathcal{R}_0$ is the basic reproduction number that we wish to estimate, and $f(\tau)$ is the infectivity kernel, which is also the probability distribution of the generation interval.

For an emerging infection we have little information about f. We may have observations of the latent period (the time from exposure to infection to becoming infectious, T_E); the incubation period (the time from exposure to infection to the onset of symptoms) which we may in some cases assume to equal T_E; or the infectious period T_I. Given these we may wish to impose a particular form on the kernel, and use our limited knowledge to estimate parameter values for the distribution. These estimates may be revised as more information becomes available.

One quantity of interest is the mean generation interval of the epidemic, which is taken here to be the mean time from an individual's exposure to infection to exposing others to infection (see [10], for an insightful exposition). We refer not to the time to the first occurrence of a secondary infection, but to the average time to all secondary infections. Alternatively, and equivalently, it can be defined as the expected duration of the primary infection at the time that a secondary infection occurs (see [22]). The mean generation interval may be determined from the formula

$$T_G = \int_0^\infty t f(t) \, dt \tag{2}$$

Given a probability distribution for the generation interval, $f(t)$, and an estimated initial rate of exponential increase for the epidemic, r, we approximate the initial stages of the epidemic by $i(t) = e^{rt}$ with $S(t) \simeq N$. Equation (1) then leads to a model-consistent estimate of the basic reproduction number via the formula

$$\mathcal{R}_0 \int_0^\infty e^{-rt} f(t) \, dt = 1 \tag{3}$$

(see [6]). If $f(t)$ were a delta function, then Eqs. (2, 3) would lead to the estimate $\mathcal{R}_0 = \mathcal{R}_0^+$. For the *SIR* model, where $f(t) = \gamma e^{-\gamma t}$, Eqs. (2, 3) lead to the estimate $\mathcal{R}_0 = \mathcal{R}_0^-$. We now compute $\mathcal{R}_0$ for three distribution functions which may be used as kernels: those with a fixed, exponentially or trapezoidally distributed infectious period (see Fig. 1). We refer to these as $\mathcal{R}_0^{\text{fix}}$, $\mathcal{R}_0^{\text{exp}}$ and $\mathcal{R}_0^{\text{trap}}$, respectively. We also compute $\mathcal{R}_0$ for the model with latent and infectious periods that each have gamma distributions, referred to as $\mathcal{R}_0^{(m,n)}$. We have $\mathcal{R}_0^{\text{exp}} = \mathcal{R}_0^{(1,1)}$ and $\mathcal{R}_0^{\text{fix}} = \lim_{m,n\to\infty} \mathcal{R}_0^{(m,n)}$.

2.1 Fixed infectious period

Given fixed latent and infectious periods, T_E and T_I respectively, and assuming f constant when non-zero, we have $f(\tau) = 1/T_I$ for $T_E < \tau < T_E + T_I$ and $f(\tau) = 0$ otherwise. For this distribution $T_G = T_E + T_I/2$ and

$$\mathcal{R}_0 = \mathcal{R}_0^{\text{fix}} = \frac{r(T_G - T_E)}{\sinh r(T_G - T_E)} e^{rT_G} \tag{4}$$

$\mathcal{R}_0^+$ is useful as an estimator for $\mathcal{R}_0^{\text{fix}}$ when the latent period may be regarded as fixed and the infectious period is short relative to the timescale $1/r$ (rT_I is small). As $\sinh x > x$ whenever $x > 0$, and $\lim_{x\to 0} \frac{\sinh x}{x} = 1$ we have $\mathcal{R}_0^{\text{fix}} \leqslant \mathcal{R}_0^+$, and $\lim_{T_E \to T_G} \mathcal{R}_0^{\text{fix}} = \mathcal{R}_0^+$. Wallinga and Lipsitch [23] showed that $\mathcal{R}_0^+$ is an upper bound on estimates of $\mathcal{R}_0$ for any distribution $f(t)$.

2.2 Trapezoidal infection kernel

Consider the kernel

$$f(\tau) = \begin{cases} \frac{1}{T_I} \frac{\tau - \tau_a}{\tau_b - \tau_a} & : \quad \tau \in (\tau_a, \tau_b) \\ \frac{1}{T_I} & : \quad \tau \in (\tau_b, \tau_c) \\ \frac{1}{T_I} \frac{\tau_d - \tau}{\tau_d - \tau_c} & : \quad \tau \in (\tau_c, \tau_d) \\ 0 & : \quad \text{otherwise} \end{cases} \tag{5}$$
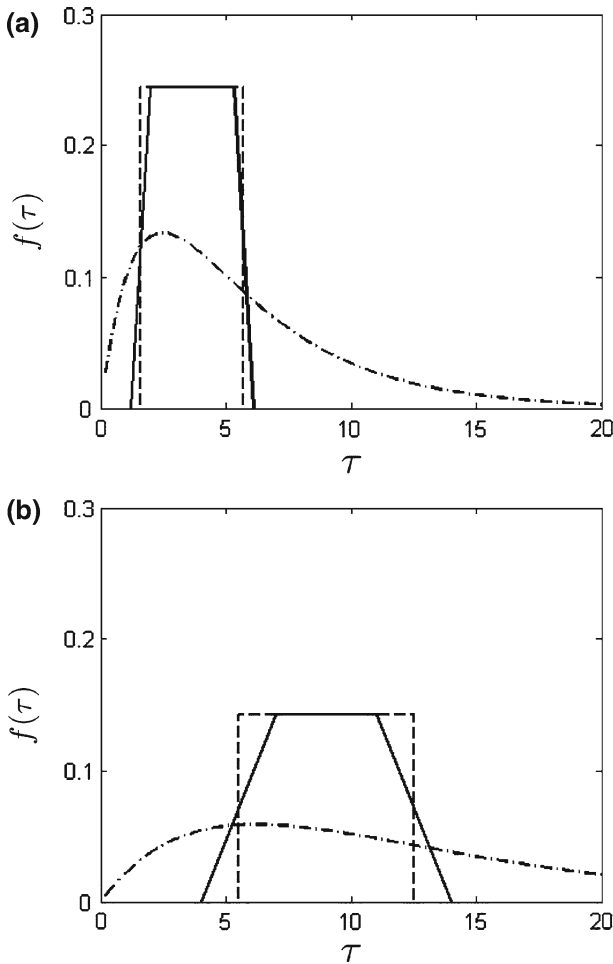
Fig. 1 A selection of kernels, $f(\tau)$ that may be used to model emerging infections: kernels due to fixed and exponentially distributed latent and infectious periods are shown as *dashed* and *dashed/dotted lines*, respectively; and the trapezium kernel is shown as a *solid line*. All kernels are scaled to have integral one. **a** Kernels suitable for modelling influenza, with latent period $T_E = 1.6$ days and infectious period $T_I = 4.0$ days. **b** Kernels suitable for modelling SARS, with latent period $T_E = 5.5$ days and infectious period $T_I = 7.0$ days

This is a suitable approximation to an infectivity function where nobody is infectious before τ_a time units or after τ_d time units post-exposure, maximum infectivity occurs between τ_b and τ_c time units after exposure, and contact rates are constant. The distribution is consistent with a mean latent period of $T_E = \frac{\tau_a + \tau_b}{2}$, a mean infectious period of $T_I = (\tau_d + \tau_c - \tau_b - \tau_a)/2$ and a mean generation interval of

$$T_G = T_E + \frac{T_I}{2} + \frac{(\tau_d - \tau_c)^2 - (\tau_b - \tau_a)^2}{12(\tau_d + \tau_c - \tau_b - \tau_a)}$$

Hence, if the trapezium is symmetric $T_G = T_E + \frac{T_I}{2}$, which is the same relationship as that for the fixed infectious period. The basic reproduction number solves $\mathcal{R}_0^{\text{trap}} \bar{f}(r) = 1$, where $\bar{f}(s)$ is the Laplace transform of $f(t)$ (see Appendix 1)

2.3 SEIR differential equation models

In an extended *SEIR* differential equation model the population of size N is made up of S susceptibles, E that have been exposed to infection but are not yet infectious, I infectious and R that have been infected and recovered. If the epidemic processes have a much faster timescale than the demographic processes, we obtain the equations

$$\frac{dE_1}{dt} = \beta \frac{S}{N} \sum_{j=1}^{n} I_j - m\nu E_1$$

for $i = 2, \ldots, m$
$$\frac{dE_i}{dt} = m\nu E_{i-1} - m\nu E_i$$

$$\frac{dI_1}{dt} = m\nu E_m - n\gamma I_1$$

for $j = 2, \ldots, n$
$$\frac{dI_j}{dt} = n\gamma I_{j-1} - n\gamma I_j$$

$$\frac{dR}{dt} = n\gamma I_n \tag{6}$$

The exposed and infectious classes have been subdivided $E = \sum_{i=1}^{m} E_i$ and $I = \sum_{j=1}^{n} I_j$, respectively. The times spent in the exposed and infectious classes are gamma distributed with means $T_E = \frac{1}{\nu}$ and $T_I = \frac{1}{\gamma}$, respectively, and $\mathcal{R}_0 = \frac{\beta}{\gamma}$. The mean generation interval is $T_G = T_E + \frac{n+1}{2n} T_I$ (see Appendix 2). If the initial rate of exponential increase of the epidemic is r, then

$$\mathcal{R}_0^{(m,n)} = \frac{\frac{2nr}{n+1} (T_G - T_E) \left(1 + \frac{r}{m} T_E\right)^m}{1 - \left(1 + \frac{2r}{n+1} (T_G - T_E)\right)^{-n}}$$

This result is derived in Appendix 2, where it is also shown that given values of r, T_E and T_G, $\mathcal{R}_0^{(m,n)}$ is an increasing function of both m and n.

2.4 Exponentially distributed infectious period

The well-known *SEIR* differential equation model is the special case of Eqs. (6) with $m = n = 1$. For this model the times spent in the exposed and infectious classes are exponentially distributed with means $T_E = \frac{1}{\nu}$ and $T_I = \frac{1}{\gamma}$ respectively, and the appropriate kernel function in Eqs. (2, and 3) is

$$f(\tau) = \frac{\gamma \nu}{\gamma - \nu} \left(e^{-\nu\tau} - e^{-\gamma\tau}\right)$$

(see [6]). The mean generation interval is $T_G = T_E + T_I$, and given r we have

$$\mathcal{R}_0^{\exp} = \mathcal{R}_0^{(1,1)} = 1 + r\left(\frac{1}{\nu} + \frac{1}{\gamma}\right) + \frac{r^2}{\nu\gamma} = 1 + rT_G + r^2 T_E (T_G - T_E) \quad (7)$$

The approximation $\mathcal{R}_0^- = 1 + rT_G \leqslant \mathcal{R}_0^{\exp}$ is appropriate for the *SIR* model, for which $\nu \to \infty$, $T_E \to 0$ and $T_G \to T_I = \frac{1}{\gamma}$. Hence $\mathcal{R}_0^-$ performs best as an estimate when either the latent period T_E or the infectious period T_I is small compared to T_G, and performs worst when they are equal.

3 Method and results

We assumed that we had estimated values of the initial rate of exponential increase of infection incidence, r, the mean latent period, T_E, and the mean generation interval, T_G. We then used Eqs. (4, and 7) to calculate model-based estimates of the basic reproduction number using the assumptions of a fixed or exponentially distributed infectious period, leading to $\mathcal{R}_0^{\text{fix}}$ and $\mathcal{R}_0^{\exp}$, respectively. We did this for values of the ratio of the latent period to the generation interval, T_E/T_G, in the range zero to 0.99. The values of $\mathcal{R}_0^{\text{fix}}$ and $\mathcal{R}_0^{\exp}$ are plotted as functions of T_E/T_G for $rT_G = 0.5, 1.0, 1.5, 2.0$ in Fig. 2, and compared with the values of the estimators $\mathcal{R}_0^- = 1 + rT_G$ and $\mathcal{R}_0^+ = e^{rT_G}$ in those cases. When $rT_G = 0.5, 1.0, 1.5$ or 2.0 we have $\mathcal{R}_0^- = 1.5, 2.0, 2.5$ or 3.0 and $\mathcal{R}_0^+ = 1.65, 2.72, 4.48$ or 7.39, respectively.

The results shown in Fig. 2 illustrate that for fixed values of r, T_E and T_G, the values of $\mathcal{R}_0^-$ and $\mathcal{R}_0^+$ are lower and upper bounds, respectively for both $\mathcal{R}_0^{\exp}$ and $\mathcal{R}_0^{\text{fix}}$. In Sects. (2.1 and 2.4) it was shown that $\mathcal{R}_0^{\text{fix}} \leqslant \mathcal{R}_0^+$ and $\mathcal{R}_0^- \leqslant \mathcal{R}_0^{\exp}$, respectively. It is proved in Appendix 2 that $\mathcal{R}_0^{(m,n)}$ is an increasing function of both m and n. Putting these results together we obtain the inequality

$$\mathcal{R}_0^- \leqslant \mathcal{R}_0^{\exp} \leqslant \mathcal{R}_0^{(m,n)} < \mathcal{R}_0^{\text{fix}} \leqslant \mathcal{R}_0^+ \quad (8)$$

where m and n are any finite positive integers.

Table 1 was constructed to illustrate the results that may be obtained for some specific infections. Parameters were chosen from the literature to be representative of influenza [20], severe acute respiratory syndrome (SARS) [19], smallpox [1] and foot and mouth disease (FMD) [11]. For each infection a trapezium distribution was constructed for $f(\tau)$, and used together with an estimate of $\mathcal{R}_0$ to calculate an estimate of the initial exponential increase, r. The function $f(\tau)$ was also used to calculate values for the mean generation interval T_G, the mean latent period T_E and the mean infectious period T_I.

If these values of r, T_G and T_E had been estimated from data, and then $\mathcal{R}_0$ had been estimated by $\mathcal{R}_0^-$, $\mathcal{R}_0^+$, $\mathcal{R}_0^{\text{fix}}$, $\mathcal{R}_0^{\exp}$ or $\mathcal{R}_0^{\text{trap}}$, the estimates presented in Table 1 would have resulted. By construction, the estimated value of $\mathcal{R}_0^{\text{trap}}$ then corresponds to the assumed value of $\mathcal{R}_0$. Hence Table 1 must be regarded as a comparison of estimates that may be made; the relative values are important rather than the absolute values.
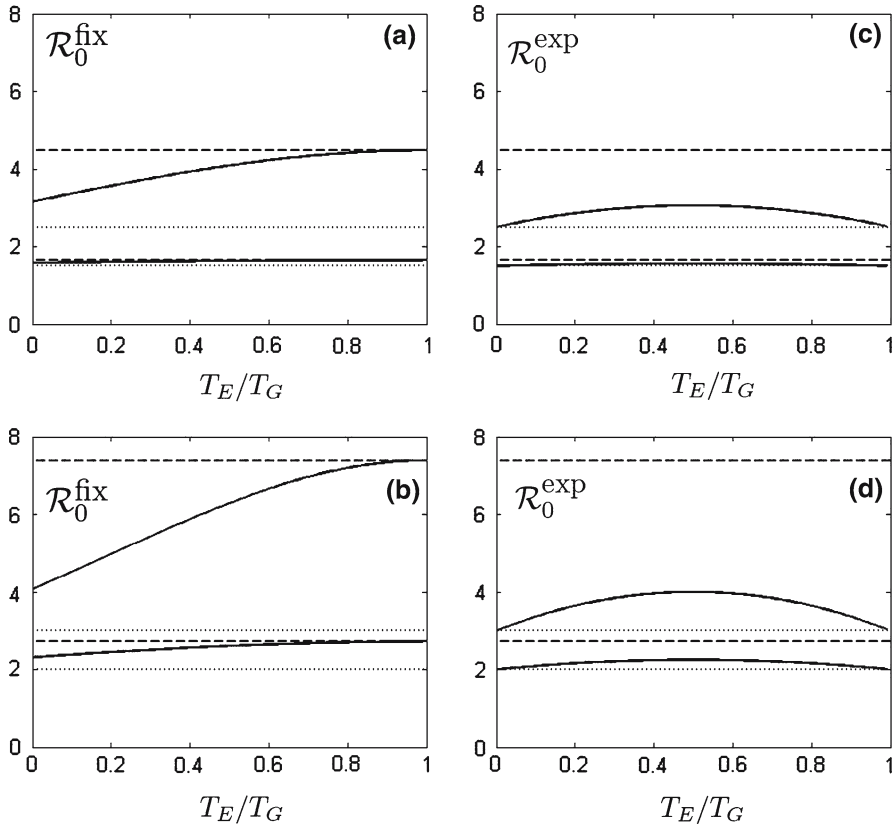
Fig. 2 Estimates of the basic reproduction number $\mathcal{R}_0$ as a function of the ratio of the latent period to the mean generation interval, T_E/T_G. The solid plots are **a,b** fixed infectious period: $\mathcal{R}_0^{\text{fix}}$ with $rT_G = 0.5$ (*lower*), $rT_G = 1.5$ (*upper*); and $rT_G = 1.0$ (*lower*), $rT_G = 2.0$ (*upper*) respectively, and **c,d** exponentially distributed infectious period: $\mathcal{R}_0^{\text{exp}}$ with $rT_G = 0.5$ (*lower*), $rT_G = 1.5$ (*upper*); and $rT_G = 1.0$ (*lower*), $rT_G = 2.0$ (*upper*), respectively. The *horizontal dotted* and *dashed lines* indicate the corresponding values of $\mathcal{R}_0^- = 1 + rT_G$ and $\mathcal{R}_0^+ = e^{rT_G}$, respectively

Table 1 Estimates of $\mathcal{R}_0$ that could be made for emerging infections

Infection	r	T_G	T_E	T_I	$\mathcal{R}_0^-$	$\mathcal{R}_0^+$	$\mathcal{R}_0^{\text{fix}}$	$\mathcal{R}_0^{\text{exp}}$	$\mathcal{R}_0^{\text{trap}}$
Influenza	0.198	3.65	1.60	4.10	1.72	2.06	2.00	1.85	2.00
SARS	0.134	9.00	5.50	7.00	2.21	3.34	3.22	2.55	3.20
Smallpox	0.0576	20.5	15.0	11.0	2.18	3.26	3.20	2.45	3.20
FMD	0.165	6.00	2.00	8.00	1.99	2.70	2.51	2.21	2.50

The initial rate of exponential increase (r day^{-1}), mean generation interval (T_G days), mean latent period (T_E days) and mean infectious period (T_I days) that could be observed for epidemics of influenza, SARS, smallpox; and for foot and mouth disease (FMD) spreading between farms; together with the corresponding estimates of the basic reproduction number made using the approximations $\mathcal{R}_0^- = 1 + rT_G$ and $\mathcal{R}_0^+ = e^{rT_G}$, or assuming a rectangular, exponential or trapeziodal distribution for the infectious period, leading to $\mathcal{R}_0^{\text{fix}}$, $\mathcal{R}_0^{\text{exp}}$ and $\mathcal{R}_0^{\text{trap}}$, respectively

4 Conclusions and discussion

We have derived and discussed model-consistent methods for estimating the basic reproduction number ($\mathcal{R}_0$) for an infectious disease from the initial rate of exponential growth of incidence of infection (r) at the beginning of an epidemic. These methods can only be applied to incidence data from the period where it is reasonable to assume that the whole population may be regarded as susceptible ($S(t) \simeq N$).

Among the first pieces of information obtained for an emerging infection are observations of the latent and infectious periods. These may be used as estimates to construct a rectangular kernel for an integral equation model, or to derive rate parameters for a differential equation model. Examples of infectivity kernels with the same latent and infectious periods are shown in Fig. 1. These kernels appear to be quite different, with the exponential kernel allowing some transmission of infection from time zero, and exhibiting a long infection tail. These features can lead to disparities in the results from modelling exercises. The transmission at early times mitigates against the success of control methods based on contact tracing, or any other method with inherent delays. The tail can lead to transmission appearing to continue in the model long after control measures should have eliminated the infection. The trapezoidal kernels also shown in Fig. 1 allow for some variability in the fixed and latent periods to be incorporated in the model. For example, the kernel shown as suitable for modelling influenza (Fig. 1a) is consistent with a latent period uniformly distributed between 1.2 and 2.0 days, and an infectious period of 4.0 days. The kernel shown as suitable for modelling SARS (Fig. 1b) is consistent with nobody being infectious before 4 days, everybody infectious by 7 days, everybody still infectious at 11 days and nobody infectious after 14 days; with the proportion infectious at intermediate times determined by linear interpolation. As well as allowing for some variability, the trapezoidal kernel has the advantage over the rectangular one that it is a continuous function, and this avoids problems with numerical schemes that do not allow discontinuities. Of course, if further information is available, then other distributions may be more appropriate.

Figure 2 compares estimates of the basic reproduction number based on fixed (Fig. 2a, b) and exponentially distributed (Fig. 2c, d) infectious periods, $\mathcal{R}_0^{\text{fix}}$ and $\mathcal{R}_0^{\text{exp}}$, respectively, with the estimates based only on mean generation interval, $\mathcal{R}_0^- = 1 + r T_G$ and $\mathcal{R}_0^+ = e^{r T_G}$. The estimate $\mathcal{R}_0^-$ is inaccurate whenever $r T_G$ is not small. Using the fixed infectious period model, $\mathcal{R}_0^+$ approximates $\mathcal{R}_0^{\text{fix}}$ when T_E / T_G is near to one, that is when the infection has a long latent period and a short infectious period. The estimate $\mathcal{R}_0^-$ is an approximation to $\mathcal{R}_0^{\text{exp}}$ if either the latent period or infectious period are very short, but $\mathcal{R}_0^+$ is never a good estimator for $\mathcal{R}_0^{\text{exp}}$ and it's use is therefore inconsistent with an *SEIR* model.

The estimates of $\mathcal{R}_0$ derived in this paper have the ordering $\mathcal{R}_0^- \leqslant \mathcal{R}_0^{\text{exp}} \leqslant \mathcal{R}_0^{(m,n)} < \mathcal{R}_0^{\text{fix}} \leqslant \mathcal{R}_0^+$. This inequality establishes that, given the same values of r, T_E and T_I, $\mathcal{R}_0^-$ provides a closer estimate for $\mathcal{R}_0^{\text{exp}}$ than for $\mathcal{R}_0^{\text{fix}}$, and $\mathcal{R}_0^+$ provides a closer estimate for $\mathcal{R}_0^{\text{fix}}$ than for $\mathcal{R}_0^{\text{exp}}$. Even though results derived from the gamma distributed kernel, $\mathcal{R}_0^{(m,n)}$, are not displayed in Fig. 2, we have established that $\mathcal{R}_0^{\text{fix}}$ and $\mathcal{R}_0^{\text{exp}}$ are upper and lower bounds respectively for $\mathcal{R}_0^{(m,n)}$. Note that for the model with

a fixed infectious period, $T_G = T_E + T_I/2$, but for the model with an exponentially distributed infectious period, $T_G = T_E + T_I$. The inequality, and the results presented in Fig. 2 are derived on the assumption that T_E and T_G are the same in both models; T_I is defined consistently with the model, and hence differs between models. This is in contrast to the distributions presented in Fig. 1, which have the same latent and infectious periods, T_E and T_I, and hence different mean generation intervals T_G.

Table 1 shows results that could be obtained when estimating $\mathcal{R}_0$ for emerging infections, with parameters suitable for these infections. The parameter values indicated for each infection should be regarded as sensible values rather than exact estimates, and the results are presented to indicate the relevance of Fig. 2. For all four examples there is close agreement between $\mathcal{R}_0^{\text{fix}}$ and $\mathcal{R}_0^{\text{trap}}$. This is not surprising for the *influenza* example where the trapezium kernel has steep sides (Fig. 1a), but also applies for examples such as *SARS* (Fig. 1b). The results in Table 1 confirm that if rT_G is small, hence $\mathcal{R}_0^-$ is close to one, then $\mathcal{R}_0^-$ may be used as an estimator for $\mathcal{R}_0^{\text{exp}}$ when an exponential model is appropriate. If a model with a fixed infectious period is more appropriate then $\mathcal{R}_0^+$ is a better estimate for $\mathcal{R}_0^{\text{fix}}$, especially when T_E/T_G is closer to one: compare for example the relative values of $\mathcal{R}_0^+$ and $\mathcal{R}_0^{\text{fix}}$ for *smallpox* where $T_E/T_G = 0.73$ and *FMD* where $T_E/T_G = 0.33$.

Recently other estimation methods have been suggested as improvements on $\mathcal{R}_0^-$ and $\mathcal{R}_0^+$. Wearing et al. [25] compared estimates based on $\mathcal{R}_0^{\text{exp}}$ with results obtained using gamma-distributed infection kernels. Lloyd [15] found similar results for models of within-host virus dynamics. Heffernan and Wahl [13] also examined the problem, and provided correction factors for estimates of $\mathcal{R}_0$ based on both the mean and variance of observed transition times. Wallinga and Lipsitch [23] used a similar approach to ours, and derived estimates of $\mathcal{R}_0$ for a selection of infectivity kernels $f(t)$, including those derived from a gamma-distributed infectious period but only with $T_E = 0$. They also considered the case where $f(t)$ is a Normal distribution; if employed though this distribution should be truncated to avoid the possibility of negative generation intervals.

Even though we selected a number of particular kernels for our study, and these cover a reasonable range of first choices, our method is applicable to all biologically sensible kernels. When estimating kernels from data, one should be careful. Estimates made early in an epidemic are likely to be based on household studies, and may be truncated due to local saturation of contacts. In addition, it is unclear how valid such estimates are when extrapolated to the wider community with multiple levels of mixing.

We must be careful in attempting to draw conclusions from our analysis. As a new infection emerges the appropriate model is speculative, and in any situation there is no such thing as the correct model. For example, in the context of pandemic influenza, Ferguson et al. [8] had $T_G = 2.6$ and $T_E = 1.48$ days, and estimated $\mathcal{R}_0 \approx \mathcal{R}_0^-$, obtaining values in the range $1 - 2$. Their model was more complex than those discussed here, but we have seen that for low values of rT_G, $\mathcal{R}_0^-$ provides a reasonable estimate. Mills et al. [18] used an *SEIR* model, and estimated $\mathcal{R}$ from $\mathcal{R}_0^{\text{exp}}$ which is consistent with their model. Our approach is to advocate using an estimate of $\mathcal{R}_0$ that is consistent with the model used to evaluate control strategies.

Appendices

Appendix 1: The derivation of $\mathcal{R}_0^{\text{trap}}$

To find the Laplace transform of the trapezoidal distribution (5), define the functions

$$\phi(\tau_a, \tau_b, s) = \int_{\tau_a}^{\tau_b} e^{-s\tau} \, d\tau = \frac{e^{-s\tau_a} - e^{-s\tau_b}}{s}$$

and

$$\psi(\tau_a, \tau_b, s) = \int_{\tau_a}^{\tau_b} \tau e^{-s\tau} \, d\tau = \frac{e^{-s\tau_a} - e^{-s\tau_b}}{s^2} + \frac{\tau_a e^{-s\tau_a} - \tau_b e^{-r\tau_b}}{s}$$

Then, given an estimated value of r, the basic reproduction number solves

$$\frac{\mathcal{R}_0^{\text{trap}}}{T_I} \left(\frac{\psi(\tau_a, \tau_b, r) - \tau_a \phi(\tau_a, \tau_b, r)}{\tau_b - \tau_a} + \phi(\tau_b, \tau_c, r) \right.$$
$$\left. + \frac{\tau_d \phi(\tau_c, \tau_d, r) - \psi(\tau_c, \tau_d, r)}{\tau_d - \tau_c} \right) = 1$$

Appendix 2: The proof of inequality (8)

For the extended *SEIR* model (6)the probability that an individual infected at time zero is in one of the infected classes (E_i or I_j) at time t is found by solving the differential equations with $\beta = 0$, $E_1(0) = 1$, $E_i(0) = 0$ for $i = 2, \ldots, m$ and $I_j(0) = 0$ for $j = 1, \ldots, n$. In the Laplace transform domain the solutions are

$$\bar{E}_i(s) = \frac{(mv)^{i-1}}{(s + mv)^i} \qquad \bar{I}_j(s) = \frac{(mv)^m}{(s + mv)^m} \frac{(n\gamma)^{j-1}}{(s + n\gamma)^j}$$

The probability that an individual infected at time zero has become infectious or has ceased to be infectious by time t is $h(t)$ or $g(t)$ respectively, where

$$\bar{h}(s) = \frac{(mv)^m}{s(s + mv)^m} \qquad \bar{g}(s) = \frac{(mv)^m (n\gamma)^n}{s(s + mv)^m (s + n\gamma)^n}$$

Back-transforming,

$$h(t) = F_{m,v}(t) \qquad g(t) = \int_0^t \frac{d\,F_{m,v}(x)}{dx} F_{n,\gamma}(t-x)\,dx$$

where

$$F_{m,v}(t) = P(m, mvt) = \frac{1}{(m-1)!} \int_0^{mvt} x^{m-1} e^{-x}\,dx$$

is a regularized incomplete gamma function (see 6.5.1, [2]). The infectivity kernel has Laplace transform

$$\bar{f}(s) = \gamma\left(\bar{h}(s) - \bar{g}(s)\right)$$
$$= \frac{\gamma\,(mv)^m}{s\,(s+mv)^m}\left(1 - \frac{(n\gamma)^n}{(s+n\gamma)^n}\right)$$

The generation interval is given by the formula

$$T_G = \int_0^\infty t f(t)\,dt = -\lim_{s\to 0}\frac{d\bar{f}}{ds} = \frac{1}{v} + \frac{n+1}{2n\gamma}$$

If the initial rate of exponential increase of the epidemic is r, then $\mathcal{R}_0^{(m,n)}\,\bar{f}(r) = 1$, hence

$$\mathcal{R}_0^{(m,n)} = \frac{\frac{r}{\gamma}\left(1 + \frac{r}{mv}\right)^m}{1 - \left(1 + \frac{r}{n\gamma}\right)^{-n}}$$

This result may be found in Wearing et al. [25]. The function $\left(1 + \frac{x}{m}\right)^m$ is positive for positive x, and increases monotonically from $1 + x$ to e^x as m increases from 1 to ∞. Hence, given values of r, v and γ, $\mathcal{R}_0^{(m,n)}$ is an increasing function of m, but a decreasing function of n. However, substituting $\frac{1}{v} = T_E$ and $\frac{1}{\gamma} = \frac{2n}{n+1}(T_G - T_E)$ we obtain

$$\mathcal{R}_0^{(m,n)} = \frac{\frac{2nr}{n+1}(T_G - T_E)\left(1 + \frac{r}{m}T_E\right)^m}{1 - \left(1 + \frac{2r}{n+1}(T_G - T_E)\right)^{-n}}$$

The function

$$f_n(x) = \frac{\frac{nx}{n+1}}{1 - \left(1 + \frac{x}{n+1}\right)^{-n}}$$

is positive for positive x, and increases monotonically from $1 + \frac{x}{2}$ to $\frac{xe^x}{e^x - 1}$ as n increases from 1 to ∞. The proof is a straightforward but tedious manipulation of expressions: we multiply the numerator and denominator of the expression for $f_n(x)$ by $(1 + x/(n+1))^n$ and then show that $\frac{f_{n+1}(x)}{f_n(x)} > 1$ for $n \geq 1$ and $x > 0$. Hence, given values of r, T_E and T_G, $\mathcal{R}_0^{(m,n)}$ is an increasing function of both m and n. In the limit as m and n tend to infinity, $\mathcal{R}_0^{(m,n)}$ tends to $\mathcal{R}_0^{\text{fix}}$. Hence for all positive finite integers m and n, $\mathcal{R}_0^{\text{exp}} \leqslant \mathcal{R}_0^{(m,n)} < \mathcal{R}_0^{\text{fix}}$, completing the proof of inequality (8).

References

1. Aldis, G.K., Roberts, M.G.: An integral equation model for the control of a smallpox outbreak. Math. Biosci. **195**, 1–22 (2005)
2. Abramowitz, M., Stegun, I.A.: Handbook of Mathematical Functions. Dover, New York (1964)
3. Anderson, R.M., May, R.M.: Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, New York (1991)
4. Cauchemez, S., Boëlle, P.-Y., Donnelly, C.L., Ferguson, N.M., Thomas, G., Lueng, G.M., Hedley, A.J., Anderson, R.M., Valeron, A.-J.: Real-time estimates in early detection of SARS. Emerg. Infect. Dis. **12**, 110–113 (2006)
5. Choi, B.C.K., Pak, A.W.P.: A simple approximate mathematical model to predict the number of severe acute respiratory syndrome cases and deaths. J. Epidemiol. Commun. Health **57**, 831–835 (2003)
6. Diekmann, O., Heesterbeek, J.A.P.: Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. Wiley, New York (2000)
7. Diekmann, O., Heesterbeek, J.A.P., Metz, J.A.J.: On the definition and computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. J. Math. Biol. **28**, 365–382 (1990)
8. Ferguson, N., Cummings, D.A.T., Cauchemez, S., Fraser, C., Riley, S., Meeyai, A., Iamsirithaworn, S., Burke, D.S.: Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature **437**, 209–214 (2005)
9. Ferrari, M.J., Bjørnstad, O.N., Dobson, A.P.: Estimation and inference of $\mathcal{R}_0$ of an infectious pathogen by a removal method. Math. Biosci. **198**, 14–26 (2005)
10. Fine, P.E.M.: The interval between successive cases of an infectious disease. Am. J. Epidemiol. **158**, 1039–1047 (2003)
11. Haydon, D., Chase-Topping, M., Shaw, D.J., Matthews, L., Friar, J., Wilesmith, J., Woolhouse, M.E.J.: The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. Proc. R. Soc. Lond B **270**, 121–127 (2003)
12. Heesterbeek, J.A.P.: A brief history of $\mathcal{R}_0$ and a recipe for its calculation. Acta. Biotheor. **50**, 189–204 (2002)
13. Heffernan, J.M., Wahl, L.M.: Improving Estimates of the basic reproductive ratio: using both the mean and dispersal of transition times. Theor. Popul. Biol. **70**, 135–145 (2006)
14. Lipsitch, M., Cohen, E., Cooper, B., Robins, J.M., Ma, S., James, L., Gopalakrishna, G., Chew, S.K., Tan, C., Samore, M.H., Fisman, D., Murray, M.: Transmission dynamics and control of severe acute respiratory syndrome. Science **300**, 1966–1970 (2003)
15. Lloyd, A.L.: The dependence of viral parameter estimates on the assumed viral life cycle: limitations of studies of viral load data. Proc. R. Soc. B **268**, 847–854 (2001)
16. Lloyd-Smith, J.O., Schreiber, S.J., Kopp, P.E., Getz, W.M.: Superspreading and the effects of individual variation on disease emergence. Nature **438**, 355–359 (2005)
17. Meester, R., de Koning, J., de Jong, M.C.M., Diekmann, O.: Modeling and real-time prediction of classical swine fever epidemics. Biometrics **58**, 178–184 (2002)
18. Mills, C.E., Robins, J.M., Lipsitch, M.: Transmissibility of 1918 pandemic influenza. Nature **432**, 904–906 (2004)
19. Roberts, M.G.: Modelling strategies for minimizing the impact of an imported exotic infection. Proc. R. Soc. B **271**, 2411–2415, (2004)

20. Roberts, M.G., Baker, M., Jennings, L.C., Sertsou, G., Wilson, N.: A model for the spread and control of pandemic influenza in an isolated geographical region. J. R. Soc. Interface **4**, 325–330 (2007)
21. Stegeman, J.A., Elbers, A.R.W., Smak, J., de Jong, M.C.M.: Quantification of the transmission of classical swine fever virus between herds during the 1997–1998 epidemic in the Netherlands. Prevent. Veterinary Med. **42**, 219–234 (1999)
22. Svensson, A.: A note on generation intervals in epidemic models. Math. Biosci. **208**, 300–311 (2007)
23. Wallinga, J., Lipsitch, M.: How generation intervals shape the relationship between growth rates and reproductive numbers. Proc. R. Soc. Ser. B **274**, 599–604 (2007)
24. Wallinga, J., Teunis, P.: Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am. J. Epidemiol. **160**, 509–516 (2004)
25. Wearing, H.J., Rohani, P., Keeling, M.J.: Appropriate models for the management of infectious diseases. PLoS Med. **7**, 621–627 (2005)
26. Zhou, G., Yan, G.: Severe acute respiratory syndrome epidemic in Asia. Emerg. Infect. Dis. **9**, 1608–1610 (2003)