

Simple Approximations for Epidemics with Exponential and Fixed Infectious Periods

A. C. Fowler & T. Déirdre Hollingsworth

Bulletin of Mathematical Biology

A Journal Devoted to Research at the Junction of Computational, Theoretical and Experimental Biology Official Journal of The Society for Mathematical Biology

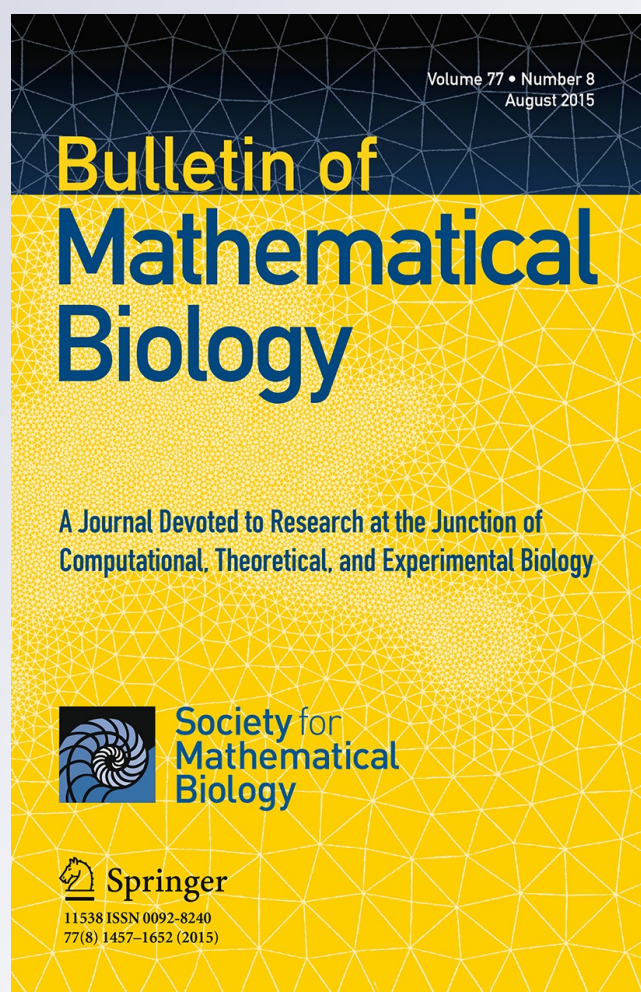
ISSN 0092-8240

Volume 77

Number 8

Bull Math Biol (2015) 77:1539-1555

DOI 10.1007/s11538-015-0095-3



Your article is protected by copyright and all rights are held exclusively by Society for Mathematical Biology. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".



Simple Approximations for Epidemics with Exponential and Fixed Infectious Periods

A. C. Fowler^{1,2} · T. Déirdre Hollingsworth^{3,4,5}

Received: 28 April 2015 / Accepted: 21 August 2015 / Published online: 3 September 2015
© Society for Mathematical Biology 2015

Abstract Analytical approximations have generated many insights into the dynamics of epidemics, but there is only one well-known approximation which describes the dynamics of the whole epidemic. In addition, most of the well-known approximations for different aspects of the dynamics are for the classic susceptible–infected–recovered model, in which the infectious period is exponentially distributed. Whilst this assumption is useful, it is somewhat unrealistic. Equally reasonable assumptions are that the infectious period is finite and fixed or that there is a distribution of infectious periods centred round a nonzero mean. We investigate the effect of these different assumptions on the dynamics of the epidemic by deriving approximations to the whole epidemic curve. We show how the well-known sech-squared approximation for the infective population in ‘weak’ epidemics (where the basic reproduction rate $R_0 \approx 1$) can be extended to the case of an arbitrary distribution of infectious periods having finite second moment, including as examples fixed and gamma-distributed infectious periods. Further, we show how to approximate the time course of a ‘strong’ epidemic, where $R_0 \gg 1$, demonstrating the importance of estimating the infectious period distribution early in an epidemic.

✉ A. C. Fowler
fowler@maths.ox.ac.uk

¹ MACSI, University of Limerick, Limerick, Ireland

² OCIAM, University of Oxford, Oxford, UK

³ Warwick Mathematics Institute, University of Warwick, Coventry, UK

⁴ School of Life Sciences, University of Warwick, Coventry, UK

⁵ Department of Clinical Medicine, Liverpool School of Tropical Medicine, Liverpool, UK

Keywords Epidemics · SIR model · Delay equation · Kermack–McKendrick model · Soper model · Asymptotic methods

1 Introduction

The dynamics of the classic susceptible–infected–recovered (SIR) infectious disease transmission model framework, as first outlined by [Kermack and McKendrick \(1927\)](#), underlies much of our understanding of infectious disease epidemiology ([Anderson and May 1991](#); [Diekmann and Heesterbeek 2000](#); [Keeling and Rohani 2007](#)). Important insights from this model framework include the threshold properties of the basic reproductive number, R_0 ([Kermack and McKendrick 1927](#)), the critical vaccination proportion ([Smith 1964](#)) and the relationship between the epidemic growth rate, r_g , the generation time, T_g , and R_0 . Within their classic paper [Kermack and McKendrick \(1927\)](#) not only derived R_0 but also derived an approximation to the epidemic curve for R_0 close to 1 (a ‘weak’ epidemic). Understanding of these various quantities, although they apply only to relatively simple homogeneous models, has proved very useful in developing our understanding of the characteristics of an epidemic.

The classic SIR model was derived as a special case of a more general formulation with infectiousness varying over the course of the time since infection, or across the population ([Kermack and McKendrick 1927](#); [Diekmann and Heesterbeek 2000](#)). This form of epidemic model in slightly different forms is also known as the ‘Lotka–Euler’ formulation ([Wallinga and Lipsitch 2007](#)) or ‘renewal equation’ ([Fraser 2007](#)). Within this framework, the classic SIR model emerges under the assumption that infectious periods are exponentially distributed across the population. This is, of course, unlikely to be the case in reality. Arguably, the most parsimonious representation of a more realistic infectious period is to assume that the infectious period is limited and is the same across all individuals [the Soper model, [Soper \(1929\)](#)]. The assumption of different distributions of infectious periods is known to affect the relationship between the exponential growth rate and the generation time distribution, and estimates of the reproductive number from the epidemic growth rate ([Fraser 2007](#); [Wallinga and Lipsitch 2007](#); [Wearing et al. 2005](#); [Lloyd 2001a](#); [Hethcote and Tudor 1980](#)), a crucial estimate in the early stages of a new outbreak.

Despite the impact of different infectious period distributions on the dynamics of the early stages of an epidemic, the ‘final epidemic size’ or the total number of infectives over the course of an epidemic has been shown to be invariant under different assumptions on the distribution of infectious periods and disease course within individuals ([Kermack and McKendrick 1927](#); [Bailey 1975](#); [Anderson and Watson 1980](#); [Anderson and May 1991](#); [Andersson and Britton 2000](#); [Diekmann and Heesterbeek 2000](#)), provided there is homogeneous mixing ([Ma and Earn 2006](#); [Anderson and May 1991](#); [Diekmann and Heesterbeek 2000](#); [Andreasen 2011](#)).

In the declining stages of an epidemic, or during the decline of a seasonally forced epidemic, the distribution of infectious periods has been shown to destabilise the dynamics ([Lloyd 2001a](#)) and to change dependence of persistence on the population size ([Lloyd 2001b](#)), results which were first derived for a model in which the infectious period was exponentially distributed ([Keeling and Grenfell 1997](#)). Keeling and

Grenfell and Lloyd came to opposite conclusions concerning persistence, but their results were reconciled by [Conlan et al. \(2010\)](#).

Given these insights, it is surprising that there has not been more investigation of the impact of the infectious period distribution on the peak and decline of the SIR-type models. Here we formulate a general approximation to the epidemic curve for any infectious period distribution within a unified framework. We derive approximations to the time course for R_0 close to 1 ('weak' epidemics) and, innovatively, for larger R_0 ('strong' epidemics). Using these novel approximations, we are able to characterise the impact of infectious period distribution on the time course of the epidemic, including the time to and magnitude of peak prevalence. Despite the simplicity of obtaining numerical solutions of these models, analytic approximations such as those highlighted above are a useful way of characterising the impact of different assumptions on epidemic dynamics, as we demonstrate below.

2 The Generalised Infectious Period Model

We first formulate the general transmission model in which individuals are either susceptible, S , infectious, I or recovered, R for a general infectious period distribution; the epidemic is assumed to occur on a fast timescale, so that births and deaths are not modelled. This type of model formulation has been described and analysed several times, most notably by [Kermack and McKendrick \(1927\)](#).

To proceed, we denote by $i(a, t)$ the number density of the infected cohort having had the disease for a period a . Then

$$I = \int_0^{\infty} i \, da \quad (1)$$

is the total number of infectives, assuming that recovery or removal is inevitable (i.e., $i \rightarrow 0$ as $a \rightarrow \infty$). As with age-dependent population models, or time since infection models, i satisfies the partial differential equation

$$\frac{\partial i}{\partial t} + \frac{\partial i}{\partial a} = -r(a)i, \quad (2)$$

where $r(a)$ is the recovery rate, and is taken to be a function of the time since infection. Suitable initial conditions are

$$\begin{aligned} i &= 0 \quad \text{at} \quad t = 0, \\ i &= i_0(t) \quad \text{at} \quad a = 0, \end{aligned} \quad (3)$$

where the 'recruitment' or incidence rate is

$$i_0(t) = -\dot{S} = kSI, \quad (4)$$

just as in the classic SIR model (Kermack and McKendrick 1927). Integration of (2) leads to

$$\dot{I} = i_0(t) - \int_0^\infty r(a)i \, da. \quad (5)$$

We solve (2) using the method of characteristics. In $t < a$, we have $i = 0$, whilst for $t > a$, we find

$$i = i_0(t - a) \exp \left[- \int_0^a r(a') \, da' \right]. \quad (6)$$

Putting $i_0 = -\dot{S}$ in this, we find, after integrating by parts, that I is given by

$$I = S_0 F(t) - S + \int_0^t K(a)S(t - a) \, da, \quad (7)$$

where we can define S_0 to be the total (pre-infection) population of susceptibles. We use the following notation with respect to the recovery, or infectious, period distribution:

$$F(a) = \exp \left[- \int_0^a r(a') \, da' \right], \quad K(a) = -F'(a) = r(a) \exp \left[- \int_0^a r(a') \, da' \right]. \quad (8)$$

Note that $K(a)$ is the infection time probability density, and that [from (8)]

$$\int_0^\infty K(a) \, da = 1. \quad (9)$$

Using (4), we thus have the generalised Soper model, following the early work by Soper (1929), and its exposition by Wilson and Burke (1942) and Wilson and Worcester (1944),

$$\dot{S} = -kSI, \quad I = S_0 F(t) - S + \int_0^t K(a)S(t - a) \, da. \quad (10)$$

The pre-infection state $S = S_0$, $I = 0$ for $t < 0$ is also described by (10), providing we take $K(a) = 0$ and thus $F(a) = 1$ for $a < 0$. The onset of the epidemic is enabled by initial conditions

$$I = I_0, \quad S = S_0 - I_0 \quad \text{at} \quad t = 0+, \quad (11)$$

and typically we suppose $I_0 \ll S_0$.

2.1 Infectious Period Distributions

Different assumptions regarding the infectious period distributions can be represented by different functional forms of $K(a)$. In these formulations we set the functions to have the same mean infectious period,

$$T = \int_0^\infty a K(a) \, da. \quad (12)$$

We additionally define the second moment,

$$K_2 = \int_0^\infty a^2 K(a) da, \quad (13)$$

for future use.

In this formulation the classic SIR model, with its exponential decay in infectiousness, corresponds to a recovery rate $r = 1/T$ which is independent of age, and a consequent delay kernel

$$K = \frac{1}{T} \exp\left(-\frac{a}{T}\right) \quad (14)$$

with mean T and second moment $2T^2$.

Another plausible assumption is that the infectious period is a fixed constant T . Since $F(a)$ is the fraction of an initial inoculate who still have the disease after period a , we can take $F = 1 - H(a - T)$, where H is the Heaviside step function, and thus K is a delta function,

$$K = \delta(a - T). \quad (15)$$

with mean T and second moment T^2 .

More general kernels can be analysed in the same way, including, for example, the gamma distribution

$$K = \frac{1}{T} \frac{\gamma^\gamma}{\Gamma(\gamma)} \left(\frac{a}{T}\right)^{\gamma-1} \exp\left(-\frac{\gamma a}{T}\right), \quad (16)$$

which has mean T and second moment $T^2(\gamma + 1)/\gamma$ and which takes the limits (14) and (15) when $\gamma = 1$ and $\gamma \rightarrow \infty$, respectively.

2.2 Nondimensionalisation

We analyse (6), (7) and (10) by first rescaling the variables, thus

$$\begin{aligned} i &= \frac{w}{kT^2}, \quad I = \frac{v}{kT}, \quad S = \frac{u}{kT}, \quad K_2 = T^2 \kappa_2, \\ t &\sim T, \quad a \sim T, \quad K \sim \frac{1}{T}, \quad r \sim \frac{1}{T}. \end{aligned} \quad (17)$$

Then we have the dimensionless integrals

$$\int_0^\infty K(a) da = 1, \quad \int_0^\infty a K(a) da = 1, \quad \int_0^\infty a^2 K(a) da = \kappa_2, \quad (18)$$

and the dimensionless equations for u , v and w can thus be written in the form [bearing in mind (18)]

$$\begin{aligned}\dot{u} &= -uv, \\ v &= R_0 - u - \int_0^t K(a)[R_0 - u(t-a)] da, \\ w &= -\dot{u}(t-a)F(a), \quad v = \int_0^\infty w(t,a) da,\end{aligned}\tag{19}$$

where

$$R_0 = kT S_0.\tag{20}$$

The initial values for u and v are, from (11),

$$u = R_0 - v_0, \quad v = v_0 = kT I_0 = \frac{R_0 I_0}{S_0}.\tag{21}$$

Note also that the initial value of v is assumed small and nonzero. The dimensionless kernels for exponentially distributed, gamma-distributed and fixed infectious periods are

$$K(a) = e^{-a}, \quad K(a) = \frac{\gamma^\gamma}{\Gamma(\gamma)} a^{\gamma-1} e^{-\gamma a}, \quad K(a) = \delta(a-1),\tag{22}$$

respectively. Note that the mean of each dimensionless kernel is one and the dimensionless second moments are $\kappa_2 = 2$, $(\gamma + 1)/\gamma$ and 1, respectively.

2.3 Initial Growth

We can find the initial growth rate of the epidemic for general infectious period distributions. We first put $u = R_0 - v_0 e^{\lambda t}$, and expanding (19) for small v_0 and large t , we find

$$\lambda = R_0 \left[1 - \int_0^\infty K(a) e^{-\lambda a} da \right],\tag{23}$$

which has a unique positive root if $R_0 > 1$; we thus identify R_0 as the basic reproduction rate of the epidemic for the general infectious period distribution.

For the gamma-distributed infectious period kernel in (22), the dimensionless epidemic growth rate satisfies

$$\lambda = R_0 \left[1 - \left(\frac{\gamma}{\gamma + \lambda} \right)^\gamma \right],\tag{24}$$

and for the particular cases $\gamma = 1$ (SIR model) and $\gamma = \infty$ (Soper model), we find

$$\begin{aligned}\lambda_{\text{SIR}} &= R_0 - 1, \quad \gamma = 1, \\ \lambda_{\text{Soper}} &= R_0 (1 - e^{-\lambda}), \quad \gamma = \infty.\end{aligned}\tag{25}$$

Whilst these approximations to the early epidemic growth rate are useful, they do not tell us about the dynamics of the whole epidemic. We now present approximations to the whole epidemic curve firstly for epidemics for which R_0 is close to one (“weak” epidemics), and then for large R_0 (“strong” epidemics).

3 Weak Epidemics

For the case where $R_0 \approx 1$, Kermack and McKendrick derived a classic approximation to the epidemic curve for the model with exponentially distributed infectious periods. Soper derived a similar expression for the model with fixed infectious periods. We rederive these expressions by showing that this approximation can be generalised for any infectious period distribution with finite second moment.

We first define

$$R_0 = 1 + \varepsilon, \quad (26)$$

and take $\varepsilon \ll 1$. We then rescale the variables by writing

$$t = \frac{\tau}{\varepsilon}, \quad u = 1 - \varepsilon U(\tau), \quad v = \varepsilon^2 V. \quad (27)$$

Substituting these changes into (19), using (18), we obtain

$$\begin{aligned} \dot{U} &= (1 - \varepsilon U)V, \\ V &= \frac{1}{\varepsilon} \left[U - \int_0^\infty K(a)U(\tau - \varepsilon a) da \right], \end{aligned} \quad (28)$$

where the overdot denotes differentiation with respect to τ , and we have replaced the upper limit on the integral by ∞ on the basis that we have

$$u(t) \equiv R_0, \quad v(t) \equiv 0 \quad \text{for } t < 0 \quad (29)$$

in (19) (the epidemic is initiated at $t = 0$). The initial conditions are, from (21),

$$U \approx -1, \quad V = \frac{v_0}{\varepsilon^2} \quad \text{at } t = 0, \quad (30)$$

and we assume that $v_0 \ll \varepsilon^2$.

We now expand $U(\tau - \varepsilon a)$ in the integral in a Taylor series, and, using (18), this leads to

$$V = \dot{U} - \frac{1}{2}\varepsilon\kappa_2\ddot{U} + \dots; \quad (31)$$

substituting this into (28)₁, using the boundary conditions

$$U \approx -1, \quad \dot{U} \approx 0 \quad \text{at } \tau = 0 \quad (32)$$

[the latter from (31)] then leads to the leading order equation

$$\dot{U} \approx \frac{(1 - U^2)}{\kappa_2}, \quad (33)$$

providing the second moment κ_2 exists, essentially equivalent to requiring that $K(a) \ll \frac{1}{a^3}$ for large a . For a heavy-tailed distribution with unbounded second moment, a more

elaborate procedure would be necessary. We do not pursue this here, but note that the breakdown of the method is associated with the nonuniform convergence of the Taylor expansion of $U(\tau - \varepsilon a)$ for large a , because of (29). The correct procedure can be obtained by replacing the upper limit in the integral in (28)₂ by τ/ε .

The solution to the Eq. (33) is

$$U = \tanh\left(\frac{\tau - \tau_p}{\kappa_2}\right). \quad (34)$$

Therefore

$$u = 1 - \varepsilon \tanh\left(\frac{\tau - \tau_p}{\kappa_2}\right) \quad (35)$$

and

$$v = \frac{\varepsilon^2}{\kappa_2} \operatorname{sech}^2\left(\frac{\tau - \tau_p}{\kappa_2}\right). \quad (36)$$

This shows that the approximation to the epidemic curve for low R_0 by [Kermack and McKendrick \(1927\)](#) for the exponential distribution and [Wilson and Worcester \(1944\)](#) for fixed infectious periods is generalisable to any infectious period distribution with a finite second moment.

If we first compare the result of this approximation for the SIR model and the constant infectious period (Soper) model, the approximations are

$$\begin{aligned} v_{\text{SIR}} &\approx \frac{(R_0 - 1)^2}{2} \operatorname{sech}^2\left\{\frac{1}{2}(R_0 - 1)(t - t_p)\right\}, \\ v_{\text{Soper}} &\approx (R_0 - 1)^2 \operatorname{sech}^2\{(R_0 - 1)(t - t_p)\}. \end{aligned} \quad (37)$$

Of note here is the factor of two, which means that the epidemic with a constant infectious period will grow (and therefore decay) more rapidly than that with an exponentially distributed infectious period. It also means that the approximate maximum prevalence for the constant infectious period model, $(R_0 - 1)^2$, is twice as big as for the SIR model. This effect of a constant infectious period on shortening the ‘generation time’ (time from infection to onward transmission) has been previously noted by, amongst others, [Diekmann and Heesterbeek \(2000\)](#) and [Wallinga and Lipsitch \(2007\)](#), but its effect on peak prevalence has not been previously approximated.

For gamma-distributed infectious periods, the epidemic curve is approximated by

$$v \approx \frac{\gamma (R_0 - 1)^2}{\gamma + 1} \operatorname{sech}^2\left\{\frac{\gamma (R_0 - 1)(t - t_p)}{(\gamma + 1)}\right\}, \quad (38)$$

and the bigger the shape parameter, γ (resulting in smaller variance in infectious periods) the higher the peak prevalence and the shorter is the duration of the outbreak.

3.1 Peak Prevalence and Time to Peak

From (36), (17), (20) and (26), the peak prevalence P , defined as the ratio of the maximum infected number I_{\max} to the total population S_0 is, for a weak epidemic,

$$P = \frac{I_{\max}}{S_0} = \frac{(R_0 - 1)^2}{R_0 \kappa_2}. \quad (39)$$

If the initial infected population consists of I_0 individuals, then the initial value of v is given by (21), and since by assumption this is very small, we can suppose v reaches its maximum when τ is large, in which case we can use the approximation $\operatorname{sech}(-\theta) \approx 2e^{-\theta}$, and the dimensionless time to peak prevalence (scaled with T), is then found from (36) to be

$$t_p = \frac{\kappa_2}{2(R_0 - 1)} \ln \left[\frac{4(R_0 - 1)^2 S_0}{\kappa_2 R_0 I_0} \right]. \quad (40)$$

4 Strong Epidemics

Now we consider the case $R_0 \gg 1$, for which we devise an asymptotic method similar to that used by Fowler (1982). First we rescale the variables as follows:

$$u = R_0 U, \quad v = R_0 V, \quad (41)$$

so that

$$\begin{aligned} \delta \dot{U} &= -UV, \\ V &= 1 - U - \int_0^t K(a) [1 - U(t - a)] da, \end{aligned} \quad (42)$$

where

$$\delta = \frac{1}{R_0} \ll 1. \quad (43)$$

There is an initial phase where $U \approx 1$, and we have

$$U \approx 1 - \frac{I_0}{S_0} e^{\lambda t}, \quad (44)$$

where λ is given by (23), using also the fact that $u + v = R_0$ at $t = 0$, and v_0 is given by (21). Since $R_0 \gg 1$, the application of Laplace's method to (23) shows that

$$\lambda \approx R_0. \quad (45)$$

Note that (44) can thus be written in the form

$$U = 1 - \exp\left(\frac{t - t_0}{\delta}\right), \quad (46)$$

where

$$t_0 = \frac{1}{R_0} \ln\left(\frac{S_0}{I_0}\right). \quad (47)$$

The approximation becomes invalid when $t \approx t_0$, and the appropriate rescaling of (42) is done by choosing

$$t = t_0 + \delta\tau. \quad (48)$$

The equation (42) become

$$\begin{aligned} U' &= -UV, \\ V &= 1 - U - \int_0^{t_0 + \delta\tau} K(a) [1 - U(t_0 + \delta\tau - a)] da, \end{aligned} \quad (49)$$

where the prime denotes differentiation with respect to τ .

For small δ , (46) implies that $U(t) \approx 1$ for $t < t_0$, and this implies that the integral in (49) is small, so that V can be approximated by

$$V \approx 1 - U, \quad (50)$$

and therefore

$$U' \approx -U(1 - U). \quad (51)$$

Note that (46) implies

$$U = 1 - e^\tau \quad (52)$$

for $\tau < 0$, and the solution of (51) which matches to this as $\tau \rightarrow -\infty$ is

$$U \approx \frac{1}{1 + e^\tau}, \quad V \approx \frac{e^\tau}{1 + e^\tau}. \quad (53)$$

The solution in (53) is a monotonic solution in which the number of infectives rapidly increases to a peak at $V \approx 1$, i.e., $v \approx R_0$, whilst U decreases towards zero: everybody gets infected! However, the approximation (50) and therefore (51) clearly break down when $\tau \sim 1/\delta$, and a further rescaling is then necessary.

As τ becomes large, we rescale back to the original time scale $t = t_0 + \delta\tau$. Note that then $U \sim e^{-\tau} = e^{-(t-t_0)/\delta}$, and this suggests we write

$$U = \exp\left(-\frac{\phi}{\delta}\right), \quad t > t_0, \quad (54)$$

with

$$\phi \sim t - t_0 \quad \text{as } t \rightarrow t_0; \quad (55)$$

then (42) becomes

$$\begin{aligned}\dot{\phi} &= V, \\ V &= 1 - \exp\left[-\frac{\phi}{\delta}\right] - \int_0^t K(a) [1 - U(t-a)] da,\end{aligned}\quad (56)$$

together with the matching condition (55).

In the integral, we may take $U(t-a) \approx 1$ for $t-a < t_0$, whilst $U(t-a) = \exp\left[-\frac{\phi(t-a)}{\delta}\right]$ for $t-a > t_0$. The exponential terms are small and can be neglected, and therefore

$$\dot{\phi} = V \approx 1 - \int_0^{t-t_0} K(a) da = \int_{t-t_0}^{\infty} K(a) da, \quad (57)$$

with $\phi \sim t - t_0$ as $t \rightarrow t_0$, and thus, interchanging the order of integration in the quadrature for ϕ ,

$$\phi = \int_0^{\infty} \min(a, t - t_0) K(a) da, \quad t > t_0, \quad (58)$$

and $\phi \rightarrow 1$ as $t \rightarrow \infty$. Thus U reaches equilibrium and V declines to zero; no further approximations are necessary.

Because the approximation has two distinct phases, it is less easy to extract such quantities as peak prevalence and time to peak. To do this, we can write a uniformly asymptotic approximation. We write the small and large time approximations in terms of t , thus

$$\begin{aligned}u &= \frac{R_0}{1 + e^{R_0(t-t_0)}}, \quad v = \frac{R_0}{1 + e^{-R_0(t-t_0)}}, \quad t \lesssim t_0, \\ u &= R_0 \exp\left[-R_0 \int_0^{\infty} \min(a, t - t_0) K(a) da\right], \quad v = R_0 \int_{t-t_0}^{\infty} K(a) da, \quad t > t_0.\end{aligned}\quad (59)$$

A uniform approximation is essentially obtained by adding the two approximations and subtracting the common part; for details see [Dyke \(1975\)](#). In the present case we can write a uniform approximation by inspection. This is

$$u \approx \frac{R_0}{1 + \exp\left[R_0 \int_0^{\infty} \min(a, t - t_0) K(a) da\right]}, \quad v = \frac{R_0}{1 + e^{-R_0(t-t_0)}} \int_{t-t_0}^{\infty} K(a) da, \quad (60)$$

providing we extend the definition of K so that $K(a) = 0$ for $a < 0$; it is clear that these expressions reduce to both approximations in (59) in the appropriate time range.

4.1 Peak Prevalence and Time to Peak

The peak time is approximately t_0 given by (47), but the peak value is not well constrained. To find this, we use the uniform approximation for v to find the time where it is maximum; this is the peak time t_p . It is given implicitly by

$$t_p = t_0 + t', \quad t' = \frac{1}{R_0} \ln \left[\frac{R_0 F(t') - K(t')}{K(t')} \right], \quad (61)$$

thus

$$t_p \approx \frac{1}{R_0} \ln \left[\frac{S_0(R_0 F - K)}{I_0 K} \right]. \quad (62)$$

Evidently t' is small, so that $F \approx 1$, but the precise expression for t' depends critically on the behaviour of the distribution kernel $K(a)$ near $a = 0$. For the gamma distribution (22), we have

$$K(a) \approx \frac{\gamma^\gamma c^{\gamma-1}}{\Gamma(\gamma) R_0^{\gamma-1}} \quad \text{for } a = \frac{c}{R_0} \ll 1, \quad (63)$$

and in that case

$$t' \approx \frac{c}{R_0}, \quad c = \ln \left[\frac{R_0^\gamma \Gamma(\gamma)}{\gamma^\gamma c^{\gamma-1}} \right], \quad (64)$$

so that

$$t' \approx \frac{\gamma \ln R_0}{R_0}, \quad t_p \approx 1 - \frac{1}{R_0} \ln \left[\frac{S_0 R_0^\gamma}{I_0} \right]. \quad (65)$$

From (60), the maximum of v is approximately $R_0 F - K$, so that the peak infected population is to leading order the whole population. More accurately, the peak prevalence

$$P = \frac{I_{\max}}{S_0} = \frac{v_{\max}}{R_0} = F - \frac{K}{R_0} \approx 1 - \frac{(\gamma c)^{\gamma-1} (c + \gamma)}{\Gamma(\gamma) R_0^\gamma}, \quad (66)$$

this last expression being for the gamma distribution. For the SIR problem for which $\gamma = 1$, and $F(a) = K(a) = e^{-a}$, we have more directly from (62) that

$$t_p \approx \frac{1}{R_0} \ln \left[\frac{(R_0 - 1) S_0}{I_0} \right], \quad (67)$$

and using this directly in (60) yields the peak prevalence as

$$P \approx \frac{(R_0 - 1)^{(R_0-1)/R_0}}{R_0}. \quad (68)$$

The limit in which $\gamma \rightarrow \infty$ corresponds to the Soper problem where $K(a) = \delta(a - 1)$, and (61) is irrelevant. Direct inspection of (60) shows that in this case v rapidly rises, reaches a maximum $\approx R_0(1 - e^{-R_0})$ at $t = t_0 + 1$, and is then instantly extinguished. This last result (from (60)) is not quite right, as it ignores the corrective terms in (56). More precisely, we have from (42), with $K(a) = \delta(a - 1)$ and taking $t > 1$,

$$V = U(t - 1) - U(t), \quad (69)$$

and we can use (53) throughout, since although it is inaccurate for $t > t_0$, U is in any case very small then. Thus the uniform approximate solution for the Soper case is

$$v = \frac{R_0}{1 + e^{-R_0(t-t_0)}} \left[\frac{1 - e^{-R_0}}{1 + e^{R_0(t-t_0-1)}} \right], \quad (70)$$

and the term in square brackets provides the correction to the step function in (60). From this we find the time to peak is

$$t_p = t_0 + \frac{1}{2}, \quad (71)$$

and the peak prevalence is given by

$$P = \frac{v_{\max}}{R_0} = \frac{1 - e^{-R_0}}{\left(1 + e^{-\frac{1}{2}R_0}\right)^2}. \quad (72)$$

4.2 Accuracy of the Approximations

Figure 1 compares numerical simulations of the model with the weak approximation we have given for the case $R_0 = 1.5$. The shapes of the curves and the time to peak are surprisingly well represented, despite the fact that $\varepsilon = 0.5$ is not that small, though the peak values are overestimated.

For large $R_0 \gg 1$ the difference between the dynamics for the two extremes of the infectious period distribution is striking (Fig. 2). The assumption of a constant infectious period results in a much faster decline of the epidemic following peak prevalence than for an exponentially distributed infectious period. The approximation to the epidemic curve is almost exact for large R_0 and constant infectious period. In particular, it captures both peak prevalence and the time at which it occurs very well

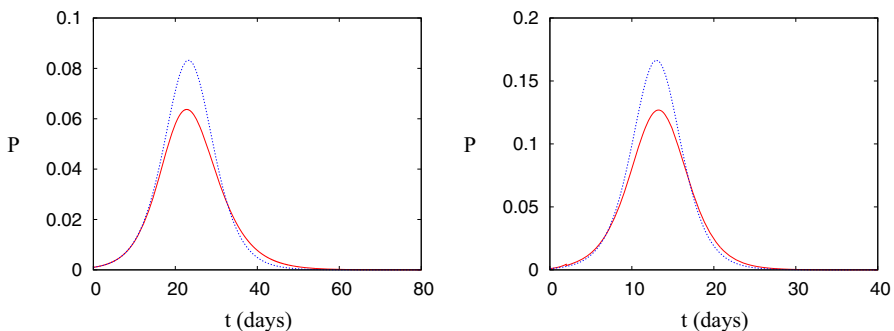


Fig. 1 Comparison between numerical simulations of the model (solid (red) lines) and approximations to the epidemic curve for $P = I/S_0$ for the weak approximation (dashed (blue) lines), using a value of $R_0 = 1.5$ and a mean recovery time of 2 days. The left hand figure is for the SIR model, and the right hand one is for the Soper model. The initial fraction of infectives $I_0/S_0 = 10^{-3}$. Notice the different scales on the axes (Color figure online)

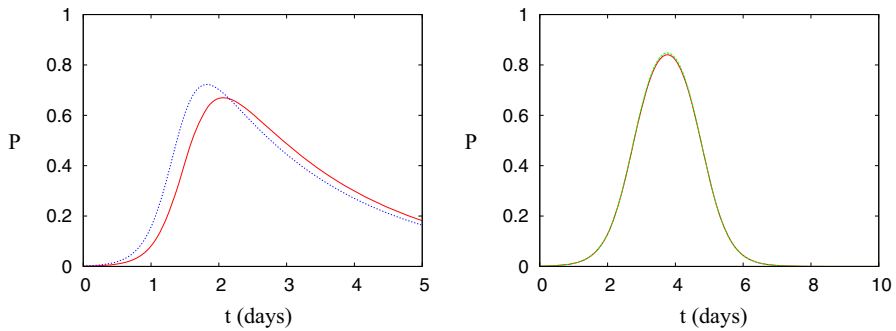


Fig. 2 Comparison between numerical simulations of the model (*solid (red) lines*) and approximations to the epidemic curve for $P = I/S_0$ for the strong approximation (*dashed (blue or green) lines*), using a value of $R_0 = 10$ for the SIR model (*left*) and $R_0 = 5$ for the Soper model (*right*); the mean recovery time is 2 days. The initial fraction of infectives $I_0/S_0 = 10^{-3}$. Note that the Soper approximation is almost exact (Color figure online)

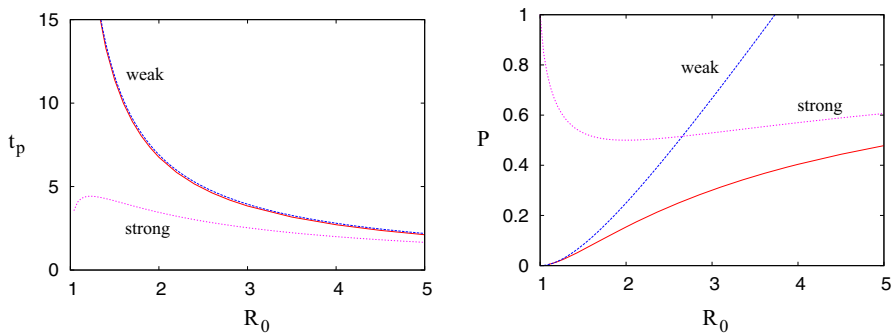


Fig. 3 Comparison of time to peak t_p approximations given by (40) (with $\kappa_2 = 2$) and (67) with direct numerical simulations for the SIR model (*left figure*), and the equivalent approximations to the peak prevalence $P = v_{\max}/R_0$ given by (39) (with $\kappa_2 = 2$) and (68) together with the direct numerical simulation. The *solid (red) curves* are the numerical solution, *dashed (blue) curves* are the weak approximations, and the *dotted (mauve) curves* are the strong approximation. Note that the times to peak here are dimensionless (Color figure online)

once R_0 is larger than about 3. This is shown in Fig. 3, where we plot the time to peak and peak prevalence for both numerical and asymptotic results as a function of R_0 .

These figures provide a gloss on the examples shown in Figs. 1 and 2. It can be seen in Fig. 3 that the weak approximation gives a uniformly excellent approximation to t_p for the SIR model. In fact it deviates at smaller R_0 (!), due to the fact that for fixed I_0 , the initial value of V in (30) increases as ε is reduced. This is more clearly visible in Fig. 4 for the Soper model. The strong approximation (for the SIR model), on the other hand, is only reasonably accurate for $R_0 \gtrsim 5$. The peak prevalence is not well approximated in either limit: the weak approximation is useful for $R_0 \lesssim 0.4$, and the strong approximation only becomes useful for $R_0 \gtrsim 10$, as also illustrated by Fig. 2.

For the Soper model, and more generally when there is a peaked infection time distribution, the weak and strong approximations cope better between them. The peak

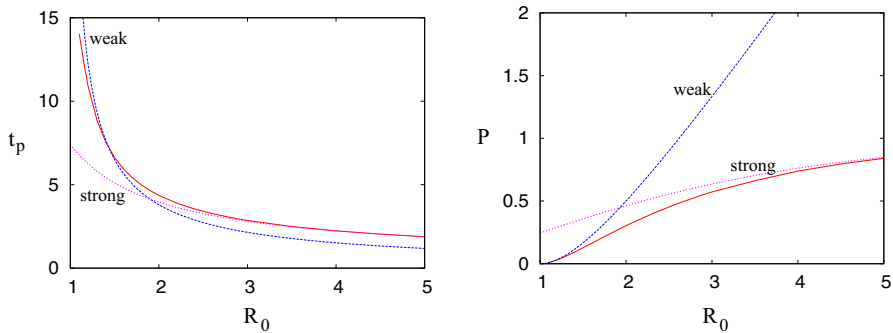


Fig. 4 Comparison of time to peak t_p approximations given by (40) (with $\kappa_2 = 1$) and (71) with direct numerical simulations for the Soper model (*left figure*), and the equivalent approximations to the peak prevalence $P = v_{\max}/R_0$ given by (39) (with $\kappa_2 = 1$) and (72) together with the direct numerical simulation. The solid (red) curves are the numerical solution, dashed (blue) curves are the weak approximations, and the dotted (mauve) curves are the strong approximation. Note that the times to peak here are dimensionless (Color figure online)

time t_p is well served by one or other approximation either side of $R_0 = 2$; the peak prevalence is less well captured, but still fares much better than the exponential distribution of the SIR model.

5 Conclusions

We have derived analytical approximations to the epidemic curve for small and large R_0 for general infectious period distributions. The weak epidemic limit is well known, but not in its application to such distributions, and particularly, its inadmissibility for heavy-tailed distributions provides a new insight. The extreme version of such a distribution corresponds to an immune carrier which can thus act as a reservoir for the infection. The strong epidemic limit has not, to our knowledge, been studied before.

Approximations to the epidemic curve are not only useful for developing our understanding of the dynamics, but may also be used in situations where large numbers of simulations are required, e. g., for parameter estimation or identifying optimal control strategies for a given set of parameters. The approximations presented here capture the general shape of the epidemic, as characterised by the peak prevalence, the time at which the peak occurs, and the rate of increase and decline of the epidemic. In the case of the constant infectious period and large R_0 , the approximation is almost exact.

These analytical expressions extend previous observations on the effect of assumptions regarding the infectious period on the dynamics of the epidemic. For an infection with a given R_0 and mean infectious period, the distribution of this infectious period has an impact on all aspects of the epidemic curve. The differences in the shape of the epidemic curves due to different infectious period distributions is most notable for large R_0 , where the epidemic curve for the SIR model is asymmetric. For $R_0 \approx 1$, the epidemic shape is symmetric due to the slow epidemic growth smoothing out the effect of different infectious period distributions. However, different distributions do affect the height and width of this symmetric curve.

The assumption of an exponentially distributed infectious period results in a broad distribution in times from one infection to another (the generation time) meaning that some individuals infect susceptibles much later than others. This serves to smooth out the profile, resulting in a less intense epidemic than one in which the infectious period is fixed. For low R_0 , the shapes are qualitatively similar, but both peak prevalence and the time to peak are approximately half that of an epidemic with a constant infectious period. An exponential distribution of infectious periods results not only in a slower epidemic with a lower peak prevalence but it lengthens the tail of the epidemic, giving a more gradual decline in prevalence following saturation than a fixed infectious period. This effect of different infectious periods on the generation time and exponential growth rate has been discussed before (Fraser 2007; Wallinga and Lipsitch 2007; Wearing et al. 2005; Lloyd 2001a; Hethcote and Tudor 1980), but the impact on the second half of the epidemic has not. These extreme differences in the tail of an epidemic demonstrate the importance of quantifying this distribution during a novel outbreak in order to estimate peak prevalence and the timescale of the decline of the epidemic.

There are, of course, limitations to the application of an SIR-type approach, whether with constant or exponentially distributed periods, since more detailed biological effects, such as a period prior to infectivity represented in SEIR models, will affect the dynamics (Wearing et al. 2005). Further, homogeneous models of epidemic spread necessarily make a number of unrealistic assumptions about contact patterns, which have been previously been shown to affect the final size of the epidemic (Ma and Earn 2006; Anderson and May 1991; Diekmann and Heesterbeek 2000; Andreasen 2011), and will therefore affect the dynamics during the epidemic. However, the insights about the timing and magnitudes of peak prevalence and the rate of decline of the epidemic are useful for understanding and validating the dynamics of more complex models.

Acknowledgments A. C. F. acknowledges the support of the Mathematics Applications Consortium for Science and Industry (www.macsi.ul.ie) funded by the Science Foundation Ireland grant 12/1A/1683. T. D. H. thanks Imperial College for provision of an Imperial College Junior Research Fellowship.

References

- Anderson D, Watson R (1980) On the spread of a disease with gamma distributed latent and infectious periods. *Biometrika* 67(1):191–198
- Anderson RM, May RM (1991) *Infectious diseases of humans*. Oxford University Press, Oxford
- Andersson H, Britton T (2000) Stochastic epidemic models and their statistical analysis. Lecture notes in statistics, vol 151. Springer, New York
- Andreasen V (2011) The final size of an epidemic and its relation to the basic reproduction number. *Bull Math Biol* 73:2305–2321
- Bailey NTJ (1975) *The mathematical theory of infectious diseases and its application*, 2nd edn. Griffin, London
- Conlan AJK, Rohani P, Lloyd AL, Keeling M, Grenfell BT (2010) Resolving the impact of waiting time distributions on the persistence of measles. *J R Soc Interface* 7:623–640
- Diekmann O, Heesterbeek JAP (2000) *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. Wiley, Chichester
- Fowler AC (1982) An asymptotic analysis of the logistic delay equation when the delay is large. *IMA J Appl Math* 28:41–49

- Fraser C (2007) Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS ONE* 2(8):e758
- Hethcote HW, Tudor DW (1980) Integral equation models for endemic infectious diseases. *Theor Popul Biol* 18:204–243
- Keeling MJ, Grenfell BT (1997) Disease extinction and community size: modeling the persistence of measles. *Science* 275:65–67
- Keeling MJ, Rohani P (2007) Modeling infectious diseases in humans and animals. Princeton University Press, Princeton
- Kermack WO, McKendrick AG (1927) Contributions to the mathematical theory of epidemics. *Proc R Soc A* 115:700–721
- Lloyd AL (2001a) Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. *Proc R Soc B* 268:985–993
- Lloyd AL (2001b) Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics. *Theor Popul Biol* 60:59–71
- Ma J, Earn DJD (2006) Generality of the final size formula for an epidemic of a newly invading infectious disease. *Bull Math Biol* 68:679–702
- Smith CEG (1964) Factors in the transmission of virus infections from animals to man. *Sci Basis Med Ann Rev* 1964:125–150
- Soper HE (1929) The interpretation of periodicity in disease prevalence. *J R Stat Soc* 92:34–73
- Van Dyke MD (1975) Perturbation methods in fluid mechanics. Parabolic Press, Stanford
- Wallinga J, Lipsitch M (2007) How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc R Soc B* 274:599–604
- Wearing HJ, Rohani P, Keeling MJ (2005) Appropriate models for the management of infectious diseases. *PLoS Med* 7:e174
- Wilson EB, Burke MH (1942) The epidemic curve. *Proc Nat Acad Sci* 28:361–367
- Wilson EB, Worcester J (1944) A second approximation to Soper's epidemic curve. *Proc Nat Acad Sci* 30:37–44