Journal of Theoretical Biology 262 (2010) 107–115

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Epidemics with general generation interval distributions

Joel C. Miller [a,b,c,*], Bahman Davoudi [c], Rafael Meza [c], Anja C. Slim [e], Babak Pourbohloul [c,d]

[a] Fogarty International Center, National Institutes of Health, Bethesda, MD, USA
[b] Harvard School of Public Health, Boston, MA, USA
[c] Division of Mathematical Modeling, British Columbia Centre for Disease Control, Vancouver, British Columbia, Canada
[d] School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, Canada
[e] School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA

ARTICLE INFO

Article history:
Received 13 May 2009
Received in revised form
5 August 2009
Accepted 5 August 2009
Available online 11 August 2009

Keywords:
Epidemic
Generation interval
Stochastic growth
Deterministic growth

ABSTRACT

We study the spread of susceptible-infected-recovered (SIR) infectious diseases where an individual's infectiousness and probability of recovery depend on his/her "age" of infection. We focus first on early outbreak stages when stochastic effects dominate and show that epidemics tend to happen faster than deterministic calculations predict. If an outbreak is sufficiently large, stochastic effects are negligible and we modify the standard ordinary differential equation (ODE) model to accommodate age-of-infection effects. We avoid the use of partial differential equations which typically appear in related models. We introduce a "memoryless" ODE system which approximates the true solutions. Finally, we analyze the transition from the stochastic to the deterministic phase.

1. Introduction

Despite many medical advances in recent history, infectious diseases continue to impact public health. The recent SARS epidemic, the ongoing pandemic of novel H1N1 (swine) influenza, and the simmering threat of H5N1 avian influenza or other diseases call attention to the need to develop simple modeling tools in preparation for future emerging pandemics. Such a pandemic could have typical generation interval measured in days or weeks, spread worldwide, and grow quickly. In the face of such an emerging disease, there would be little time to develop and implement interventions.

The ability to predict the timing and maximum patient load imposed by an epidemic is essential to intervention design. Overestimating the preparation time available or underestimating the peak may result in well-designed measures which are implemented too late or are too small.

The ability of an infectious disease to spread depends strongly on the proportion of the population that is susceptible S/N. We will find that the details of the spread are more sensitive to changes in N/S than changes in S/N (as S decreases, a small change in S/N may correspond to a large change in N/S), and so we couch most of our discussion in terms of changes in N/S.

Fig. 1 shows the course of an epidemic of an infectious disease whose characteristics are discussed later (Section 3.2.2 with $c=0.9$). At very early times the disease spreads as a branching process and stochastic effects are important. As the outbreak grows, the spread continues as a branching process, but stochastic effects lose importance. However, the epidemic timing always feels the initial stochastic impact. Eventually the proportion of the population still susceptible decreases and the epidemic dies out.

We also consider $i(t,\tau)$, the number of people infected at time t who have been infected for τ units of time (their 'infection-age'). We plot the cumulative infection-age distribution in Fig. 1 at small t (center) and larger t (right). At small t the distributions are noisy, and converge to a steady-state distribution as t increases. As spread continues, N/S begins to change perceptibly and the steady-state adjusts quasistatically if N/S changes slowly enough. If N/S does not change slowly, the system cannot adjust to the changing equilibrium. During the growing phase of the epidemic, the infected individuals are weighted toward more recent infections, while during the declining phase the infected individuals have disproportionately older infections.

We focus on several stages in this paper: the early stochastic phase, the later deterministic phase, and the transition phase between these two. If S is initially small, then N/S can change significantly during the stochastic phase. We do not address this case.

Typically disease outbreaks are either subcritical (meaning $\mathscr{R}_0<1$) for which epidemics have zero probability because an average infected person infects fewer than one individual, or supercritical (meaning $\mathscr{R}_0>1$) for which epidemics are possible.

* Corresponding author at: Fogarty International Center, National Institutes of Health, Bethesda, MD, USA.
E-mail address: joel.c.miller.research@gmail.com (J.C. Miller).

0022-5193/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2009.08.007

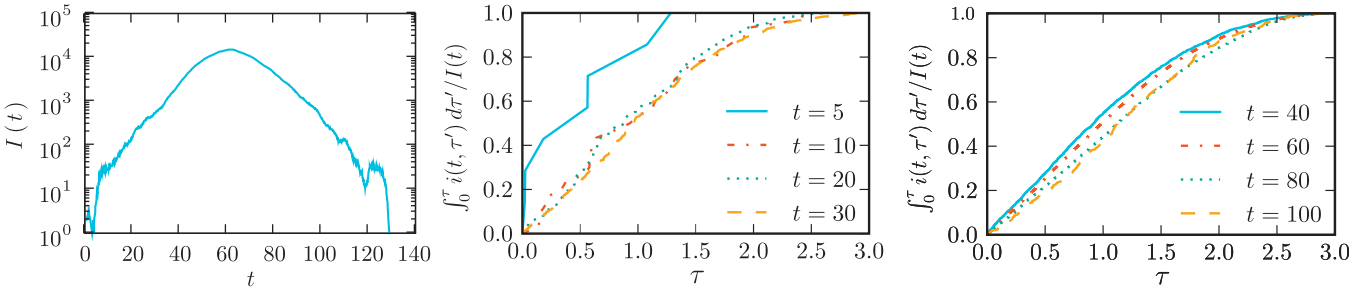

Fig. 1. The course of an epidemic with vertical logscale (left). The cumulative age-of-infection distribution $\int_0^\tau i(t,\tau')\,d\tau'/I(t)$ at different times (center and right).

We consider only supercritical outbreaks. Early in an outbreak's spread, growth is dominated by stochastic effects, and it may die out stochastically. If it persists, it may grow faster or slower than "average". As long as N/S does not change significantly, the spread can be modeled using Crump-Mode-Jagers (CMJ) processes (Crump and Mode, 1968,; 1969; Jagers, 1975; Haccou et al., 2005). A subcritical CMJ process dies out, while a supercritical CMJ process either dies out or converges to $Ce^{\phi t}$ where C is a random value and ϕ depends on the process.

If a supercritical outbreak becomes sufficiently large the spread is effectively deterministic. The usual equations for this phase are the susceptible-infected-recovered (SIR) equations

$$\dot{S} = -\beta IS/N \tag{1}$$

$$\dot{I} = \beta IS/N - \gamma I \tag{2}$$

$$\dot{R} = \gamma I \tag{3}$$

These equations assume that infected people cause infections at rate β and recover at rate γ, giving an exponentially distributed infection duration. The process is "memoryless". In contrast, for real diseases the "age" of an individual's infection affects his/her infectiousness and probability of recovering.

Ignoring "age-of-infection" effects loses important details. During the growth of an epidemic the infections are biased toward young infection ages. If young infections are more (or less) infectious, the SIR equations under- (or over-) estimate the growth rate. Similar observations hold during decay.

Several approaches have been developed to study age-of-infection models. Some explicitly track the history of the epidemic (Breban et al., 2005; Hethcote and van den Driessche, 2000; Brauer, 2005, 2008; Li and Brauer, 2008; Castillo-Chavez et al., 1989; Thieme and Castillo-Chavez, 1993). Others maintain the memoryless feature of Eqs. (1)–(3) by introducing a chain of infected compartments $I_1,\ldots,I_n$ in order to approximate the infectious period distribution (Anderson and Watson, 1980; Wearing et al., 2005; Ma and Earn, 2006; Gunther et al., 2008; Lloyd, 2001a, b). These chains of compartments usually do not have biological meaning, but instead are a simplifying "trick". Typically these assume constant β and that each of n infected classes recovers at rate γn, resulting in gamma-distributed infectious periods with constant infectiousness.

In this paper we investigate the growth of an outbreak from a single infection to a full-scale epidemic, without the restrictive assumptions underlying Eqs. (1)–(3). In Section 2, we show how to model the early stochastic phase and give comparison with deterministic predictions. In Section 3 we show how to find deterministic equations governing the epidemic's growth. We take a different approach from most previous studies and arrive at a system similar to the standard Eqs. (1)–(3) rather than a partial differential equation. If the change in N/S is not large during a typical infectious period, we can approximate the infectious population as being in equilibrium given N/S and arrive at a memoryless system that captures the dynamics well. In Section 4 we examine what it means for the outbreak to be large enough to be effectively deterministic.

2. Stochastic phase

We assume that the disease spreads from individual to individual in such a way that the ability of individual u to infect a susceptible individual depends only on how long u has been infected and whether or not u has recovered. We let $P(\tau)$ be the probability u is still infected τ units of time after becoming infected. If u is still infected, the rate u causes new infections is $\beta(\tau)S/N$. This enforces a possibly unrealistic assumption that infectiousness is independent of total infection duration. It would be straightforward to modify the model to incorporate this effect, but we do not do it here.

We have $P(0)=1$ and—assuming no one remains infectious forever—$P(\infty)=0$. We assume P is differentiable. The probability of recovering in a short interval $(\tau,\tau+\Delta\tau)$ is $-P'(\tau)\Delta\tau+\mathcal{O}(\Delta\tau^2)$, and so as $\Delta\tau\to 0$, we may assume the probability approaches $-P'(\tau)\Delta\tau$. We let $P_{rec}(\tau)$ be the probability density function (pdf) for recovery: $P_{rec}(\tau)=-P'(\tau)\geq 0$.

2.1. The equations

We have full derivations of the equations in Appendix A. If $p_k(t)$ is the probability that k individuals are infected at time t, then the probability generating function (pgf) $f(x,t)=\sum_{k=0}^{\infty}p_k(t)x^k$ provides a useful tool to help calculate p_k. We get

$$f(x,t)=xP(t)g(x,t|t)+\int_0^t g(x,t|\tau)P_{rec}(\tau)\,d\tau \tag{4}$$

Here $g(x,t|\tau)=\sum q_k(t|\tau)x^k$ is the pgf for the number of descendants an individual has t units of time after its infection given that it recovers $\tau\leq t$ units of time after infection. That is $q_k(t|\tau)$ is the probability an individual has k infectious descendants t units of time after becoming infected given that it recovers after $\tau\leq t$ units of time.

We find (for $\tau\leq t$)

$$g(x,t|\tau)=\exp\left(\int_0^\tau [f(x,t-\theta)-1]\beta(\theta)\,d\theta\right) \tag{5}$$

To find equations for $p_k(t)$ we take the k-th derivative of f, divide by $k!$, and evaluate at $x=0$. We solve the equations as described in Appendix B.

We compare the solutions with 50 000 simulations in Fig. 2. We take the pdf of the infection duration to be a Weibull distribution, $W(5.8, 2.59)$, so $P(\tau)=e^{-(\tau/5.8)^{2.59}}$. We take constant

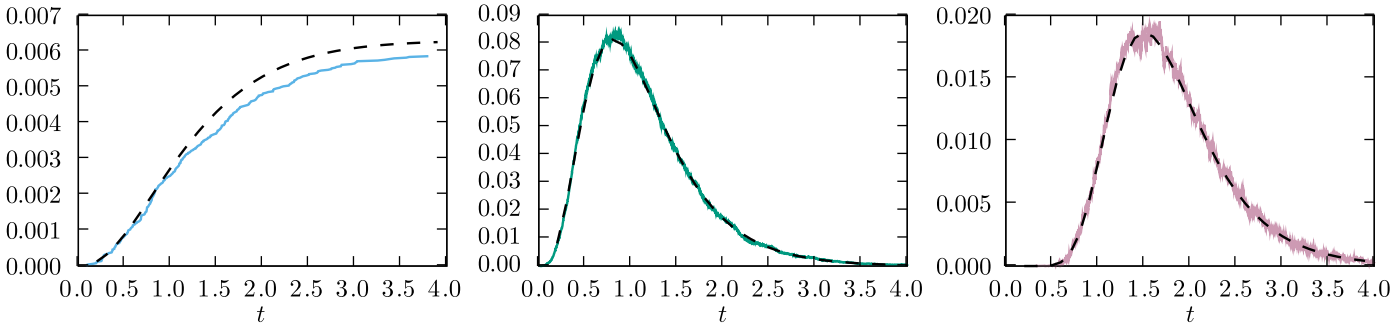

Fig. 2. The probability of having 0, 5, or 20 people infected as functions of time beginning with a single index case: comparison of theory (dashed) and 50 000 simulations for Weibull distributed infectious durations $W(5.8,2.59)$ with greek symbol beta = 2.

$\beta = 2$. Although there is considerable noise in simulations, we find close match with analytic results.

2.2. Asymptotic behavior at large I

If $S(0)$ is large, then N/S may still be approximately constant even as I becomes much larger than 1. We are interested in the behavior of I as it becomes large, but before N/S has changed significantly. If we assume $N/S = 1$ remains fixed, then under weak assumptions it can be shown (Feller, 1941; Crump and Mode, 1968, 1969) that $I(t)$ either becomes zero at some finite time or it converges to $C e^{\phi t}$ where C is a random number determined by stochastic effects and ϕ solves

$$1 = \int_0^\infty e^{-\phi\tau}\beta(\tau)P(\tau)\,\mathrm{d}\tau$$

This equation is the *Euler–Lotka* (EL) equation, which we derive in Section 3. The solution ϕ is unique and known as the *Malthusian parameter*. The most significant assumption we require for convergence is that the infection is not a "lattice" process, that is, possible times of infection are not discretized so I can change continuously.[1] This result guarantees that if the susceptible population is sufficiently large, the outbreak either dies out or grows and becomes effectively deterministic.

We have shown that Eqs. (4) and (5) accurately predict the probability of having a given number of infections as a function of time. Once the outbreak is sufficiently large, the impact of stochastic effects is reduced and the infected population size scales like $C e^{\phi t}$ for fixed ϕ. The random value of C determines how much time is available to prepare for the epidemic.

2.3. Distribution of epidemic onset times

We use a simpler disease process to investigate the impact of the stochastic phase on how quickly an epidemic "takes off". We consider a population with constant infectiousness and exponentially distributed infection durations (corresponding to a constant recovery rate). We compare predictions of Eqs. (4) and (5) with predictions from the deterministic Eqs. (1)–(3) which are exactly valid precisely for this infection process. We take $\beta = 1.5$ and $\gamma = 1$.

Fig. 3 shows that if the initial number of infections is low, it is relatively likely that the number infected becomes large before the deterministic equations predict it should. This has a number of implications for interpreting early stages of an outbreak. If we attempt to predict the present size of an outbreak given a known introduction date using the assumption of deterministic growth,

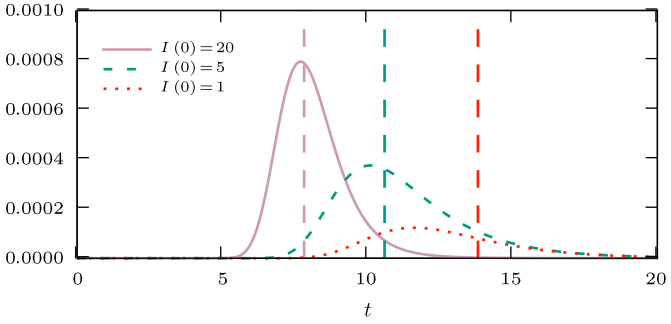

Fig. 3. A comparison of the deterministically predicted time at which 1000 individuals are infected (vertical dashed lines) with the actual probabilities of having 1000 individuals infected at each time given different numbers of initial infections using an exponentially distributed infection duration with $\gamma = 1$ and constant infectiousness with $\beta = 1.5$.

we are likely to underpredict the current size. Consequently if we make preparations to introduce interventions under the assumption of deterministic growth, we may be using interventions that are too small and implemented too late.

The mismatch decreases as the initial number of infections increase. We explain this observation by noting that different outbreaks with only a few infections grow *on average* at the deterministically predicted rate. However, those at the lower range of growth often go extinct, while those at the higher range tend to become epidemics quickly. This leads to the important conclusion that if an epidemic happens, it is likely to happen faster than the deterministic equations predict.

Although we have been considering the distribution of times given outbreak size, we could also consider the distribution of sizes given time. The first is relevant to situations in which we have an intervention which we plan to implement if the outbreak reaches a given size, and so we want to know how long it may be until we implement it. The second is relevant to situations in which we will be able to implement an intervention at a given time, and we want to know how large the outbreak may be.

We can make some analytic progress on calculating the expected size of epidemics at sufficiently large times that the non-epidemic outbreaks have died out. In general the expected size of epidemics will be the expected size of all outbreaks divided by the probability of an epidemic. We demonstrate this calculation under the assumption infectiousness is constant and equal to β and infection duration is exponentially distributed with parameter γ. Assuming a single index case, during the exponentially growing phase the expected size of outbreaks is

$$E[I(t)] = e^{(\beta-\gamma)t}$$

If we set $\mathscr{Y}$ to be the probability an epidemic happens given a single index case and restrict our attention to those outbreaks

[1] For lattice processes, similar results apply with discrete rather than continuous time.

which will become epidemics we get

$$E[I(t)|\text{epidemic}] = \frac{\gamma}{\beta-\gamma}\left(\frac{\beta}{\gamma}\,e^{(\beta-\gamma)t} - e^{-(\beta-\gamma)t}\right) \approx \frac{\beta}{\beta-\gamma}\,e^{(\beta-\gamma)t} \approx E[I(t)]/\mathscr{Y}$$

because epidemics occur with probability $\mathscr{Y} = (\beta-\gamma)/\beta$. Thus epidemics are shifted forward in time by an average of d units of time where $e^{\phi d} = 1/\mathscr{Y}$.

If the disease does not have exponentially distributed infection durations or constant infectiousness, then these calculations will be modified, but the time shift will be the same.

3. Deterministic phase

In this section we develop the deterministic equations governing epidemics once stochastic effects are unimportant. Our exact equations are equivalent to many previous age-of-infection models (Breban et al., 2005; Hethcote and van den Driessche, 2000; Brauer, 2005, 2008; Li and Brauer, 2008; Castillo-Chavez et al., 1989; Thieme and Castillo-Chavez, 1993), but we avoid the usual use of PDEs. A related approach also avoiding PDEs was used by Brauer (2005), but we cast our equations in a form similar to the standard SIR equations (1)–(3). We then introduce an approximation to these equations. We discuss the transition from the stochastic phase to the deterministic phase in Section 4.

In the stochastic phase analysis, we assumed that infectiousness is independent of the recovery time (except that after recovery infectiousness is zero). We can drop this assumption here without any additional complications and redefine $\beta(\tau)$ as the average rate of infection τ units of time after infection for those individuals still infected. The product $\beta(\tau)P(\tau)$ represents the expected rate of new infections (of which a fraction S/N are successful) caused by an individual u infected τ units of time previously, where the expectation is taken without prior knowledge of whether u has recovered. We normalize this by $\mathscr{R}_0 = \int_0^\infty \beta(\tau)P(\tau)\,\mathrm{d}\tau$ to arrive at the generation interval distribution $\beta(\tau)P(\tau)/\mathscr{R}_0$ (Svensson, 2007; Wallinga and Lipsitch, 2007).

Let $b(t)$ denote the rate of new infections occurring at time t and $d(t)$ the rate of recoveries. Let $i(t,\tau)$ denote the number of people who became infected at time $t-\tau$ and are still infected at time t. Then $i(t,\tau) = b(t-\tau)P(\tau)$. We can find b in terms of i by $b(t) = \int_0^\infty i(t,\tau)(S/N)\beta(\tau)\,\mathrm{d}\tau$ and d in terms of b by $d(t) = \int_0^\infty b(t-\tau)P_{rec}(\tau)\,\mathrm{d}\tau$.

We look for a solution of the form $b(t) = C\,e^{\xi(t)}$. We have

$$C\,e^{\xi(t)} = \int_0^\infty C\,e^{\xi(t-\tau)}\frac{S(t)}{N}\beta(\tau)P(\tau)\,\mathrm{d}\tau$$

Rearrangement gives

$$e^{\xi(t)}\frac{N}{S(t)} = \mathscr{F}[\xi,t]$$

where we define $\mathscr{F}[\xi,t] = \int_0^\infty e^{\xi(t-\tau)}\beta(\tau)P(\tau)\,\mathrm{d}\tau$.

We derive equations for I and S in terms of ξ as follows: the derivative of S is $-b(t) = -C\,e^{\xi(t)}$. We multiply by $1 = I/\int_0^\infty i(t,\tau)\,\mathrm{d}\tau$, using $i(t,\tau) = b(t-\tau)P(\tau) = C\,e^{\xi(t-\tau)}P(\tau)$ to get

$$\dot{S} = -\frac{Ie^{\xi(t)}}{\mathscr{G}[\xi,t]} = -\frac{\mathscr{F}[\xi,t]\,IS}{\mathscr{G}[\xi,t]\,N}$$

where $\mathscr{G}[\xi,t] = \int_0^\infty e^{\xi(t-\tau)}P(\tau)\,\mathrm{d}\tau$. Repeating this for $\dot{I} = b(t) - d(t)$ we get

$$\dot{I} = \frac{I}{\mathscr{G}[\xi,t]} - \frac{\mathscr{H}[\xi,t]}{\mathscr{G}[\xi,t]}I = \frac{\mathscr{F}[\xi,t]\,IS}{\mathscr{G}[\xi,t]\,N} - \frac{\mathscr{H}[\xi,t]}{\mathscr{G}[\xi,t]}I$$

where $\mathscr{H}[\xi,t] = \int_0^\infty e^{\xi(t-\tau)}P_{rec}(\tau)\,\mathrm{d}\tau$. This can be written in a similar form to the standard SIR equations, except that the coefficients change in time and depend on the history of the epidemic

$$\dot{S} = -\hat{\beta}(t)\frac{IS}{N} \tag{6}$$

$$\dot{I} = \hat{\beta}(t)\frac{IS}{N} - \hat{\gamma}(t)I \tag{7}$$

$$\dot{R} = \hat{\gamma}(t)I \tag{8}$$

$$\mathscr{F}[\xi,t] = \frac{N}{S}\,e^{\xi(t)} \tag{9}$$

where $\hat{\beta}(t) = \mathscr{F}[\xi,t]/\mathscr{G}[\xi,t]$ and $\hat{\gamma}(t) = \mathscr{H}[\xi,t]/\mathscr{G}[\xi,t]$. Because of the similarity in notation, we distinguish $\hat{\beta}(t)$ to be the average rate of causing infection of all individuals infected at time t, while $\beta(\tau)$ is the average rate of causing infection by an individual still infected τ units of time after becoming infected. To initialize the problem we need $\xi(t)$ for all $t<0$ as well as $S(0)$ and $I(0)$. Typically we will assume that $\xi(t) = -\infty$ for $t<0$ so that $e^{\xi(t)} = 0$. As we solve forward, new values of ξ are calculated based on the change in S. The history of $\xi(t-\tau)$ for $\tau>0$ encodes all information needed about the age-of-infection distribution at t. A less intuitive, but simpler formulation of these equations appears in Appendix C.

Note that if β is constant, then the ratio $\hat{\beta} = \mathscr{F}/\mathscr{G}$ is constant and equal to β. Similarly, if infection durations are exponentially distributed with parameter γ, then $\hat{\gamma} = \mathscr{H}/\mathscr{G}$ is also constant and equal to γ.

3.1. Approximating the solution

Storing the history of an outbreak introduces some mild analytical and computational difficulties. It is convenient to work with a system that depends only on its current state.

If N/S is constant the age-of-infection distribution converges to a steady-state where $i(t,\tau)/I(t)$ is independent of t. The infected population size grows or decays exponentially, so $b(t) = C\,e^{t\phi}$ where ϕ solves the modified EL equation

$$C\,e^{t\phi} = \int_0^\infty C\,e^{(t-\tau)\phi}\frac{S}{N}\beta(\tau)P(\tau)\,\mathrm{d}\tau$$

$$\Rightarrow \frac{N}{S} = \int_0^\infty e^{-\tau\phi}\beta(\tau)P(\tau)\,\mathrm{d}\tau$$

This has been used at early times (Wallinga and Lipsitch, 2007) when $N/S \approx 1$ to relate the exponential growth in time ϕ with $\mathscr{R}_0$.

Of course N/S is not constant, but if it varies slowly relative to how quickly ξ changes, we can assume that the system responds quasistatically to changes in N/S and so the age-of-infection distribution is at equilibrium with the current value of N/S. This assumption will allow us to create equations analogous to Eqs. (1)–(3) with changing coefficients, which may be solved by standard ODE methods. This approach will break down if N/S changes significantly during a typical infectious period. Fortunately, we can use the results of the approximation to identify when the approximation fails.

We replace $\xi(t-\tau)$ by $\xi(t) - \int_0^\tau \phi(t-\theta)\,\mathrm{d}\theta$ where $\phi(t) = \xi'(t)$ and approximate $\mathscr{F}/e^{\xi}$, $\mathscr{G}/e^{\xi}$, and $\mathscr{H}/e^{\xi}$ by $F(\phi)$, $G(\phi)$, and $H(\phi)$ respectively assuming that $\phi(t-\tau) \approx \phi(t)$ for the range of τ which make a significant contribution to the integral

$$F(\phi) = \int_0^\infty e^{-\tau\phi(t)}\beta(\tau)P(\tau)\,\mathrm{d}\tau$$

$$G(\phi) = \int_0^\infty e^{-\tau\phi(t)}P(\tau)\,\mathrm{d}\tau$$

$$H(\phi) = \int_0^\infty e^{-\tau\phi(t)} P_{rec}(\tau)\, d\tau$$

Note that each of these is a Laplace transform. The resulting approximating equations are

$$\dot{S} = -\hat{\beta}_0(t)\frac{IS}{N} \tag{10}$$

$$\dot{I} = \hat{\beta}_0(t)\frac{IS}{N} - \hat{\gamma}_0(t)I \tag{11}$$

$$\dot{R} = \hat{\gamma}_0(t)I \tag{12}$$

$$F(\phi) = \frac{N}{S(t)} \tag{13}$$

where $\hat{\beta}_0(t) = F(\phi)/G(\phi)$ and $\hat{\gamma}_0(t) = H(\phi)/G(\phi)$.

Computationally this system of equations is only mildly more difficult than the standard SIR equations. We can either find the functional forms of the Laplace transforms, or simply calculate them for various ϕ in advance. Once that is done, then at each time step, we need only look at N/S, identify ϕ such that $F(\phi) = N/S$, and then find G and H. Then the integration proceeds as in the standard SIR equations.

The approximation is valid as long as the amount of change of N/S during a typical infectious period is small, and is therefore valid well into the nonlinear regime after the exponential growth phase has ended.

3.2. Examples

3.2.1. The usual suspects

If we make the usual assumptions of constant infectiousness and exponentially distributed recovery time (β constant and $P_{rec}(\tau) = \gamma e^{-\gamma\tau}$) the system is memoryless. The function ξ encodes the age-of-infection distribution, which is irrelevant in a memoryless system. Thus the equations for I and S should not depend on ξ. We find $\mathscr{F}[\xi, t] = \beta\mathscr{G}[\xi, t]$, and so $\dot{S} = -\beta IS/N$. We similarly find $\mathscr{H}[\xi, t]/\mathscr{G}[\xi, t] = \gamma$ and so $\dot{I} = \beta IS/N - \gamma I$. So in this special case the exact age-of-infection model (6)–(9) reduces to the standard SIR equations (1)–(3). This holds even for our approximate system (10)–(13).

3.2.2. A piecewise continuous example

We take

$$\beta(\tau) = \begin{cases} c & 0 \le \tau \le 1 \quad \text{or} \quad 2 \le \tau \le 3 \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

$$P_{rec}(\tau) = \begin{cases} 1/2 & 1 \le \tau \le 3 \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

So people are initially infectious, then stop being infectious at $\tau = 1$ and begin to recover. At $\tau = 2$, they continue recovering, but become infectious once more. By $\tau = 3$ all individuals have recovered. Such a system could model a disease in which individuals are infectious before and possibly after having symptoms, but self-isolate during the symptomatic phase. The generation interval distribution is given by

$$\frac{\beta(\tau)P(\tau)}{\mathscr{R}_0} = \begin{cases} 4/5 & 0 \le \tau \le 1 \\ \dfrac{2(3-\tau)}{5} & 2 \le \tau \le 3 \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

In Fig. 4 we find that the exact model (6)–(9) fits the simulations well (with the discrepancy due to stochastic shifts in time). The difference in timing between the exact and the

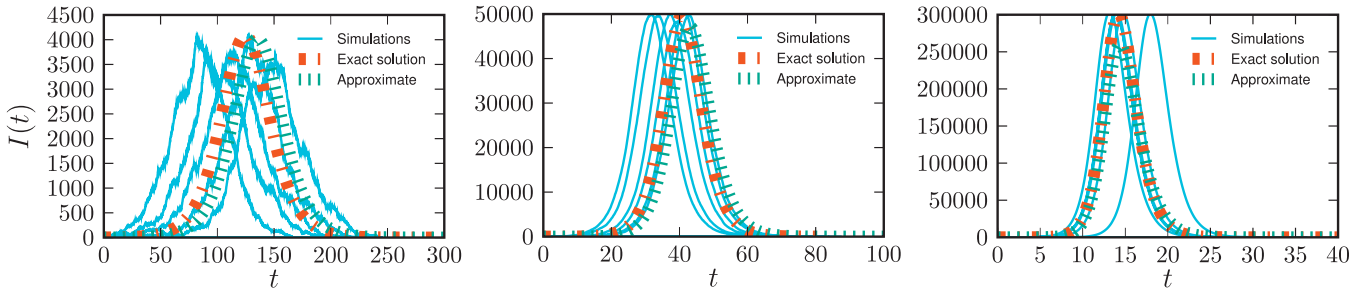

Fig. 4. Comparison of simulations with exact age-of-infection model and approximation for the system of section 3.2.2 with $c = 0.85$, 1, and 1.5 in a population of size 10^6. The temporal shift of the exact and approximate solutions is a result of difference in initial condition. The exact solution takes the initial condition that $\xi(t) = 0$ for $t < 0$ while the approximate solution assumes that $\xi(t) = t\phi(0)$ for $t < 0$. The temporal shift relative to simulations results from the variable time simulations take to become deterministic.

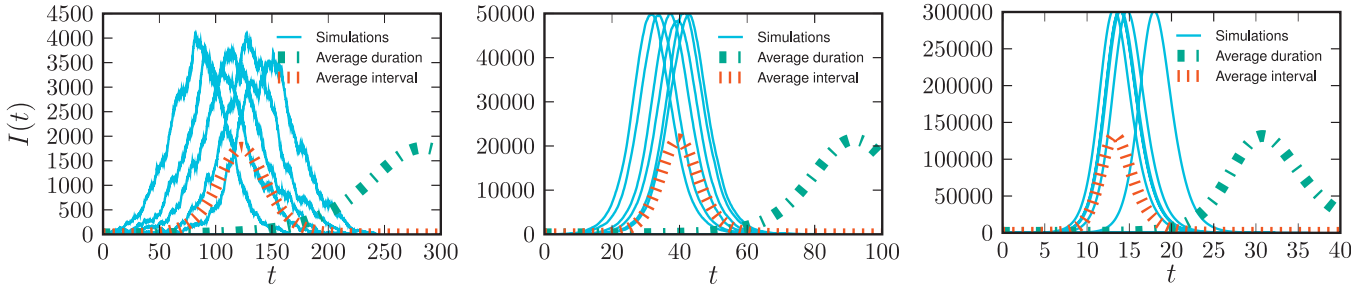

Fig. 5. The standard SIR equations cannot closely capture the dynamics of the disease spread, regardless of whether we preserve the average duration of infection or the average generation interval.

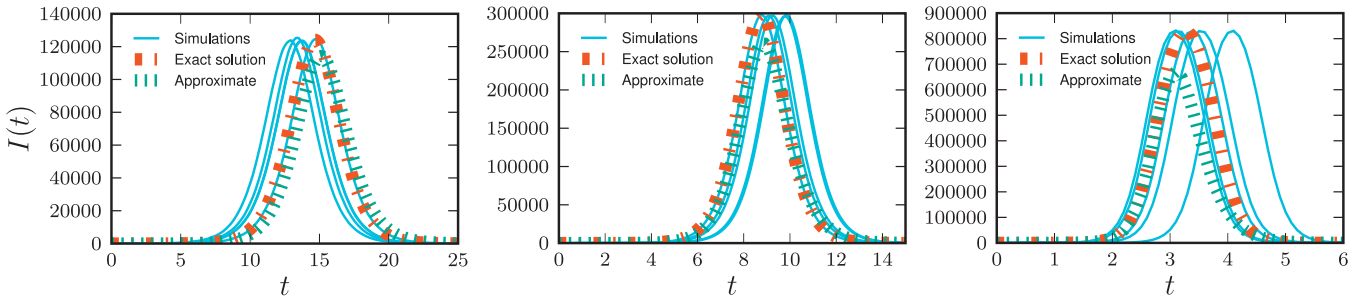

Fig. 6. For gamma distributed recovery time corresponding to an average duration of 1 with 100 exponentially distributed intermediate phases. Infectiousness is constant $\beta = 1.5$, 2, and 5 in a population of size 10^6. the exact system differs from simulations only in time shifts. The approximation closely matches the initial growth phase, but begins deviating close to the peak.

approximate solution (10)–(13) is due to differences in initial conditions: the exact calculation assumes a single infection beginning at $t = 0$ while the approximate solution assumes that the epidemic begins with the equilibrium age-distribution already reached by $t = 0$. The approximate model is a good fit for the behavior at early times and remains a good approximation until the change in N/S becomes significant over the duration of an infection. The approximation performs best in those situations where the number of infections remains smaller.

If we attempt to approximate the epidemic course using the standard SIR model (1)–(3), then we have two free parameters, β and γ. We can identify (at least) three constraints: $\mathscr{R}_0$, average duration of infection, and average generation interval. We can only match two of these at a time, which we show in Fig. 5. If we choose to match $\mathscr{R}_0$ and average duration of infection then the total number of infected person-days is correct, but the timing is far off. If we choose to match $\mathscr{R}_0$ and average generation interval, then the timing is much closer, but the peak patient load is significantly underestimated.

3.2.3. Gamma-distributed recovery times

Recently Wearing et al. (2005) investigated some of the role the distribution of infection duration has on the dynamics of an epidemic. They considered a gamma-distributed infectious period with constant infectiousness. The model they studied corresponds to a chain of 100 exponentially distributed infectious classes, each with infectiousness β and expected duration $\frac{1}{100}$. They showed that the standard SIR equations (1)–(3) provide a poor approximation.

For this system, $P_{rec}(\tau) = \tau^{n-1}\exp(-n\tau)n^n/(n-1)!$ where $n = 100$. The Laplace transform of this is $(1+\phi/n)^{-n}$. From this we can derive the transforms of P and βP, which allows us to define the coefficients for our approximation. Fig. 6 shows that the approximation closely follows the early growth even after the exponential phase ends. It finally deviates close to the epidemic peak, but it gives a reasonable estimate of the timing and maximum load of the epidemic.

3.3. Comment

The approximation we have developed takes into account details about the distribution of generation interval and infection duration, and has nonconstant coefficients. We might alternately try to use a chain progression model where infected individuals travel through a sequence of infectious compartments in such a way that each compartment has its own (fixed) infection and recovery rates. This gives a system of constant coefficient equations.

We saw above that using a single infectious compartment does not give a good fit to the true dynamics. At first glance, with careful choice of the recovery and infection rates for two infectious compartments, we could match $\mathscr{R}_0$, the average duration of infection, and two moments of the generation interval distribution. However, such an approach sometimes fails to give non-negative (or even real) coefficients, resulting in a system of equations whose physical interpretation is difficult or impossible.

4. Transition phase

We have shown that stochastic effects play an important role on whether an epidemic occurs and the timing of an epidemic if it does occur. We have also seen that once the epidemic is sufficiently large, it follows the deterministic predictions. We borrow an approach from Gillespie (2000) to identify when the transition from the stochastic phase to the deterministic phase occurs. For simplicity in our analysis, we will assume that the generation interval distribution is not highly peaked. This allows us to assume that $i(t,\tau)/I(t)$ is close to its equilibrium state.

In order to treat the dynamics as deterministic over a time interval Δt, we must satisfy two competing conditions. First, we need the time interval to be large enough that the number of infections and recoveries that happen in that interval is well approximated by the expected number. That is, we need the expected error to be small compared to the expected value, and so the *coefficient of variation* (the square root of the variance divided by the expectation) is small. Assuming that the rates remain constant, the infection and recovery processes are both Poisson, and so their difference is a Skellam distribution, which has variance $I\Delta t(\hat{\beta}+\hat{\gamma})$ (Skellam, 1946; Johnson et al., 2005). Consequently the condition we need is that $\sqrt{I\Delta t(\hat{\beta}+\hat{\gamma})}/I\Delta t|\hat{\beta}-\hat{\gamma}| \ll 1$. So

$$\Delta t \gg \frac{\hat{\beta}+\hat{\gamma}}{I(\hat{\beta}-\hat{\gamma})^2} \tag{17}$$

Second, we need the time interval to be small enough that the rate at which the infectious population size changes is not affected by changes in the infectious population. That is we need $\Delta I \approx (\beta-\gamma)I\Delta t \ll I$. So

$$\Delta t \ll \frac{1}{|\hat{\beta}-\hat{\gamma}|} \tag{18}$$

For small values of I, conditions (17) and (18) cannot be satisfied simultaneously.

Combining these conditions we need that

$$I \gg \frac{\hat{\beta}+\hat{\gamma}}{|\hat{\beta}-\hat{\gamma}|}$$

More strictly, we actually require that $\sqrt{I} \gg (\hat{\beta}+\hat{\gamma})/|\hat{\beta}-\hat{\gamma}|$.

The analysis we have done does not apply close to the peak of the epidemic (where $\hat{\beta}=\hat{\gamma}$). Here we can replace condition (17) with the requirement that the error in the number of new infections is small compared to the number of new infections and similarly for the number of recoveries. In general we need condition (18) combined with either this pair of conditions or condition (17) to guarantee that the deterministic equations apply. For practical purposes, once the deterministic equations hold, we expect them to hold through the peak until I decays at which point we can use Eq. (17) again.

If the generation interval distribution were highly peaked around some typical time, then we could still argue that the system is deterministic, but we would have to explicitly set the history of ξ rather than assuming that it takes the equilibrium form. By assuming the equilibrium distribution we can treat infections as occurring at a slowly changing rate.

5. Discussion

A typical disease outbreak begins small and whether it grows or becomes extinct is strongly influenced by stochastic effects. If it grows, it generally does so faster than predicted deterministically because those outbreaks which are most likely to not die out stochastically are those which initially grow faster than average. Consequently if we observe an epidemic, it is likely to have grown to an epidemic faster than deterministic equations predict.

Once an outbreak becomes large, it transitions to a deterministic phase. We can estimate the size an outbreak must reach to be deterministic by identifying a time interval which is large enough that many events happen in the interval (and so the error of a deterministic prediction is small compared to the prediction), while at the same time the interval is small enough that the size of I and S does not change significantly. Such a time interval can only exist if I is sufficiently large.

Once an outbreak is deterministic, we can use the deterministic equations to accurately model the spread once a correcting time shift is applied. These equations are somewhat difficult because they require saving the history of an epidemic, and so it may be more convenient to use approximate models. We have introduced an approximate model of the same form as the standard SIR equations. We assume that the system responds quasistatically to changes in the susceptible fraction. It uses a single infectious class, but has coefficients that change in time. It provides a good estimate of the early behavior, but may deviate close to the peak. We can estimate when it deviates by looking at how quickly the susceptible fraction changes during a typical infectious period.

We have assumed throughout that the infectious population can be modeled in continuous time. If the generation interval is discrete, then these assumptions fail, but similar approaches work in discrete time. A more complicated situation arises when the generation interval distribution is close to discrete: if the distribution is tightly peaked about a mean which is sufficiently far from zero, then it may take many generations for the infectious population to reach equilibrium. The dynamics may become deterministic before the age distribution of the infected population reaches equilibrium, in which case our exact equations will provide a good model (assuming appropriate initial conditions) while our approximation may fail badly.

The models we have developed are straightforward to adapt to SIR with birth or death, SIS, or SIRS. In fact, such situations may be more amenable to our approximating method because the rate of change of N/S is reduced.

Acknowledgments

BP would like to acknowledge the support of Canadian Institutes of Health Research (CIHR) (Grants nos. PTL-93146, PAP-93425 and MOP-81273) and the Michael Smith Foundation for Health Research (MSFHR) (Senior Scholar Funds). JCM, BD, and RM were supported by these grants.

JCM was additionally supported by the RAPIDD Program of the Science & Technology Directorate, Department of Homeland Security and the Fogarty International Center, National Institutes of Health.

Appendix A. Probability generating functions

A probability generating function (pgf) is a function $f(x)$ which encodes a probability distribution of non-negative integers (Herbert, 2005). Given that the probability of k is p_k we define the function

$$f(x) = p_0 + p_1 x^1 + p_2 x^2 + \cdots$$

Probability generating functions have a number of useful properties. The product of two pgfs is itself a pgf for the sum of two numbers chosen independently from each distribution. From this fact, it can be shown that for two pgfs f and g encoding the distributions P_g and P_f respectively, the function $f(g(x))$ is the pgf for the distribution found by choosing a random number s from P_f, and then taking the sum of s independent random numbers from P_g.

This property of function composition is useful in our context to deal with taking a random number of infected (corresponding to P_f), and each of them infects a random number of susceptibles (from a distribution P_g). The resulting number of new infections is given by the composition of the corresponding pgfs.

A.1. Derivation of equations

We assume that the population is sufficiently large relative to the number of infections, that no infections are prevented by the depletion of susceptibles. We focus our attention on a single infected individual u and its descendants. We can assume that $t=0$ when u becomes infected. Let $f(x,t)$ be the time-dependent pgf for the number of individuals (descended from u, including u) who are infected at t. That is $f(x,t)=\sum_{n=0}^{\infty} p_n(t)x^n$ where $p_n(t)$ is the probability that n individuals are infected at time t.

Let $g(x,t|\tau)$ be the pgf for the number of infectious descendants u has t units of time after becoming infected given that its infection lasts τ units of time. Note that if $\tau > t$, then $g(x,t|\tau)=g(x,t|t)$. Then the number of current infections is given by a weighted average of the number of descendants (plus 1 if u is still infectious). Encoding this as a statement for pgfs gives

$$f(x,t) = xP(t)g(x,t|t) + \int_0^t g(x,t|\tau)P_{rec}(\tau)\,\mathrm{d}\tau$$

The number of infections resulting from an individual v infected at time θ has pgf $f(x,t-\theta)$. This allows us to express g in terms of f.

To find g, we consider an individual who recovers at time τ and divide the duration of infectiousness into small $\Delta\theta$ sized blocks. The pgf for the number of infections at time t due to an infection that occurs in the interval $[\theta,\theta+\Delta\theta)$ is $f(x,t-\theta)+\mathcal{O}(\Delta\theta)$.

The infection occurs with probability $\beta(\theta)\Delta\theta + \mathcal{O}(\Delta\theta^2)$. The probability that infection does not occur during that time period is $1-\beta(\theta)\Delta\theta + \mathcal{O}(\Delta\theta^2)$. Consequently the pgf for the number of infections at time t resulting from infections in the time interval of interest is

$$1-[1-f(x,t-\theta)]\beta(\theta)\Delta\theta + \mathcal{O}(\Delta\theta^2)$$

The pgf for the number of infections occurring in any of the time intervals is the product of the individual generating functions. Consequently, taking $\Delta\theta\to 0$, the pgf for the number of descendants an individual has at time t given that it recovers at $\tau \le t$ is

$$\begin{aligned} g(x,t|\tau) &= \lim_{\Delta\theta\to 0}\prod_{i=0}^{\tau/\Delta\theta}(1-[1-f(x,t-i\Delta\theta)]\beta(i\Delta\theta)\Delta\theta + \mathcal{O}(\Delta\theta^2)) \\ &= \lim_{\Delta\theta\to 0}\exp\left(\sum_{i=0}^{\tau/\Delta\theta}\ln(1+[f(x,t-i\Delta\theta)-1]\beta(i\Delta\theta)\Delta\theta + \mathcal{O}(\Delta\theta^2))\right) \\ &= \lim_{\Delta\theta\to 0}\exp\left(\sum_{i=0}^{\tau/\Delta\theta}[f(x,t-i\Delta\theta)-1]\beta(i\Delta\theta)\Delta\theta + \mathcal{O}(\Delta\theta^2)\right) \\ &= \exp\left(\int_0^\tau [f(x,t-\theta)-1]\beta(\theta)\,\mathrm{d}\theta\right) \end{aligned}$$

If the individual recovers at time $\tau > t$, then the pgf for the number of descendants at time t including itself satisfies $g(x,t|\tau) = xg(x,t|t)$.

This expression for g can be derived alternately by considering a large population size N and noting that if the expected number of infections caused by v is $r=\int_0^\tau \beta(\theta)\,\mathrm{d}\theta$, then the probability of infecting each individual is $p=\int_0^\tau \beta(\theta)/N\,\mathrm{d}\theta$. The probability of infecting n people is then $\binom{N}{n}p^n(1-p)^{N-n}$. From this we can derive the pgf for the number of infections caused directly from v, and then using function composition we arrive at the same expression.

Appendix B. Notes on the numerics for the stochastic problem

We take $f(x,t)=\sum p_k(t)x^k$ and $g(x,t|\tau)=\sum q_k(t|\tau)x^k$ where p_k gives the probability of having k people infected at time t, while q_k gives the probability of having k descendants given that recovery occurs at time τ. If we take k derivatives of these equations, divide by $k!$ and evaluate at $x=0$, we get the probability of k infections. The resulting system of equations is straightforward to solve numerically. As our initial condition at $t=0$ we generally set all derivatives of f to be 0 except the first derivative, which is 1, though other options are possible.

If we make a simplifying assumption that β is constant, we can find an expression for g which reduces the dimensionality of the problem. We have

$$\begin{aligned} \int_0^\tau [f(x,t-\theta)-1]\beta\,\mathrm{d}\theta &= \beta\left[-\tau + \int_0^t f(x,t-\theta)\,\mathrm{d}\theta - \int_\tau^t f(x,t-\theta)\,\mathrm{d}\theta\right] \\ &= -\beta\tau + \beta\left[\int_0^t f(x,\theta)\,\mathrm{d}\theta - \int_0^{t-\tau} f(x,\theta)\,\mathrm{d}\theta\right] \end{aligned}$$

We define the auxiliary function $\zeta(x,s)=\int_0^s f(x,\theta)\,\mathrm{d}\theta$. Then $g(x,t|\tau)=\exp\beta[\zeta(x,t)-\zeta(x,t-\tau)-\tau]$

Our equation for f remains

$$f(x,t) = xP(t)g(x,t|t) + \int_0^t g(x,t|\tau)P_{rec}(\tau)\,\mathrm{d}\tau$$

This allows us to simplify the calculations by storing ζ at each value of s rather than needing to integrate f at each time step.

In practice we want to find arbitrary derivatives of f evaluated at $x=0$. To find this numerically, we differentiate these equations with respect to x to arrive at equations coupling derivatives of $f(x,t)$ with derivatives of ζ at $x=0$. Let us assume we know $\zeta(0,s)$ and its derivatives for $s=0,\mathrm{d}t,2\mathrm{d}t,\ldots,t$ and $f(0,t)$ and its derivatives. To find $\zeta(0,t+\mathrm{d}t)$ and $f(0,t+\mathrm{d}t)$, it is straightforward to use an implicit numerical method.

Appendix C. An equivalent formulation

Although Eqs. (6)–(9) are intuitively appealing because of their similarity to the standard SIR equations, we can reduce them to a simpler form. We first replace $e^{\xi(t)}$ with $\psi(t)$. Note that $\dot{S}=-b(t)=-C\psi(t)$. Further $\mathscr{G}=I(t)/C$, so from the initial condition at $t=0$, we can calculate C, and have no further need for g. Thus we arrive at

$$\dot{S} = -C\psi(t) \tag{19}$$

$$\dot{I} = C\psi(t) - C\int_0^\infty \psi(t-\tau)P_{rec}(\tau)\,\mathrm{d}\tau \tag{20}$$

$$\dot{R} = C\int_0^\infty \psi(t-\tau)P_{rec}(\tau)\,\mathrm{d}\tau \tag{21}$$

$$\mathscr{F}[\psi,t] = \frac{N}{S}\psi(t) \tag{22}$$

If we take as the initial condition that all infections at time $t=0$ begin their infection period at $t=0$, then $\psi(t-\tau)=0$ for $\tau>t$ and we can assume that the integrals have their upper limit at $\tau=t$. If we take some other initial condition, we may have to include the entire range of τ. Although these equations are simpler to solve, they lose some of their intuitive appeal because it is more difficult to identify the meaning of each term.

References

Anderson, D., Watson, R., 1980. On the spread of a disease with gamma distributed latent and infectious periods. Biometrika 67 (1), 191–198.

Brauer, F., 2008. Compartmental models in epidemiology. In: Lecture Notes in Mathematics, Springer, Berlin, vol. 1945, p. 19.

Brauer, F., 2005. Age of infection in epidemiology models. In: Electronic Journal of Differential Equations, Conference, vol. 12, pp. 29–37.

Breban, R., Vardavas, R., Blower, S., 2005. Linking population-level models with growing networks: a class of epidemic models. Physical Review E 72 (4), 46110.

Castillo-Chavez, C., Cooke, K., Huang, W., Levin, S.A., 1989. On the role of long incubation periods in the dynamics of acquired immunodeficiency syndrome (AIDS). Journal of Mathematical Biology 27 (4), 373–398.

Crump, K.S., Mode, C.J., 1968. A general age-dependent branching process I. Journal of Mathematical Analysis and Applications 24 (3), 494–508.

Crump, K.S., Mode, C.J., 1969. A general age-dependent branching process II. Journal of Mathematical Analysis and Applications 25 (1), 8–17.

Feller, F., 1941. On the integral equation of renewal theory. The Annals of Mathematical Statistics 12 (3), 243–267.

Gillespie, D.T., 2000. The chemical Langevin equation. The Journal of Chemical Physics 113, 297.

Gunther, O.P., Ogilvie, G., Naus, M., Young, E., Patrick, D.M., Dobson, S., Duval, B., Noel, P.A., Marra, F., Miller, D., et al., 2008. Protecting the next generation: what is the role of the duration of human papillomavirus vaccine-related immunity?. The Journal of Infectious Diseases 197 (12), 1653–1661.

Herbert, S.W., 2005. Generatingfunctionology, third ed. A K Peters, Ltd, 2005.

Haccou, P., Jagers, P., Vatutin, V.A., 2005. Branching Processes: Variation, Growth, and Extinction of Populations. Cambridge University Press, Cambridge.

Hethcote, H.W., van den Driessche, P., 2000. Two SIS epidemiologic models with delays. Journal of Mathematical Biology 40 (1), 3–26.

Jagers, P., 1975. Branching Processes with Biological Applications. Wiley, New York.

Johnson, N.L., Kotz, S., Kemp, A.W., 2005. Univariate Discrete Distributions. Wiley/Interscience, New York.

Li, J., Brauer, F., 2008. Continuous-time age-structured models in population dynamics and epidemiology. In: Lecture Notes in Mathematics, vol. 1945. Springer, Berlin, p. 205.

Lloyd, A.L., 2001a. Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. Proceedings of the Royal Society B: Biological Sciences 268 (1470), 985–993.

Lloyd, A.L., 2001b. Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics. Theoretical Population Biology 60 (1), 59–71.

Ma, J.J., Earn, D.J.D., 2006. Generality of the final size formula for an epidemic of a newly invading infectious disease. Bulletin of Mathematical Biology 68 (3), 679–702.

Skellam, J.G., 1946. The frequency distribution of the difference between two Poisson variates belonging to different populations. Journal of the Royal Statistical Society, 296–296.

Svensson, Å., 2007. A note on generation times in epidemic models. Mathematical Biosciences 208 (1), 300–311.

Thieme, H.R., Castillo-Chavez, C., 1993. How may infection-age-dependent infectivity affect the dynamics of HIV/AIDS?. SIAM Journal on Applied Mathematics 53, 1447.

Wallinga, J., Lipsitch, M., 2007. How generation intervals shape the relationship between growth rates and reproductive numbers. Proceedings of the Royal Society B: Biological Sciences 274 (1609), 599–604.

Wearing, H.J., Rohani, P., Keeling, M.J., 2005. Appropriate models for the management of infectious diseases. PLoS Medicine 2 (7), e174.