ORIGINAL ARTICLE

Generality of the Final Size Formula for an Epidemic of a Newly Invading Infectious Disease

Junling Ma*, David J.D. Earn

Department of Mathematics & Statistics, McMaster University, Hamilton, ON, Canada L8S 4K1

Abstract The well-known formula for the final size of an epidemic was published by Kermack and McKendrick in 1927. Their analysis was based on a simple susceptible-infected-recovered (SIR) model that assumes exponentially distributed infectious periods. More recent analyses have established that the standard final size formula is valid regardless of the distribution of infectious periods, but that it fails to be correct in the presence of certain kinds of heterogeneous mixing (e.g., if there is a core group, as for sexually transmitted diseases). We review previous work and establish more general conditions under which Kermack and McKendrick's formula is valid. We show that the final size formula is unchanged if there is a latent stage, any number of distinct infectious stages and/or a stage during which infectives are isolated (the durations of each stage can be drawn from any integrable distribution). We also consider the possibility that the transmission rates of infectious individuals are arbitrarily distributed—allowing, in particular, for the existence of super-spreaders—and prove that this potential complexity has no impact on the final size formula. Finally, we show that the final size formula is unchanged even for a general class of spatial contact structures. We conclude that whenever a new respiratory pathogen emerges, an estimate of the expected magnitude of the epidemic can be made as soon the basic reproduction number $\mathcal{R}_0$ can be approximated, and this estimate is likely to be improved only by more accurate estimates of $\mathcal{R}_0$, not by knowledge of any other epidemiological details.

Keywords Epidemic models · Final size · Arbitrary stage durations · Integro-differential equations 1991 MSC:92D30

1. Introduction

Whenever a serious infectious disease emerges or re-emerges in a human population, a matter of immediate interest is the likely magnitude of the outbreak. This

---

*Corresponding author.
E-mail addresses: junlingm@math.mcmaster.ca (Junling Ma), earn@math.mcmaster.ca (David J.D. Earn).

is often called the expected *final size* of the epidemic (Bailey, 1975; Anderson and Watsons, 1980; Anderson and May, 1991; Andersson and Britton, 2000; Diekmann and Heesterbeek, 2000), which we denote Z. Of course, if the emergent pathogen becomes endemic in the population then there will never really be a "final" size, but in that case the matter at issue is the likely magnitude of the first major wave of cases.

To our knowledge, a formal, mathematical argument leading to an estimate of an epidemic's expected final size Z was first published in the landmark 1927 paper by Kermack and McKendrick (1927). The formula for Z that Kermack and McKendrick obtained [Eq. (5) below] depends only on the basic reproduction number, $\mathcal{R}_0$ (the expected number of secondary cases caused by a typical primary case in a fully susceptible population). However, without further analysis, it is unclear to what extent this formula depends on the particular assumptions that Kermack and McKendrick made in constructing their model, and hence whether it bears a strong relationship to the final size in realistic situations.

One implicit assumption of the Kermack and McKendrick (1927) analysis is that infectious periods are exponentially distributed. Anderson and Watsons (1980) showed that the final size formula remains valid if infectious periods follow a Gamma distribution. In Section 1.3 of their book, Diekmann and Heesterbeek (2000) generalized this statement to cover an arbitrary distribution of infectious periods.

Another key assumption of the Kermack and McKendrick (1927) analysis is that the host population is homogeneously mixed. It is well known that if this assumption is dropped then the final size will not necessarily be given by the standard formula. In particular, the existence of a core group, as for sexually transmitted diseases, gives rise to a different formula for the final size (Anderson and May, 1991; Diekmann and Heesterbeek, 2000).

In this paper, we explore the generality of the standard Kermack-McKendrick final size formula (5). We begin with a pedagogical review of the main results that have been obtained previously, adding model structure in steps. We then proceed to generalize these results in three new directions, showing that the standard formula remains valid (i) regardless of the number of distinct infectious stages, (ii) if the mean contact rate is itself arbitrarily distributed and (iii) for a large class of spatially heterogeneous contact structures. We conclude that the Kermack-McKendrick formula (5) applies in great generality, making it an extremely useful relationship in practice.

2. The final size in the standard SIR model

The standard SIR model (Anderson and May, 1991; Diekmann and Heesterbeek, 2000; Brauer and Castillo-Chavez, 2001) is represented by a system of three ordinary differential equations,

$$\frac{dS}{dt} = -\beta SI, \tag{1a}$$

$$\frac{dI}{dt} = \beta SI - \gamma I, \tag{1b}$$

$$\frac{dR}{dt} = \gamma I. \tag{1c}$$

Here, S, I, and R denote the proportions of the population that are susceptible, infectious and recovered, respectively. Recovered individuals are assumed to be immune to reinfection. Since $S + I + R = 1$, one of the three equations is redundant. The two parameters in the model are the transmission rate (β) and the recovery rate (γ). The mean infectious period is $T = 1/\gamma$. The basic reproduction number is $\mathcal{R}_0 = \beta T$. The proportion of the population that is infected can increase, and hence an epidemic can occur, if and only if $\mathcal{R}_0 > 1$.

In Eqs. (1), recruitment of new susceptibles through birth or immigration is ignored, as is loss of individuals through mortality or emmigration. This approximation is reasonable provided the timescale of the epidemic is much shorter than the timescale over which demographic turnover occurs.

No recruitment ensures that the disease will eventually burn out, i.e., $I(\infty) = 0$. To see formally that we must have $I(\infty) = 0$, note that the positive orthant is invariant so all solutions of Eqs. (1) lie in the non-negative, bounded set defined by $S, I, R \geq 0$ and $S + I + R = 1$. Observing that

$$\frac{d}{dt}(S + I) = -\gamma I, \tag{2}$$

we see that $S + I$ is decreasing whenever $I > 0$. However, $S + I$ is bounded below by 0; hence it has a limit. Moreover, Eq. (2) implies that $\frac{d}{dt}(S + I)$ is bounded because I is bounded. Hence $\lim_{t \to \infty} \frac{d}{dt}(S + I) = 0$, so Eq. (2) yields $I(\infty) = 0$.

This model (1) is sufficiently simple that an exact solution can be obtained for the phase portrait. Forming dI/dS and integrating yields

$$I = I(0) + S(0) - S + \frac{1}{\mathcal{R}_0} \ln [S/S(0)]. \tag{3}$$

An epidemic ends when no infectives remain. Consequently, we can find the final size Z by setting $I = 0$ in Eq. (3) and solving for $Z = S(0) - S$. This yields the implicit relation

$$Z = S(0)\left(1 - e^{-\mathcal{R}_0[Z + I(0)]}\right) \tag{4}$$

(an explicit form is given in Appendix A).

The general formula (4) applies regardless of the proportions of the population that are susceptible and infective initially. A less than fully susceptible population must be taken into account if some individuals have been vaccinated or retain immunity from previous exposures. However, in the important special case where a new pathogen enters a fully susceptible population, we have both $I(0) \ll 1$ and $S(0) \sim 1$. In the limit $I(0) \to 0$, $S(0) \to 1$,

$$Z = 1 - e^{-\mathcal{R}_0 Z}, \tag{5}$$

which is the usual final size formula.

The practical utility of the final size formula would appear to be limited by the fact that it is derived from the overly simplistic standard SIR model. In the remainder of this paper we show that, in fact, the formula is valid under very general circumstances. We generalize the model in several steps, in order to explain the ideas clearly and to avoid exposing the reader to a notational burden at the outset.

3. Latency and multiple infectious stages

The standard SIR model ignores any delay between the time an individual is infected and the onset of infectiousness. This *latent period* is often substantial compared with the infectious period (Anderson and May, 1991). In addition, some pathogens (notably HIV) give rise to a sequence of distinct stages of infection, each yielding a different transmission rate (Redfield et al., 1986; Seligmann et al., 1987).

It is common to incorporate latency by including an "exposed" class (E), yielding an SEIR model (Schwartz and Smith, 1983; Earn et al., 2000a). However, noting that the exposed stage can be regarded as an infectious stage during which the transmission rate happens to be zero, we lose no generality by restricting attention to multiple infectious stage models, which we refer to as SI^nR models if there are n infectious stages.

Generalizing the SIR model (1) to include multiple infectious stages, the equations for an SI^nR are

$$\dot{S} = -S \sum_{j=1}^{n} \beta_j I_j, \tag{6a}$$

$$\dot{I}_1 = S \sum_{j=1}^{n} \beta_j I_j - \gamma_1 I_1, \tag{6b}$$

$$\dot{I}_i = \gamma_{i-1} I_{i-1} - \gamma_i I_i, \qquad i = 2, \ldots, n, \tag{6c}$$

$$\dot{R} = \gamma_n I_n, \tag{6d}$$

where I_i is the density of individuals in the ith infectious stage, β_i is the transmission rate for contact with individuals in this stage, and $\frac{1}{\gamma_i}$ is the mean duration of this stage.

The sum of Eqs. (6) is zero, implying that $S + \sum_{i=1}^{n} I_i + R$ is invariant (and hence that Eq. (6d) is superfluous, as in the basic SIR model). To see that the final size is well-defined in this model, we generalize the argument given for the simple SIR model in Section 2. Let $J_k = S + \sum_{i=1}^{k} I_i$. Then $\frac{d}{dt} J_k = -\gamma_k I_k$, so $I_k > 0$ implies J_k is decreasing and $\lim_{t \to \infty} \frac{d}{dt} J_k = 0$. Hence $I_k(\infty) = 0$ for all k.

Before a new disease is introduced into a population, everyone is susceptible, i.e., $S = 1$, $I_1 = \cdots = I_n = 0$ and $R = 0$, which is known as the *disease free equilibrium* (DFE). The stability of the DFE determines whether the

new disease can cause an epidemic. Hyman et al. (1999) proved that the quantity

$$\mathcal{R}_0 = \sum_{i=0}^{n} \frac{\beta_i}{\gamma_i}, \tag{7}$$

determines the stability of the DFE, i.e., the DFE is locally stable if $\mathcal{R}_0 < 1$ and unstable if $\mathcal{R}_0 > 1$. Note $\mathcal{R}_0$ as defined by (7) is the sum of the products of the transmission rate and mean duration in each infectious stage, i.e., it is the basic reproduction number for this model.

We now proceed to show that Eqs. (6) yield the same final size formula as (5). Anderson and Watsons (1980) established this for the special case in which each stage has the same transmission rate and duration ($\beta_i = \beta$ and $\gamma_i = \gamma$ for all i). Their method can easily be generalized to arbitrary β_i and γ_i as follows:

Let $G_k = \sum_{i=k+1}^{n} I_i + R$, $k = 1, \ldots, n-1$ and $G_n = R$. Then Eqs. (6c) and (6d) imply $\frac{d}{dt} G_k = \gamma_k I_k$ for each $k = 1, \ldots, n$. Consequently, Eq. (6a) implies that the function

$$F(t) = \log S(t) + \sum_{k=1}^{n} \frac{\beta_k}{\gamma_k} G_k(t) \tag{8}$$

is a constant of the motion. For a newly invading disease, $S(0) \to 1$, $I_k(0) \to 0$ and $R(0) \to 0$, so $F(0) = 0$. On the other hand, $I_k(\infty) = 0$ implies $G_k(\infty) = R(\infty)$ for all k. Hence,

$$F(\infty) = \log S(\infty) + \sum_{k=1}^{n} \frac{\beta_k}{\gamma_k} R(\infty) = F(0) = 0. \tag{9}$$

Recalling that $Z = R(\infty) = 1 - S(\infty)$, and Eq. (7), the final size formula Eq. (5) immediately follows.

4. Gamma-distributed stage durations

Consideration of a special case of the model in the previous section allows us to infer that the distribution of stage durations in any infectious stage need not be exponentially distributed. The final size formula (5) holds true if the stage durations follow a Gamma distribution, which has probability density

$$g_{k,\phi}(t) = \frac{\phi^k}{\Gamma(k)} e^{-\phi t} t^{k-1}, \qquad t > 0. \tag{10}$$

Here, the *shape parameter* k is a positive integer and the *scale parameter* $\phi > 0$. The mean is k/ϕ and the variance is k/ϕ^2. This family of distributions includes the exponential distribution ($k = 1$), nearly normal distributions (k large) and the Delta distribution, which yields a fixed duration ($k \to \infty$).

The term $-\gamma I$ in Eq. (1b) for the simple SIR model implies that the durations of infectious periods are exponentially distributed with mean $1/\gamma$ (Brauer and Castillo-Chavez, 2001). Since the sum of exponentially distributed random variables is Gamma distributed, a standard trick (Anderson and Watsons, 1980; Lloyd, 2001a,b) enables us to replace an exponential distribution with mean $1/\gamma$ by a Gamma distribution $g_{k,k\gamma}$ (which also has mean $1/\gamma$). The trick is to replace a single infectious stage with k identical exponentially distributed substages of mean duration $1/(k\gamma)$. The model then has precisely the form the SI^{n}R model (6), but with the sequence of infectious stages being mathematical artefacts rather than biologically meaningful.

Since this substage trick can be applied equally well to any infectious stage, Anderson and Watson's (1980) conclusion that the final size in an SIR model with Gamma distributed infectious periods is given by the usual formula (5) now generalizes to an arbitrary number of stages, each with Gamma distributed durations.

5. Kermack and McKendrick's special case

Although expressed differently, in their original paper Kermack and McKendrick (1927) studied a particular limit of the Gamma distributed multiple stage model described in the previous section.

Suppose we have a sequence of n infectious stages, each with the same, fixed stage duration τ. Thus, this is the limit $k \to \infty$ in the Gamma distribution (10), with $\phi = 1/\tau$. During stage i, the transmission rate is β_i $(i = 1, \ldots, n)$. If we now let the number of infectious stages increase $(n \to \infty)$ and the length of each stage decrease $(\tau \to 0)$, keeping the total infectious period $(T = n\tau)$ constant, then we arrive at a model in which the transmission rate (β) varies continuously through the infectious period. In such a model, an individual's infectivity depends on his/her *stage-age*, i.e., the amount of time since initial infection.

Using a stage-age SIR formulation, Kermack and McKendrick (1927) derived the final size formula (5) and showed that its form is independent of how the transmission rate depends on stage-age.

6. Arbitrarily distributed stage durations

We have seen that the final size formula (5) is the same if stage durations are distributed according to any member of the family of the Gamma distributions (10). This suggests that *any* distribution of stage durations will yield the same final size.

Unlike the situation for Gamma distributed infectious periods, with an arbitrary distribution the time-evolution of the infectious class can no longer be expressed using ordinary differential equations (ODEs). Instead, we must construct a system of integro-differential equations (Feng and Thieme, 2000). Let $U(t)$ be the rate at which individuals become infectious (enter class I) at time $t \geq 0$, i.e.,

$$U(t) = \beta S(t) I(t). \tag{11a}$$

The equation for the susceptible class is the same as in the standard SIR model,

$$\frac{dS}{dt} = -\beta SI. \tag{11b}$$

We also define $U(t)$ for $t < 0$ since this will allow us to specify the initial conditions. For $t < 0$, $U(t)$ is the rate at which individuals who were still infectious at the initial time (time 0) became infectious at time t. Hence, the number of infectious individuals at time $t = 0$ is

$$I(0) = \int_{-\infty}^{0} U(t)\, dt. \tag{11c}$$

Let $f(t)$ be the probability density of the infectious period (so, for the special case of a Gamma distribution, $f(t)$ is given by $g_{k,\phi}(t)$ in Eq. 10). For an individual who becomes infectious at time $\tau \geq 0$, the probability density that s/he recovers (leaves the I class) at time $t > \tau$ is

$$l(\tau, t) = f(t - \tau), \qquad \text{if } \tau \geq 0. \tag{11d}$$

However, for individuals who became infectious at time $\tau < 0$, the probability density to leave class I at time t must be conditioned by still being infectious at time 0,

$$l(\tau, t) = \frac{f(t - \tau)}{\int_{-\tau}^{\infty} f(s)\, ds}, \qquad \text{if } \tau < 0. \tag{11e}$$

Hence, the rate of change of the number of infectious individuals at time t is

$$\frac{dI}{dt} = U(t) - \int_{0}^{t} f(t - \tau)U(\tau)\, d\tau - \int_{-\infty}^{0} U(\tau)l(\tau, t)\, d\tau. \tag{11f}$$

Those who leave class I enter class R, hence

$$\frac{dR}{dt} = \int_{0}^{t} f(t - \tau)U(\tau)\, d\tau + \int_{-\infty}^{0} U(\tau)l(\tau, t)\, d\tau. \tag{11g}$$

In Appendix B, we show that if the infectious period is exponentially distributed then Eqs. (11) reduce to the the standard SIR model (1).

As in the standard SIR model (1), the population size is invariant ($S + I + R = 1$), $S \geq 0$, and $S(\infty)$ exists. Establishing that $I(\infty) = 0$ (so the disease will burn out regardless of the infectious period distribution) requires a little more work. Since $f(t)$ and $l(\tau, t)$ are probability densities, $\int_{0}^{\infty} f(t)\, dt = \int_{0}^{\infty} l(\tau, t)\, dt = 1$; consequently, switching the order of integrations, we find

$$\int_{0}^{\infty} \int_{0}^{t} f(t - \tau)U(\tau)\, d\tau\, dt = \int_{0}^{\infty} U(t)\, dt$$

and

$$\int_0^\infty \int_{-\infty}^0 l(\tau, t) U(\tau) \, d\tau \, dt = \int_{-\infty}^0 U(\tau) \, d\tau = I(0).$$

Next, integrating from 0 to ∞ on both sides of Eq. (11f), we have $I(\infty) = 0$.

To relate the final size in the present model to the usual formula (5), we need an expression for the basic reproduction number $\mathcal{R}_0$. Feng and Thieme (2000) have shown that $\mathcal{R}_0 = \beta T$, where $T = \int_0^\infty t f(t) \, dt$ is the mean infectious period.

We now divide Eq. (11b) by S and integrate, yielding

$$\log \frac{S(\infty)}{S(0)} = -\beta \int_0^\infty I(t) \, dt. \tag{12}$$

Therefore, the usual final size formula (5) will follow if $\int_0^\infty I(t) \, dt = TZ$. This is plausible, intuitively, since the integral $\int_0^\infty I(t) \, dt$ sums all infectious individuals, weighted by the time they are infectious; this must also equal the mean infectious period T times the number of individuals infected over the course of the entire epidemic (the final size Z). More rigorously, in Appendix B we prove

Lemma 6.1. *For the model specified by Eqs. (11), in the limit $I(0) \to 0$,*

$$\int_0^\infty I(t) \, dt = T \int_0^\infty U(t) \, dt.$$

Noting now that

$$\int_0^\infty U(t) \, dt = \int_0^\infty \beta S(t) I(t) \, dt = S(0) - S(\infty) = Z,$$

we have $\int_0^\infty I(t) \, dt = TZ$, as required. Thus, we have

Theorem 6.1. *Consider the SIR model with arbitrary infectious period distribution, specified by Eqs. (11). If the initial number of infectious individuals is small ($I(0) \to 0$), then the final size of the epidemic (Z) is given by the classical formula (5). In this formula, the basic reproduction number is $\mathcal{R}_0 = \beta T$, where T is the mean infectious period.*

7. Arbitrarily distributed durations in multiple stages

With modest additional effort, we can generalize Theorem 6.1 to the multiple stage SI^{n}R model, i.e., the usual final size formula (5) is still valid if there are multiple infectious stages and the durations of each stage are arbitrarily distributed. Recall that latent stages can be considered infectious stages with zero transmission rate, so we do not make any explicit reference to latency. As in all the situations considered

above, we assume that vital dynamics (births and deaths) can be ignored and that recovery entails lifelong immunity.

Suppose the ith infectious stage has duration distribution with probability density $f_i(t)$ and mean T_i. Thus, in the special case that durations in each stage follow Gamma distributions (10), we have $f_i(t) = g_{k_i \phi_i}(t)$ and $T_i = 1/\phi_i$.

The the rate of change of the susceptible class is given by Eq. (6a), as when all stage durations are exponentially distributed. Similar to Section 6, the equation for each infectious stage can be written

$$\dot{I}_i = U_i(t) - \int_{-\infty}^{t} l_i(\tau, t) U_i(\tau) \, d\tau, \qquad i = 1, 2, \ldots, n, \tag{13a}$$

where $U_i(t)$ is the number of individuals entering class I_i at time t, and

$$l_i(\tau, t) = \begin{cases} f_i(t - \tau), & \text{if } \tau \geq 0, \\ \frac{f_i(t-\tau)}{\int_{-\tau}^{\infty} f_i(s) \, ds}, & \text{if } \tau < 0. \end{cases} \tag{13b}$$

Similar to Eq. (11a),

$$U_1(t) = S \sum_{i=1}^{n} \beta_i I_i. \tag{13c}$$

Individuals who leave infectious stage i, $i < n$, enter stage $i + 1$, hence

$$U_{i+1}(t) = \int_{-\infty}^{t} l_i(\tau, t) U_i(\tau). \tag{13d}$$

The initial conditions are specified by

$$I_i(0) = \int_{-\infty}^{0} U_i(t) \, dt. \tag{13e}$$

Individuals who leave the last infectious stage recover (into class R), hence

$$\dot{R} = U_{n+1}. \tag{13f}$$

When the infectious periods and the latent period are exponentially distributed, i.e., $f_i(t) = \gamma_i \, e^{-\gamma_i t}$, the technique introduced in Appendix B can be applied to show that the model (13) is indeed the standard multiple-stage SIR model.

In their analysis of the SI^{n}R model with arbitrarily distributed stage durations (13), Feng and Thieme (2000) have shown that the basic reproduction number is

$$\mathcal{R}_0 = \sum_{i=1}^{n} \beta_i T_i \tag{14}$$

and that the disease free equilibrium is unstable if and only if $\mathcal{R}_0 > 1$.

Similar to previous sections, we find that $S + \sum_i I_i + R = 1$ is invariant, $S \geq 0$, and $S(\infty)$ exists. Integrating Eq. (13a) from 0 to t and and switching the order of integrations, we have $I(t) \geq 0$; similarly, integrating from 0 to ∞, we have $I(\infty) = 0$. Thus, the disease will also eventually burn out.

Theorem 7.1. *Consider the multiple stage SI^nR epidemic model with arbitrarily distributed stage durations, specified by Eqs. (13). If the initial number of infectious individuals is suffciently small, i.e., in the limit $\sum_i I_i(0) \to 0$, the final size Z is given by the unique solution of Eq. (5), or explicitly by Eq. (A.2), in which $\mathcal{R}_0$ is given by Eq. (14).*

Proof. To prove this, we divide by S on both sides of (6a) and integrate with respect to time,

$$\log \frac{S(\infty)}{S(0)} = -\sum_i \beta_i \int_0^\infty I_i \, dt. \tag{15}$$

Using Lemma 6.1, Eq. (15) becomes

$$\log \frac{S(\infty)}{S(0)} = -\sum_i \beta_i T_i \int_0^\infty U_i(t) \, dt. \tag{16}$$

Suppose $I_i(0) = \epsilon_i \ll 1$. Switching the order of integration,

$$\int_0^\infty \int_{-\infty}^0 U_i(\tau) l(\tau, t) \, d\tau \, dt = \int_{-\infty}^0 U_i(\tau) \, d\tau = I_i(0) = \epsilon_i \ll 1.$$

Also, integrating Eq. (13d) and switching the order of integrations, we have

$$\int_0^\infty U_{i+1}(t) \, dt = \int_0^\infty \int_0^t f_i(t - \tau) U_i(\tau) \, d\tau \, dt, \tag{17}$$

$$= \int_0^\infty U_i(\tau) \int_\tau^\infty f_i(t - \tau) \, dt \, d\tau, \tag{18}$$

$$= \int_0^\infty U_i(\tau) \, d\tau, \tag{19}$$

i.e., summed over all time, the number of individuals who enter the $(i + 1)$th stage is the same as the number who enter the ith stage. We then integrate Eq. (6a). From the definition of $U_1(t)$ in Eq. (13c), we have

$$S(\infty) - S(0) = -\int_0^\infty U_1(t) \, dt = -\int_0^\infty U_i(t) \, dt$$

for all $i = 1, \ldots, n$. Notice that the final size $Z = S(0) - S(\infty)$, so Eq. (16) becomes

$$\log\left(1 - \frac{Z}{S(0)}\right) = -\sum_i \beta_i T_i Z.$$

Finally, since $\mathcal{R}_0 = \sum_i \beta_i T_i$ [Eq. (14)], we have

$$\log\left(1 - \frac{Z}{S(0)}\right) = -\mathcal{R}_0 Z.$$

8. Super-spreaders

During the epidemic of severe acute respiratory syndrome (SARS) in 2003 (Poutanen et al., 2003; Low and McGeer, 2003), one aspect of observed transmission that received a great deal of attention was the existence of *super-spreaders* (WHO SARS Update 27, 2003), i.e., individuals who infect many more people than the average (e.g., at least an order of magnitude greater than $\mathcal{R}_0$). In this section, we prove that the existence of the type of super-spreaders that occurred during the SARS outbreak in 2003 has no effect on the final size formula.

One mechanism for the generation of super-spreaders is that the infectious period distribution could be bimodal, such that a small proportion of individuals are infectious much longer than average. This situation is already covered by Theorem 7.1, which deals with arbitrary distributions of stage durations.

In this section, we consider a different type of generalization of the SIR model. In all of the models we have considered thus far, it has been assumed implicitly that all infectious individuals have the same transmission rate. If some individuals actually have a much higher than average transmission rate then they will be super-spreaders. This could arise, for example, because of individual variation in viral load (reflecting variation in immune response) or as the result of chance occurence of environmental circumstances that promote disease spread (as apparently occurred during the SARS epidemic (WHO SARS Update 33, 2003)). Whatever the origin, we focus here on variation in transmission rate that is manifested only after an individual is infected. In the following sections we consider the effects of intrinsic variation in contact rates among individuals, which results from inhomogeneous contact network structure in the host population.

We now proceed to develop and analyze an SIR model with a distribution of transmission rates and show that the usual final size formula (5) applies. Extension of our results to a multiple stage SI^nR model is straightfoward and similar to the way we generalized our results for the SIR with arbitrary infectious period distribution to an SI^nR model with arbitrary stage duration distributions.

Let $I(\beta, t)$ be the distribution (at time t) of infectious individuals that have transmission rate β. Note that we assume implicitly that a given individual retains the same transmission rate throughout his/her infectious period, and hence that

this transmission rate is determined as soon as infection occurs. Let $q(\beta)$ be the probability that a newly infectious individual has transmission rate β. Thus, the mean transmission rate is

$$\langle \beta \rangle = \int_0^\infty q(\beta) \, \mathrm{d}\beta. \tag{20}$$

The rate of change of the proportion of the population that is susceptible is

$$\dot{S} = -S \int_0^\infty \beta I(\beta, t) \, \mathrm{d}\beta. \tag{21a}$$

Upon infection, individuals leave the susceptible class S and enter the infectious class I, a proportion $q(\beta)$ of them having a transmission rate β. Upon recovery, they leave the infectious class at rate γ. Thus,

$$\dot{I}(\beta, t) = q(\beta)S \int_0^\infty \beta I(\beta, t) \, \mathrm{d}\beta - \gamma I(\beta, t). \tag{21b}$$

$$\dot{R} = \gamma \int_0^\infty I(\beta, t) \, \mathrm{d}\beta. \tag{21c}$$

In order to ensure that the model is well posed, we must establish that if $I(\beta, 0) > 0$ then $I(\beta, t) > 0$ for all time. Suppose this is not true. Then there exists $\tilde{\beta}$ and T such that $\int_0^\infty I(\beta, T) \, \mathrm{d}\beta > 0$, $I(\tilde{\beta}, t) > 0$ for $t < T$, but $I(\tilde{\beta}, T) = 0$. Substituting these conditions into Eq. (21b), we obtain $\dot{I}(\tilde{\beta}, T) > 0$. But this contradicts the hypothesis that $I(\tilde{\beta}, t)$ decreases from $I(\tilde{\beta}, 0) > 0$ to $I(\tilde{\beta}, T) = 0$. Hence $I(\beta, t) > 0$ for all $t \geq 0$.

We now check that the disease will always burn out in this model, as expected. Denoting $I_{\text{tot}}(t) = \int_0^\infty I(\beta, t) \, \mathrm{d}\beta$, Eqs. (21) imply

$$\frac{\mathrm{d}}{\mathrm{d}t}(S + I_{\text{tot}}) = -\gamma I_{\text{tot}}, \tag{22}$$

which is strictly negative. Hence, $S + I_{\text{tot}}$ is decreasing and bounded below by zero. Hence $\lim_{t \to \infty} \frac{\mathrm{d}}{\mathrm{d}t}(S + I_{\text{tot}}) = 0$, i.e., $I_{\text{tot}}(\infty) = 0$.

For the model specified by Eqs. (21), the basic reproduction number is

$$\mathcal{R}_0 = \int_0^\infty \frac{\beta}{\gamma} q(\beta) \, \mathrm{d}\beta = \frac{\langle \beta \rangle}{\gamma}. \tag{23}$$

In Appendix B, we prove that the disease free equilibrium (DFE) is unstable if and only if $\mathcal{R}_0 > 1$.

To find the final size, we divide by S on both sides of (21a) and integrate,

$$\log \frac{S(\infty)}{S(0)} = \int_0^\infty \int_0^\infty \beta I(\beta, t) \, \mathrm{d}\beta \, \mathrm{d}t = \int_0^\infty \beta \int_0^\infty I(\beta, t) \, \mathrm{d}t \, \mathrm{d}\beta. \tag{24}$$

Substituting Eq. (21a) in Eq. (21b) we can write

$$\dot{I}(\beta, t) = -q(\beta)\dot{S} - \gamma I(\beta, t).$$

Integrating this we obtain

$$I(\beta, \infty) - I(\beta, 0) = -q(\beta)[S(\infty) - S(0)] - \gamma \int_0^\infty I(\beta, t)\, \mathrm{d}t.$$

Since we have established that $I(\beta, \infty) = 0$, we can write

$$\gamma \int_0^\infty I(\beta, t)\, \mathrm{d}t = -q(\beta)[S(\infty) - S(0)] + I(\beta, 0). \tag{25}$$

Substituting (25) into (24) and using (23), we obtain

$$\log \frac{S(\infty)}{S(0)} = -[S(\infty) - S(0)]\mathcal{R}_0 + \frac{1}{\gamma} \int_0^\infty \beta I(\beta, 0)\, \mathrm{d}\beta.$$

Hence, since the final size $Z = S(0) - S(\infty)$, we have

Theorem 8.1. *Consider the SIR model with arbitrarily distributed transmission rates, as specified by Eqs. (21). In the limit that the initial proportion of individuals that are infectious is small, i.e., $\int_0^\infty \beta I(\beta, 0)\, \mathrm{d}\beta \to 0$, the final size of the epidemic is given by the usual formula (5) with $\mathcal{R}_0$ given by Eq. (23).*

It is possible to generalize this theorem for an SI^nR model with arbitrarily distributed stage durations and arbitrarily distributed transmission rates in each stage. We do not go through the details because no new ideas are required. With Theorems 7.1 and 8.1 in hand, the extra effort required for the generalization is merely one of keeping track of notation.

9. Spatial structure

In all of the models that we have discussed so far, we have assumed that the population is homogeneously mixed. In the remainder of this paper, we explore the significance of heterogeneous mixing for the final size formula. In the present section we identify an important class of spatially structured models for which the standard formula (5) remains valid.

A simple but important example of heterogenous mixing occurs if the population is divided into a number of spatially isolated patches (e.g., cities), often called a *metapopulation* (Hanski and Gilpin, 1997). In ecological models, coupling among patches in a metapopulation usually occurs as a result of migration (Earn et al., 2000b). In the present context, inter-patch coupling occurs because individuals travel temporarily from their home patch to other patches. On such journeys, susceptible individuals might contact infectious individuals and become infected,

and infectious individuals might contact susceptibles and infect them. If we restrict attention to a single infectious stage with exponentially distributed infectious period then the model is specified as follows.

$$\dot{S_i} = -S_i \sum_{j=1}^{n} \beta_{ij} I_j, \tag{26a}$$

$$\dot{I_i} = S_i \sum_{j=1}^{n} \beta_{ij} I_j - \gamma I_i, \tag{26b}$$

$$\dot{R_i} = \gamma I_i. \tag{26c}$$

Here, S_i, I_i, and R_i ($i = 1, \ldots, n$) are the proportions of the population in patch i that are susceptible, infectious recovered, respectively. Summing Eqs. (26), we see that $S_i(t) + I_i(t) + R_i(t) = 1$ is invariant in all patches. We denote by N_i the number of individuals in patch i, so the total population size is $N = \sum_i N_i$. It is also convenient to use $B = (\beta_{ij})$ to denote the $n \times n$ transmission matrix.

Similar to our analysis of other models in previous sections, it is straightforward to prove that $S_i(\infty)$, $R_i(\infty)$, $I_i(\infty)$ exist, that $S_i(\infty) + R_i(\infty) = 1$ and that $I_i(\infty) = 0$. Since $\beta_{ij} \geq 0$, the dominant eigenvalue λ of B is real and positive (Horn and Johnson, 1985). If we linearize the model (26) at the disease free equilibrium (DFE) ($S_i = 1$, $I_i = 0$, $R_i = 0$), we see that the stability of the DFE is determined by λ. If $\lambda/\gamma > 1$, then the DFE is unstable, otherwise it is stable. Thus, the threshold for exponential growth of cases (an epidemic) is determined by the value of $\mathcal{R}_0 = \lambda/\gamma$, which can be interpreted as the basic reproduction number (van den Driessche and Watmough, 2002).

The final size in patch i is $Z_i = S_i(0) - S_i(\infty)$. We can obtain a system of n coupled algebraic equations for the Z_i if we divide by S_i on both sides of Eq. (26a) and integrate from 0 to ∞. These algebraic equations contain $\int_0^\infty I_i \, \mathrm{d}t$, which can be eliminated by integrating Eq. (26c) and noting that $Z_i = R_i(\infty) - R_i(0)$. In the limit $I_i(0) \to 0$, $S_i(0) \to 1$, we obtain

$$\log(1 - Z_i) = -\sum_j \frac{\beta_{ij}}{\gamma} Z_j, \qquad i = 1, \ldots, n. \tag{27}$$

These equations form a special case of a more general final size formula discussed by Diekmann and Heesterbeek (2000) (exercises 6.17 and 6.19). In their solution to exercise 6.19, Diekmann and Heesterbeek (2000) sketch an argument that implies that if $\mathcal{R}_0 > 1$ then there is a unique non-trivial (i.e., non-zero) solution to Eqs. (27).

The final size for the whole metapopulation, Z, is not in general the simple average of the Z_i's in Eq. (27). Rather, each Z_i must be weighted by the population size N_i of patch i, i.e.,

$$Z = \sum_{i=1}^{n} \frac{N_i}{N} Z_i. \tag{28}$$

Consequently, there will always be particular values of the N_i that make Z in Eq. (28) equal to the solution of the standard formula (5). However, since population sizes are given a priori, the problem of interest is to find necessary and sufficient conditions on the transmission matrix (B) that ensure that the sum in Eq. (28) is equal to the final size obtained from the standard formula (5), regardless of the values of the N_i.

In Eq. (28), Z will have no dependence on the patch population sizes if and and only if

$$Z_1 = Z_2 = \cdots = Z_n = Z. \tag{29}$$

Given this, Eq. (27) implies

$$\sum_{j=1}^{n} \beta_{ij} = \beta = \text{constant}, \tag{30}$$

i.e., the transmission rate is the same in all patches. If, on the other hand, we start with Eq. (30) as an assumption, then inserting Eq. (29) in Eq. (27) yields the standard final size formula (5). Thus, Eq. (27) and Eq. (5) yield the same final size Z for the population as a whole if and only if Eq. (30) is satisfied.

Restricting attention to situations in which the transmission rate is the same in every patch may seem to be an extremely stringent constraint. In fact, this is often the situation of interest. For example, individuals who normally reside in Boston, Philadelphia or New York can expect to contact similar numbers of people each day, even though the New York metropolitan area has a population that is several times larger than the other two. More generally, if patches represent large cities in a given country, then it is reasonable to expect that the number of contacts an individual has will not depend strongly on the size of the city.

Given Eq. (30), we can write

$$\beta_{ij} = \beta P_{ij}, \tag{31}$$

where P_{ij} is the probability, for an individual in patch i, that a given contact is with an individual in patch j, and β is the transmission rate (the product of the number of contacts per unit time and the probability that a contact between a susceptible and an infectious individual leads to transmission). Note that since the rows of the matrix P are probability distributions, it follows that

$$\sum_{j=1}^{n} P_{ij} = 1, \qquad i = 1, \ldots, n, \tag{32}$$

i.e., P is a *stochastic matrix* (Horn and Johnson, 1985).

For convenience, we make the mild assumption that $P_{ij} > 0$ for all i and j, i.e., for any pair of patches, there is a non-zero probability that residents of the two patches will come into contact. It then follows that if the final size Z is positive

then the final size Z_i is strictly positive in every patch. To see this, note that if $Z_i = 0$ for some i then Eq. (27) implies

$$\sum_{j \neq i} \beta_{ij} Z_j = -\log(1 - Z_i) = 0. \tag{33}$$

Since $P_{ij} > 0$ implies $\beta_{ij} > 0$, Eq. (33) implies $Z_j = 0$ for all j.

Since P is a positive, stochastic matrix, its largest eigenvalue is 1 (Horn and Johnson, 1985). Consequently, the largest eigenvalue of $B = \beta P$ is β and the basic reproduction number is $\mathcal{R}_0 = \frac{\beta}{\gamma}$.

Therefore, we have

Theorem 9.1. *Consider the multi-patch SIR epidemic model specified by Eqs. (26). Suppose the transmission rate in each patch is the same, i.e., $\sum_{j=1}^n \beta_{ij} = \beta$, for each i. Then, in the limit that the initial proportion of individuals that are infectious is small, i.e., $I_i(0) \to 0$ for each i, the final size Z of an epidemic is given implicitly by Eq. (5), or explicitly by Eq. (A.2), where $\mathcal{R}_0 = \frac{\beta}{\gamma}$.*

At this point, it will come as no surprise that with some effort in keeping track of notation, Theorem 9.1 can be generalized to a multi-patch SI^{n}R model with arbitrarily distributed stage durations and arbitrarily distributed transmission rates.

Theorem 9.2. *For a multi-patch SI^{n}R model with arbitrarily distributed stage durations, if susceptible recruitment and mortality are negligible, the distributions of stage durations are the same in each patch, and the transmission matrix B_i for each stage i is proportional to a stochastic matrix, then the final size of the epidemic Z_i in each patch i is given by the unique solution to the standard final size formula (5) where $\mathcal{R}_0 = \sum_i \beta_i / \gamma$, with β_i being the dominant eigenvalue of B_i.*

10. Social heterogeneities that affect the final size formula

Equations (26), which specify the spatially heterogeneous model we analyzed in the previous section, can be interpretted as representing other types of social heterogeneities, which are known to yield different final size formulae (Gart, 1968; Dwyer et al., 2000; Diekmann and Heesterbeek, 2000; Andreasen, 2003). In this section, we briefly discuss two other interpretations, which lead to a transmision matrix that is not proportional to a stochastic matrix and will not yield the usual final size formula (5).

Rather than specifying heterogeneities of transmission resulting from spatial structure, as in the previous section, suppose that heterogeneities arise from age-structured mixing patterns. One possible formulation is to categorize the population by discrete age cohorts, which is typically motivated by mixing patterns of children in schools. If we consider a time-scale short enough that transfer from one cohort to the next can be ignored, then the standard "realistic age-structured model" (Schenzle, 1984) reduces to Eqs. (26), where the transmission

matrix B now refers to contact patterns among different age cohorts. As in the case of the spatial patch interpretation, it might be reasonable to assume that there is a stochastic matrix P such that

$$\beta_{ij} = \beta_i P_{ij}, \tag{34}$$

where the interpretation is that individuals in cohort i have a proportion P_{ij} of their contacts with individuals in cohort j, and β_i is the transmission rate for individuals in cohort i. Unlike the spatial interpretation, however, it would be hard to justify taking β_i to be the same for each i. In particular, young children tend to come into much closer contact with their classmates than teenagers or adults do with members of their cohorts. As a result, Theorem 9.1 will not apply and the final size will not be given by the usual formula (5).

In Section 8 where we discussed super-spreaders, we specifically excluded the types of super-spreaders that occur for sexually transmitted diseases, namely individuals who always have a higher rate of contact with others, regardless of whether they happen to be infected. When such "core groups" (Yorke and Hethcote, 1984; Anderson et al., 1986) exist, Eqs. (26) can still be used, with the interpretation that the stratification of the population is by social group rather than spatial region or age cohort. Like the age-structured situation, we can safely assume that Eq. (34) holds but not that β_i is the same for every i as in Eq. (30). Indeed, a large difference in β_i is precisely what defines the core groups. Again, we have a situation where the usual final size formula will not apply. Core groups are well known to have a critical role in the transmission dynamics of sexually transmitted diseases, so it is not surprising that they will affect the final size of epidemics.

11. Discussion

The well-known formula (5) for the expected final size of an epidemic is valid in remarkably general circumstances. Previous work (e.g., Kermack and McKendrick (1927); Anderson and Watsons (1980); Diekmann and Heesterbeek (2000)) established that the formula is valid in an SIR model with an arbitrarily distributed infectious period (Theorem 6.1). Here we have shown, in addition, that the standard formula (5) is invariant to the number of latent and infectious stages of disease (Theorem 7.1), the distributions of transmission rates within stages (Theorem 8.1), and even to common spatial contact heterogeneities (Theorem 9.1).

The invariance of the final size formula has important practical implications. Typically, the time at which one wishes to estimate the expected final size of an epidemic is long before enough information has been gathered to estimate the distributions of latent and infectious periods or other epidemiological details. Our theorems provide rigorous support for estimates of the expected magnitude of epidemics based solely on estimates of the basic reproduction number $\mathcal{R}_0$, which is the only parameter that appears in the final size formula (5).

It should be noted that the final size formula refers only to the *ensemble average* size of an epidemic for a disease with a given $\mathcal{R}_0$. Different stochastic realizations of the same process will lead to different final sizes and our analysis

says nothing about the variance or higher moments of the final size distribution. Kurtz (1980) proved that in the limit that the population size goes to infinity, the ensemble mean of the stochastic SIR model with arbitrarily distributed infectious period converges to the solutions of the integro-differential equation model specified in our Eqs. (11). Hence, the standard final size formula gives the mean final size of the stochastic SIR model. However, it is difficult to deduce the distribution of the final size, and this remains an open problem for stochastic SIR models with arbitrarily distributed infectious periods. For stochastic SEIR models with Gamma distributed latent and infectious periods, Anderson and Watsons (1980) derived a normal approximation to the final size distribution that is valid in the limit of large population size.

If there is little variation in the length of the latent period, and the infectious period is short, the stochastic SEIR model can be approximated by a *chain binomial* model (Bailey, 1975), which is a discrete time Markov chain. The time step for this chain is equal to the fixed-length latent period, since the infectious period is presumed infinitesimal (all contacts occur instantaneously at the end of the latent period). For such models, the only parameter is the probability q that a contact between an infectious individual and a susceptible individual leads to infection. With the assumption that $q = 1 - e^{-\mathcal{R}_0}$, Von Bahr and Martin-Lof (1980) and Scalia-Tomba (1985) showed that final size distribution for the traditional chain binomial model is asymptotically normal in the limit of large population size. Scalia-Tomba (1986) and Andersson (1999) studied more general chain binomial models (with heterogeneous contact structures equivalent to those we discussed in Sections 9 and 10) and found asymptotic final size distributions. Their results imply that if $q = 1 - e^{-\mathcal{R}_0}$ then the ensemble mean final size in these chain binomial models is the solution of our Eqs. (27).

We emphasized two circumstances under which the usual final size formula will not apply. One was the case of age-structured heterogeneity of transmission; while this would influence the final size, it is unlikely that we would be able to parameterize the age-structured contact patterns sufficiently well to improve on the final size estimate generated with the assumption of homogeneous mixing. The other case we discussed was the existence of social core groups, which are extremely important in the epidemiological dynamics of sexually transmitted diseases; in this case, some estimate of the difference in transmission rates in different social groups would likely be needed to obtain a useful estimate of the final size.

In computing the final size, we have always assumed that there are no temporal changes in the transmission rate β. Temporal variations in β could result from intrinsic seasonality of contact rates, imposition of control measures, or behavioral change in response to epidemic alerts. Small seasonal variations in β cause only small perturbations to the final size formula, and substantial seasonality is generally associated with school-age children (London and Yorke, 1973) who form a small fraction of the susceptible pool when a new infection enters a population. Infection control measures, such as those adopted during the 2003 SARS epidemic (Lipsitch et al., 2003; Wallinga and Teunis, 2004), could have a dramatic effect on the final size; in this case, rather than $\mathcal{R}_0$, the reproduction number that is relevant for the final size formula must be calculated after sustainable precautionary

measures have been put into place. Individual behavioral change could alter the transmission rate through time-dependence ($\beta = \beta(t)$, as in the case of imposed control measures) or through density-dependence ($\beta = \beta(I)$, with β decreasing as I increases, which would arise if people tend to be more careful when they are aware of more cases). In either case, the standard final size formula might yield a poor approximation of the true final size.

All of our analysis is predicated on the assumption that vital dynamics (births and deaths), and more generally any source of new susceptible individuals, can be ignored. This approximation is not valid in circumstances where infectious periods are substantial compared with life expectancy or where immunity decays rapidly. It will, however, typically be relevant whenever a new respiratory pathogen emerges, as in the case of the SARS outbreak in 2003 (Donnelly et al., 2003) or the emergence of a new subtype of influenza (Earn et al., 2002).

Appendix A: Explicit form of the final size formula

An explicit solution for Z expressed in terms of elementary functions is not possible, but Eq. (4) can be solved explicitly for Z in terms of the Lambert W function (Weisstein),

$$Z = S(0) + \frac{1}{\mathcal{R}_0} W\big[- \mathcal{R}_0 S(0) \, e^{-\mathcal{R}_0(I(0)+S(0))}\big]. \tag{A.1}$$

Here, $W(x)$ is the unique solution of the transcendental equation $x = W \, e^{W}$, for $x \in [-\frac{1}{e}, 0]$. Note that since $\mathcal{R}_0$, $S(0)$, $I(0) \geq 0$, the argument of W in this formula lies in the interval $[-\frac{1}{e}, 0]$.

For a newly invading pathogen, Eq. (5) becomes

$$Z(\mathcal{R}_0) = 1 + \frac{1}{\mathcal{R}_0} W[-\mathcal{R}_0 \, e^{-\mathcal{R}_0}]. \tag{A.2}$$

Appendix B: Proof that model (11) with exponentially distributed infectious periods yield the standard SIR model

When the infectious period is exponentially distributed, i.e., $f(t) = \gamma \, e^{-\gamma t}$, we have $l(\tau, t) = f(t)$. Thus Eq. (11f) can be written

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta S(t)I(t) - \int_0^t f(t - \tau)U(\tau)\,\mathrm{d}\tau - I(0)f(t).$$

Let

$$x(t) = \frac{1}{\gamma}\left[\int_0^t f(t - \tau)\beta S(\tau)I(\tau)\,\mathrm{d}\tau - I(0)f(t)\right],$$

so

$$\dot{x} = S(t)I(t) - \frac{1}{\gamma} \int_0^t \gamma^2 e^{-\gamma(t-\tau)} \beta S(\tau)I(\tau) \, d\tau - I(0)\gamma^2 e^{-\gamma t},$$

$$= \beta S(t)I(t) - \gamma x.$$

Equation (11f) can also be written as

$$\dot{I} = \beta S(t)I(t) - \gamma x,$$

Thus, $\frac{d}{dt}(I(t) - x(t)) = 0$, so if we let $x_i(0) = I_i(0)$, then $I(t) \equiv x(t)$. In other words, system (11) becomes

$$\dot{S} = -\beta Sx,$$

$$\dot{x} = \beta Sx - \gamma x,$$

$$\dot{R} = \gamma x,$$

which is indeed the standard SIR model (1).

Appendix C: Proof of Lemma 6.1

Integrating Eq. (11f) we obtain

$$I(t) - I(0) = \int_0^t U(\tau) \, d\tau - \int_0^t \int_0^\tau f(\tau - s)U(s) \, ds \, d\tau$$

$$- \int_0^t \int_{-\infty}^0 U(\tau)l(\tau, t) \, d\tau. \tag{C.1}$$

Since $I(0) = \int_{-\infty}^0 U(\tau) \, d\tau \ll 1$, we can drop the rightmost term on both sides of Eq. (C.1), yielding

$$I(t) = \int_0^t U(\tau) \, d\tau - \int_0^t \int_0^\tau f(\tau - s)U(s) \, ds \, d\tau.$$

Switching the order of integration in the second term above, we have

$$\int_0^t \int_0^\tau f(\tau - s)U(s) \, ds \, d\tau = \int_0^t \int_s^t f(\tau - s) \, d\tau \, U(s) \, ds,$$

$$= \int_0^t \int_0^{t-s} f(\tau) \, d\tau \, U(s) \, ds.$$

Hence, if we let $F(t) = \int_0^t f(\tau)\,d\tau$, we can write

$$I(t) = \int_t^\infty U(\tau)\,d\tau - \int_0^t \int_0^{t-s} f(\tau)\,d\tau\,U(s)\,ds,$$

$$= \int_0^t [1 - F(t-s)]U(s)\,ds.$$

Integrating this from 0 to ∞, we obtain

$$\int_0^\infty I(t)\,dt = \int_0^\infty \int_0^t [1 - F(t-s)]U(s)\,ds\,dt.$$

Next, switching the order of the integration, we have

$$\int_0^\infty I(t)\,dt = \int_0^\infty \int_s^\infty [1 - F(t-s)]\,dt\,U(s)\,ds,$$

$$= \int_0^\infty [1 - F(t)]\,dt \int_0^\infty U(s)\,ds.$$

Noting now that

$$\int_0^\infty [1 - F(t)]\,dt = \int_0^\infty t f(t)\,dt = T,$$

we have, finally,

$$\int_0^\infty I(t)\,dt = T \int_0^\infty U(s)\,ds,$$

as required.

Appendix D: Stability of the disease free equilibrium in the model (21) with arbitrarily distributed transmission rate

The local stability of the DFE ($S = 1$, $I(\beta,) = 0\,\forall \beta$, $R = 0$) of the model (21) is determined by its linearization,

$$\dot{S} = -\int_0^\infty \beta I(\beta, t)\,d\beta, \tag{D.1a}$$

$$\dot{I}(\beta, t) = q(\beta) \int_0^\infty \beta I(\beta, t)\,d\beta - \gamma I(\beta, t). \tag{D.1b}$$

We are interested in growth or decay of $I(\beta,\ )$, which is determined only by Eq. (D.1b). Defining $X(t) = \int_0^\infty \beta I(\beta, t)\, \mathrm{d}\beta$, multiplying both sides of Eq. (D.1b) by β, and integrating with respect to β, we find

$$\frac{\mathrm{d}}{\mathrm{d}t} X = \left(\int_0^\infty \beta q(\beta)\, \mathrm{d}\beta - \gamma \right) X.$$

On the other hand, recall that $I(t) = \int_0^\infty I(\beta, t)\, \mathrm{d}t$ (the total number of infectious individuals at time t). Therefore, integrating Eq. (B.2b) with respect to β, noting that $q(\beta)$ is a probability density function so $\int_0^\infty q(\beta)\, \mathrm{d}\beta = 1$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t} I = X - \gamma I.$$

Thus,

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{bmatrix} X \\ I \end{bmatrix} = \begin{bmatrix} \int_0^\infty \beta q(\beta)\, \mathrm{d}\beta - \gamma & 0 \\ 1 & -\gamma \end{bmatrix} \begin{bmatrix} X \\ I \end{bmatrix}.$$

The origin of this system of two ordinary differential equations is locally stable if and only if $\int_0^\infty \beta q(\beta)\, \mathrm{d}\beta - \gamma < 0$ or, equivalently, $\mathcal{R}_0 < 1$. Thus, the DFE is stable if $\mathcal{R}_0 < 1$ and unstable if $\mathcal{R}_0 > 1$.

We can expand model (21) to include multiple stages. Using the same techniques, we find that the conclusion that the epidemic threshold is given by $\mathcal{R}_0 = 1$ remains valid.

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, R.M., May, R.M., 1991. Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, Oxford.

Anderson, R.M., Medley, G.F., May, R.M., Johnson, A.M., 1986. A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS. IMA J. Math. Appl. Med. Biol. 3(4), 229–263.

Andersson, H., Britton, T., 2000. Stochastic epidemic models and their statistical analysis. Lecture Notes in Statistics, vol. 151. Springer-Verlag, New York.

Andersson, M., 1999. The asymptotic final size distribution of multitype chain-binomial epidemic processes. Adv. Appl. Prob. 31, 220–234.

Andreasen, V., 2003. Dynamics of annual influenza a epidemics with immuno-selection. J. Math. Biol. 46(6), 504–536.

Bailey, N.T.J., 1975. The Mathematical Theory of Infectious Diseases and its Application, 2nd edition. Griffin, London.

Brauer, F., Castillo-Chavez, C., 2001. Mathematical models in population biology and epidemiology. Texts in Applied Mathematics, vol. 40. Springer-Verlag, New York.

Diekmann, O., Heesterbeek, J.A.P., 2000. Mathematical epidemiology of infectious diseases: Model building, analysis and interpretation. Wiley Series in Mathematical and Computational Biology. John Wiley & Sons, LTD, New York.

Donnelly, C.A., Ghani, A.C., Leung, G.M., Hedley, A.J., Fraser, C., Riley, S., Abu-Raddad, L.J., Ho, L.M., Thach, T.Q., Chau, P., Chan, K.P., Lam, T.H., Tse, L.Y., Tsang, T., Liu, S.H.,

Kong, J.H.B., Lau, E.M.C., Ferguson, N.M., Anderson, R.M., 2003. Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. Lancet 361(9371), 1761–1766.

Dwyer, G., Dushoff, J., Elkinton, J.S., Levin, S.A., 2000. Pathogen-driven outbreaks in forest defoliators revisited: Building models from exprimental data. Am. Nat. 156, 105–120.

Earn, D.J.D., Dushoff, J., Levin, S.A., 2002. Ecology and evolution of the flu. Trends Ecol. Evol. 17(7), 334–340.

Earn, D.J.D., Levin, S.A., Rohani, P., 2000b. Coherence and conservation. Science 290(5495), 1360–1364.

Earn, D.J.D., Rohani, P., Bolker, B.M., Grenfell, B.T., 2000a. A simple model for complex dynamical transitions in epidemics. Science 287(5453), 667–670.

Feng, Z.L., Thieme, H.R., 2000. Endemic models with arbitrarily distributed periods of infection I: Fundamental properties of the model. SIAM J. Appl. Math. 61(3):803–833.

Gart, J.J., 1968. The mathematical analysis of an epidemic with two kinds of susceptibles. Biometrics 557–566.

Hanski, I.A., Gilpin, M.E. (Eds.) 1997. Metapopulation Biology: Ecology, Genetics, and Evolution. Academic Press, San Diego.

Horn, R.A., Johnson, C.R., 1985. Matrix Analysis. Cambridge University Press, Cambridge.

Hyman, J.M., Li, J.A.S.E., 1999. The differential infectivity and staged progression models for the transmission of HIV. Math. Biosci. 77–109.

Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics. Proc. R. Soc. Lond. Ser. A 115, 700–721.

Kurtz, T.G., 1980. Relationships between stochastic and deterministic population models. In: Lect. Notes Biomath. 38, 449–467.

Lipsitch, M., Cohen, T., Cooper, B., Robins, J.M., Ma, S., James, L., Gopalakrishna, G., Chew, S.K., Tan, C.C., Samore, M.H., Fisman, D., Murray, M., 2003. Transmission dynamics and control of severe acute respiratory syndrome. Science 300(5627), 1966–1970.

Lloyd, A.L., 2001a. Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. Proc. R. Soc. Lond. Ser. B 268(1470), 985–993.

Lloyd, A.L., 2001b. Realistic distributions of infectious periods in epidemic models: Changing patterns of persistence and dynamics. Theor. Popul. Biol. 60(1), 59–71.

London, W., Yorke, J.A., 1973. Recurrent outbreaks of measles, chickenpox and mumps. I Seasonal variation in contact rates. Am. J. Epidemiol. 98(6), 453–468.

Low, D.E., McGeer, A., 2003. SARS—one year later. N. Engl. J. Med. 349(25), 2381–2382.

Poutanen, S.M., Low, D.E., Henry, B., Finkelstein, S., Rose, D., Green, K., Tellier, R., Draker, R., Adachi, D., Ayers, M., Chan, A.K., Skowronski, D.M., Salit, I., Simor, A.E., Slutsky, A.S., Doyle, P.W., Krajden, M., Petric, M., Brunham, R.C., McGeer, A.J., 2003. Identification of severe acute respiratory syndrome in Canada. N. Engl. J. Med. 348(20), 1995–2005.

Redfield, R.R., Wright, D.C., Tramont, E.C., 1986. The Walter Reed staging classification for HTLV-III/LAV infection. N. Eng. J. Med. 314, 131–132.

Scalia-Tomba, G., 1985. Asymptotic final size distribution for some chain-binomial processes. Adv. Appl. Prob. 17, 477–495.

Scalia-Tomba, G., 1986. Asymptotic final size distribution of the multitype Reed-Frost process. Adv. Appl. Prob. 23, 563–584.

Schenzle, D., 1984. An age-structured model of pre- and post-vaccination measles transmission. IMA J. Math. Appl. Med. Biol. 1, 169–191.

Schwartz, I., Smith, H., 1983. Infinite subharmonic bifurcation in an SEIR model. J. Math. Biol. 18, 233–253.

Seligmann, M., Pinching, A.J., Rosen, F.S.E.A., 1987. Immunology of human immunodeficiency virus and the acquired immune dificiency syndrome. Ann. Int. Med. 107(2), 234–242.

van den Driessche, P., Watmough, J., 2002. Reproduction numbers and subthreadold endemic equilibria for compartmental models fo disease transmission. Math. Biosci. 180, 29–48.

Von Bahr, B., Martin-Lof, A., 1980. Threshold limit theorems for some epidemic processes. Adv. Appl. Prob. 12, 319–349.

Wallinga, J., Teunis, P., 2004. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am. J. Epidemiol. 160, 509–516.

Weisstein, E.W., Lambert W-Function. From MathWorld–A Wolfram Web Resource. http:\\math world.wolfram.com/LambertW-Function.html.

WHO SARS Update 27, 2003. Severe acute respiratory syndrome (SARS), multi-country out-
    break. Technical report, W.H.O. http:\\www.who.int/csr/sars/archive/2003_04_11/en/print.
    html.
WHO SARS Update 33, 2003. Severe acute respiratory syndrome (SARS), multi-country out-
    break. Technical report, W.H.O. http:\\www.who.int/csr/sars/archive/2003_04_18/en/print.
    html.
Yorke, J.A., Hethcote, H.W., 1984. Gonorrhea: Transmission dynamics and control. Lecture notes
    in Biomathematics, 56, 1–105. Springer-Verlag, Berlin.