Chin. Ann. Math.
31B(3), 2010, 289–304
DOI: 10.1007/s11401-009-0454-1

Chinese Annals of Mathematics, Series B
© The Editorial Office of CAM and Springer-Verlag Berlin Heidelberg 2010

General Compartmental Epidemic Models

Fred BRAUER*

Abstract The age of infection approach introduced by Kermack and Mckendrick in 1927 gives a unified way of describing and analyzing a variety of epidemic models, including models with multiple stages, treatment, and heterogeneous mixing. The author gives a description of the main results for such models, emphasizing the use of the final size relation to estimate the size of the epidemic.

Keywords Epidemic models, Treatment models, Basic reproduction number, Final size relation
2000 MR Subject Classification 92D30

1 Introduction

An epidemic, acting on a short time scale, is an outbreak of a disease that infects a substantial portion of the population in a region before it disappears, usually leaving many members untouched. Often these attacks recur with intervals of several years between outbreaks, possibly diminishing in severity as populations develop some immunity. Epidemics have had major effects on the course of history.

One of the striking early results of mathematical epidemiology (see [18]) was the formulation of a simple model that predicted behaviour very similar to this. The Kermack-McKendrick model is a compartmental model based on relatively simple assumptions on the rates of flow between different classes of members of the population.

The purpose of this note is to give a unified description of the properties of compartmental epidemic models. Since the results have appeared elsewhere, we give an outline without including details of proofs. The models we describe do not include any demographic effects, in keeping with the idea of epidemics on a short time scale.

2 The Age of Infection Epidemic Model

The general epidemic model described by Kermack and McKendrick [18] included a dependence of infectivity on the time since becoming infected (age of infection). We let $S(t)$ denote the number of susceptibles at time t and let $\varphi(t)$ be the total infectivity at time t, defined as the sum of products of the number of infected members with each infection age and the mean infectivity for that infection age. We also let $\varphi_0(t)$ be the total infectivity at time t of those

Manuscript received January 6, 2010. Published online April 20, 2010.
*Department of Mathematics, University of British Columbia, Vancouver, B.C. V6T 1Z2, Canada.
E-mail: brauer@math.ubc.ca

members of the population who were infected at the start of the epidemic, $t = 0$. We assume that on the average members of the population make a constant number a of contacts in unit time. We let $B(\tau)$ be the fraction of infected members remaining infected at infection age τ and let $\pi(\tau)$ with $0 \le \pi(\tau) \le 1$ be the mean infectivity at infection age τ. Then we let

$$A(\tau) = \pi(\tau)B(\tau),$$

the mean infectivity of members of the population with infection age τ. We assume that there are no disease deaths, so that the total population size is a constant N.

The age of infection epidemic model is

$$\begin{aligned} S' &= -\beta S\varphi \\ \varphi(t) &= \varphi_0(t) + \int_0^t \beta S(t-\tau)\varphi(t-\tau)A(\tau)\mathrm{d}\tau \\ &= \varphi_0(t) + \int_0^t [-S'(t-\tau)]A(\tau)\mathrm{d}\tau. \end{aligned} \tag{2.1}$$

The basic reproduction number is

$$\mathcal{R}_0 = \beta N \int_0^\infty A(\tau)\mathrm{d}\tau.$$

Integration of

$$-\frac{S'(t)}{S(t)} = \beta\varphi_0(t) + \beta \int_0^t [-S'(t-\tau)]A(\tau)\mathrm{d}\tau$$

with respect to t from 0 to ∞ as in [3] gives the final size relation

$$\ln \frac{S_0}{S_\infty} = \mathcal{R}_0\Big[1 - \frac{S_\infty}{N}\Big] - \beta \int_0^\infty [(N - S_0)A(t) - \varphi_0(t)]\mathrm{d}t.$$

If all initial infectives have infection age zero at $t = 0$, $\varphi_0(t) = [N - S_0]A(t)$, and

$$\int_0^\infty [\varphi_0(t) - (N - S_0)A(t)]\mathrm{d}t = 0.$$

Then (2.2) takes the form

$$\ln \frac{S_0}{S_\infty} = \mathcal{R}_0\Big(1 - \frac{S_\infty}{N}\Big), \tag{2.2}$$

and this is the commonly used form of the final size relation. Note that the final size of the epidemic, the number of members of the population who are infected over the course of the epidemic, is $N - S_\infty$. This is often described in terms of the attack ratio $\big(1 - \frac{S_\infty}{N}\big)$.

If there are initial infectives with infection age greater than zero, the initial term satisfies

$$\int_0^\infty [(N - S_0)A(t) - \varphi_0(t)]\mathrm{d}t \ge 0.$$

The final size relation is sometimes presented in the form

$$\ln \frac{S_0}{S_\infty} = \mathcal{R}_0\Big(1 - \frac{S_\infty}{S_0}\Big), \tag{2.3}$$

with an initial term which is assumed small and omitted (see for example [1, 5, 15]).

It is not difficult to prove that there is a unique solution of the final size relation (2.2), and if $\mathcal{R}_0 > 1$ this solution satisfies

$$S_\infty < \frac{S_0}{\mathcal{R}_0}.$$

The special case $I(t) = \varphi(t),\ A(\tau) = \mathrm{e}^{-\alpha\tau}$ gives the model

$$\begin{aligned} S' &= -\beta SI, \\ I' &= \beta SI - \alpha I, \end{aligned} \tag{2.4}$$

which is the simplest example of an epidemic model, and has often been used to fit epidemic data. It is sometimes called the Kermack-McKendrick model, ignoring the fact that the original Kermack-McKendrick model was actually the much more general (2.1).

3 The Beginning of a Disease Outbreak

The Kermack–McKendrick compartmental epidemic model assumes that the sizes of the compartments are large enough that the mixing of members is homogeneous, or at least that there is homogeneous mixing in each subgroup if the population is stratified by activity levels. However, at the beginning of a disease outbreak, there is a very small number of infective individuals. The transmission of infection is a stochastic event depending on the pattern of contacts between members of the population. A description should take this pattern into account.

Our approach is to give a stochastic branching process description of the beginning of a disease outbreak to be applied so long as the number of infectives remains small, distinguishing a (minor) disease outbreak confined to this stage from a (major) epidemic which occurs if the number of infectives begins to grow at an exponential rate. Once an epidemic has started, we may switch to a deterministic compartmental model, arguing that in a major epidemic contacts would tend to be more homogeneously distributed.

We describe the network of contacts between individuals by a graph with members of the population represented by vertices and with contacts between individuals represented by edges. The study of graphs originated with the abstract theory of Erdös and Rényi [8–10], and has become important more recently in many areas, including social contacts and computer networks, as well as the spread of communicable diseases.

An edge is a contact between vertices that can transmit infection. The number of edges of a graph at a vertex is called the degree of the vertex. The degree distribution of a graph is $\{p_k\}$, where p_k is the fraction of vertices having degree k. The degree distribution is fundamental in the description of the spread of disease.

We think of a small number of infectives in a population of susceptibles large enough that in the initial stage we may neglect the decrease in the size of the susceptible population. Our development begins along the lines of that in [5] and then develops along the lines in [21, 23]. We assume that the infectives make contacts independently of one another and let p_k denote the probability that the number of contacts by a randomly chosen individual is exactly k,

with $\sum_{k=0}^{\infty} p_k = 1$. In other words, $\{p_k\}$ is the degree distribution of the vertices of the graph corresponding to the population network.

For convenience, we define the generating function

$$G_0(z) = \sum_{k=0}^{\infty} p_k z^k.$$

Since $\sum_{k=0}^{\infty} p_k = 1$, this power series converges for $0 \le z \le 1$, and may be differentiated term by term. Thus

$$p_k = \frac{G_0^{(k)}(0)}{k!}, \quad k = 0, 1, 2, \cdots .$$

It is easy to verify that the generating function has the properties

$$G_0(0) = p_0, \quad G_0(1) = 1, \quad G_0'(z) > 0, \quad G_0''(z) > 0.$$

The mean degree, which we denote by $\langle k \rangle$, is

$$\langle k \rangle = \sum_{k=1}^{\infty} k p_k = G_0'(1).$$

This is also the mean number of neighbors of a randomly chosen vertex, and is also denoted by z_1.

When a disease is introduced into a network, we think of it as starting at a vertex (patient zero) who transmits infection to every individual to whom this individual is connected, that is, along every edge of the graph from the vertex corresponding to this individual. We assume that this initial vertex has been infected by a contact outside the population (component of the network) being studied. For transmission of disease after this initial contact we need to use the excess degree of a vertex. If we follow an edge to a vertex, the excess degree of this vertex is one less than the degree. We use the excess degree because infection can not be transmitted back along the edge whence it came. The probability of reaching a vertex of degree k, or excess degree $(k-1)$, by following a random edge is proportional to k, and thus the probability that a vertex at the end of a random edge has excess degree $(k-1)$ is a constant multiple of kp_k with the constant chosen to make the sum over k of the probabilities equal to 1. Then the probability that a vertex has excess degree $(k-1)$ is

$$q_{k-1} = \frac{k p_k}{\langle k \rangle}.$$

This leads to a generating function $G_1(z)$ for the excess degree

$$G_1(z) = \sum_{k=1}^{\infty} q_{k-1} z^{k-1} = \sum_{k=1}^{\infty} \frac{k p_k}{\langle k \rangle} z^{k-1} = \frac{1}{\langle k \rangle} G_0'(z),$$

and the mean excess degree, which we denote by $\langle k_e \rangle$, is

$$\langle k_e \rangle = \frac{1}{\langle k \rangle} \sum_{k=1}^{\infty} k(k-1) p_k = \frac{1}{\langle k \rangle} \sum_{k=1}^{\infty} k^2 p_k - \frac{1}{\langle k \rangle} \sum_{k=1}^{\infty} k p_k = \frac{\langle k^2 \rangle}{\langle k \rangle} - 1 = G_1'(1).$$

We let $\mathcal{R}_0 = G_1'(1)$ be the mean excess degree. This is the mean number of secondary cases infected by patient zero and is the basic reproduction number as usually defined; the threshold for an epidemic is determined by $\mathcal{R}_0$. We may also express the basic reproduction number in terms of the mean number of neighbors and second neighbors of an arbitrarily chosen vertex. The mean number of second neighbors of a randomly chosen vertex, denoted by z_2, is

$$\sum_{k=1}^{\infty} k(k-1)p_k = \langle k^2 \rangle - \langle k \rangle,$$

and thus we have

$$\mathcal{R}_0 = \frac{z_2}{z_1}.$$

The first main result for this description of a disease outbreak is that if $\mathcal{R}_0 < 1$ the probability that the infection will die out is 1 but on the other hand, if $\mathcal{R}_0 > 1$ there is a solution $z_\infty < 1$ of

$$G_1(z) = z$$

and there is a probability $1 - G_0(z_\infty) > 0$ that the infection will persist, and will lead to a major epidemic. However, there is a positive probability $G_0(z_\infty)$ that the infection will increase initially but will produce only a minor outbreak and will die out before triggering a major epidemic. This distinction between a minor outbreak and a major epidemic, and the result that if $\mathcal{R}_0 > 1$ there may be only a minor outbreak and not a major epidemic are aspects of stochastic models not reflected in deterministic models. In distinguishing between a minor outbreak and a major epidemic, implicitly we are thinking of a population of infinite size and a major epidemic is a disease outbreak that spreads to a non-zero fraction of this population.

If contacts between members of the population are random, corresponding to the assumption of mass action in the transmission of disease, then the probabilities p_k are given by the Poisson distribution

$$p_k = \frac{\mathrm{e}^{-c} c^k}{k!}$$

with $c = \mathcal{R}_0$.

It has been observed that in many situations there is a small number of long range connections in the graph, allowing rapid spread of infection. There is a high degree of clustering (some vertices with many edges) and there are short path lengths. Such a situation may arise if a disease is spread to a distant location by an air traveller. This type of network is called a small world network. Long range connections in a network can increase the likelihood of an epidemic dramatically.

The commonly observed situation that most infectives do not pass on infection but there are a few "superspreading events" (see [25]) corresponds to a probability distribution that is quite different from a Poisson distribution, and could give a quite different probability that an epidemic will occur. Examples demonstrate that the probability of a major epidemic depends strongly on the nature of the contact network. Simulations suggest that for a given value of the basic reproduction number the Poisson distribution is the one with the maximum probability

of a major epidemic. We will not explore network models further here, but we point out that this is an actively developing field of science. Some basic references are [21–23, 26].

4 Models with Disease Deaths

The assumption in the model (2.4) of a rate of contacts per infective which is proportional to population size N, called mass action incidence or bilinear incidence, was used in all the early epidemic models. However, it is quite unrealistic, except possibly in the early stages of an epidemic in a population of moderate size. It is more realistic to assume a contact rate which is a non-increasing function of total population size. For example, a situation in which the number of contacts per infective in unit time is constant, called standard incidence, is a more accurate description for sexually transmitted diseases. If there are no disease deaths, so that the total population size remains constant, such a distinction is unnecessary.

We generalize the model (2.4) by making the assumption that an average member of the population makes $C(N)$ contacts in unit time with $C'(N) \geq 0$ (see [7]), and we define

$$\beta(N) = \frac{C(N)}{N}.$$

It is reasonable to assume $\beta'(N) \leq 0$ to express the idea of saturation in the number of contacts. Then mass action incidence corresponds to the choice $C(N) = \beta N$, and standard incidence corresponds to the choice $C(N) = \lambda$. The assumptions $C(N) = N\beta(N),\ C'(N) \geq 0$ imply

$$\beta(N) + N\beta'(N) \geq 0. \tag{4.1}$$

Some epidemic models (see [7]) have used a Michaelis-Menten type of interaction of the form

$$C(N) = \frac{aN}{1 + bN}.$$

Another form based on a mechanistic derivation for pair formation (see [14]) leads to an expression of the form

$$C(N) = \frac{aN}{1 + bN + \sqrt{1 + 2bN}}.$$

Data for diseases transmitted by contact in cities of moderate size (see [20]) suggests that data fits the assumption of a form

$$C(N) = \lambda N^a$$

with $a = 0.05$ quite well. All of these forms satisfy the conditions $C'(N) \geq 0,\ \beta'(N) \leq 0$.

Because the total population size is now present in the model, we must include an equation for total population size in the model. This forces us to make a distinction between members of the population who die of the disease and members of the population who recover with immunity against reinfection. We assume that a fraction f of the αI members leaving the infective class at time t recover and the remaining fraction $(1 - f)$ die of disease. We use $S,\ I$

and N as variables, with $N = S + I + R$. We now obtain a three-dimensional model

$$\begin{aligned} S' &= -\beta(N)SI, \\ I' &= \beta(N)SI - \alpha I, \\ N' &= -(1-f)\alpha I. \end{aligned} \tag{4.2}$$

Since N is now a decreasing function, we define $N(0) = N_0 = S_0 + I_0$. We also have the equation $R' = -f\alpha I$, but we need not include it in the model since R is determined when S, I and N are known. We should note that if $f = 1$ the total population size remains equal to the constant N, and the model (4.2) reduces to the simpler model (2.4) with β replaced by the constant $\beta(N_0)$.

For the model (4.2), the basic reproduction number is given by

$$\mathcal{R}_0 = \frac{N_0\beta(N_0)}{\alpha},$$

and there is a final size inequality

$$\ln\frac{S_0}{S_\infty} = \int_0^\infty \beta(N(t))I(t)\mathrm{d}t \geq \beta(N_0)\int_0^\infty I(t)\mathrm{d}t = \mathcal{R}_0\Big[1 - \frac{S_\infty}{N_0}\Big].$$

If the disease death rate is small, the final size inequality is an approximate equality. In the remainder of these notes, we will always assume that there are no disease deaths and thus the total population size is constant. This will allow us to assume throughout that β is a constant.

5 Example: The *SEIR* Model

The age of infection model includes models with multiple infective and treatment stages. For example, consider the standard $SEIR$ epidemic model but with individuals in E having infectivity reduced by a factor ε. The model can be described by the system

$$\begin{aligned} S' &= -\beta S(I + \varepsilon E), \\ E' &= \beta S(I + \varepsilon E) - \kappa E, \\ I' &= \kappa E - \alpha I, \end{aligned} \tag{5.1}$$

with initial conditions

$$S(0) = S_0, \quad E(0) = E_0, \quad I(0) = I_0, \quad N(0) = N = S_0 + E_0 + I_0.$$

The initial condition assumption is that there are no removed members at the start of the epidemic. Since we are assuming that there are no disease deaths, the total population size $N(t)$ is a constant N.

The basic reproduction number may be calculated directly, and

$$\mathcal{R}_0 = \beta N\Big(\frac{1}{\alpha} + \frac{\varepsilon}{\kappa}\Big).$$

The model can be viewed as an age of infection model with

$$\varphi = \varepsilon E + I.$$

It is not difficult to calculate

$$A(\tau) = \varepsilon \mathrm{e}^{-\kappa\tau} + \frac{\kappa}{\kappa - \alpha}[\mathrm{e}^{-\alpha\tau} - \mathrm{e}^{-\kappa\tau}],$$

and it is easy to calculate

$$\int_0^\infty A(\tau)\mathrm{d}\tau = \frac{1}{\alpha} + \frac{\varepsilon}{\kappa}.$$

This gives the same value for $\mathcal{R}_0$ as was calculated directly.

6 Staged Progression Models

In epidemic models there is often a sequence of stages of different lengths and infectivities, known as staged progression models (see [17]), where individuals pass from one stage to the next. The simplest example is an $SEIR$ model with an exposed stage, possibly with some infectivity, before the development of symptoms. To describe such a model, we suppose that there is a finite sequence of n infected stages $I_1(t), \cdots, I_n(t)$, with relative infectivity parameters $\varepsilon_1, \cdots, \varepsilon_n$, and infectivity distributions $P_1(\tau), \cdots, P_n(\tau)$. It should be noted that $P_i(\tau)$ represents the fraction of members who were infected initially τ time units earlier who are in the stage I_i.

The total infectivity at time t is the sum of the infectivities of each infected compartment,

$$\varphi(t) = \sum_{i=1}^{n} \varepsilon_i I_i(t).$$

The general age-of-infection model with a sequence of infected stages is

$$\begin{aligned} S'(t) &= -\beta S(t)\varphi(t), \\ \varphi(t) &= \int_0^\infty [-S'(t-\tau)] \sum_{i=1}^{n} \varepsilon_i A_i(\tau)\,\mathrm{d}\tau. \end{aligned}$$

Then

$$\mathcal{R}_0 = \beta N \sum_{i=1}^{n} \varepsilon_i \int_0^\infty A_i(\tau)\mathrm{d}\tau.$$

In order to calculate $\mathcal{R}_0$, we need to find

$$\int_0^\infty A_i(\tau)\,\mathrm{d}\tau.$$

It is known (see [28]) that the reproduction number is

$$\mathcal{R}_0 = \beta N \sum_{i=1}^{n} \varepsilon_i \int_0^\infty P_i(\tau)\,\mathrm{d}\tau.$$

This implies the following result.

Theorem 6.1 *The basic reproduction number $\mathcal{R}_0$ depends only on the mean period in each infective stage, regardless of its distribution. General epidemic models without treatment behave the same as models with exponentially distributed periods.*

There is no difficulty in extending the approach of this section to models, in which at the end of a stage individuals may proceed to one of two stages, such as the influenza model of [1, 2]. In this model, there is a latent period after which a fraction p of latent individuals L proceeds to an infective stage I, while the remaining fraction $(1-p)$ proceeds to an asymptomatic stage A, with infectivity reduced by a factor δ and a different period $\frac{1}{\eta}$.

With exponentially distributed latent, infective and asymptomatic periods, the model is

$$\begin{aligned} S' &= -\beta S[I + \delta A], \\ L' &= \beta S[I + \delta A] - \kappa L, \\ I' &= p\kappa L - \alpha I, \\ A' &= (1-p)\kappa L - \eta A \end{aligned} \tag{6.1}$$

and

$$\mathcal{R}_0 = \beta N\Big[\frac{p}{\alpha} + \frac{\delta(1-p)}{\eta}\Big].$$

According to Theorem 6.1, for a model with the same mean infective and asymptomatic periods the basic reproduction number has the same value.

The model (6.1) is an example of a differential infectivity model. In such models, also used in the study of HIV/AIDS (see [16, 17]), individuals enter a specific group when they become infected and stay in that group over the course of the infection. Different groups may have different parameter values. For example, for influenza infective and asymptomatic members may have different infectivities and different periods of stay in the respective stages. Theorem 6.1 is applicable to such models, and shows that the basic reproduction number depends on the mean stay in each compartment, not on the specific form of the distribution.

In the next section, we examine treatment models that include the rate at which members are removed during a stage and transferred to a treatment stage. Such models differ from staged progression models in that members are removed from a compartment during their stay in the compartment rather than proceeding at the end of their stay in the compartment.

7 Example: A Treatment Model

Consider an SIR model in which a fraction γ per unit time of infectives are selected for treatment, and the treatment reduces infectivity by a fraction δ. Suppose that the rate of removal from infective class is η. The $SITR$ model, where T is the treatment class, is given by

$$\begin{aligned} S' &= -\beta S[I + \delta T], \\ I' &= \beta S[I + \delta T] - (\alpha + \gamma)I, \\ T' &= \gamma I - \eta T. \end{aligned} \tag{7.1}$$

In order to calculate the basic reproduction number, we may argue that an infective in a totally susceptible population causes βN new infections in unit time, and the mean time spent in the infective compartment is $\frac{1}{\alpha+\gamma}$. In addition, a fraction $\frac{\gamma}{\alpha+\gamma}$ of infectives is treated. While in the treatment stage the number of new infections caused in unit time is $\delta\beta N$, and the mean time in the treatment class is $\frac{1}{\eta}$. Thus $\mathcal{R}_0$ is

$$\mathcal{R}_0 = \frac{\beta N}{\alpha+\gamma}\Big[1+\frac{\delta\gamma}{\eta}\Big]. \tag{7.2}$$

While both the models (6.1) and (7.1) contain bifurcations, there is an important difference. In (6.1) the splitting between the compartments I and A comes at the end of the stay in the compartment L, while in (7.1), some individuals are removed from the compartment I during their stay and are sent to the compartment T, while others remain in the compartment I until the end of their stay and then proceed to the compartment R.

We now extend the model (7.1) to an age of infection model with general infective and treatment stage distributions. Assume that the distribution of infective periods is given by $P(\tau)$, and the distribution of periods in treatment is given by $Q(\tau)$. Then the $SITR$ model becomes

$$\begin{aligned} S'(t) &= -\beta(N)S(t)[I(t)+\delta T(t)],\\ I(t) &= I_0(t)+\int_0^t [-S'(t-\tau)]\mathrm{e}^{-\gamma\tau}P(\tau)\,\mathrm{d}\tau,\\ T(t) &= \int_0^t \gamma I(t-\sigma)Q(\sigma)\,\mathrm{d}\sigma. \end{aligned} \tag{7.3}$$

Then

$$\varphi(t) = I(t)+\delta T(t).$$

We see from the second equation of (7.3) that the contribution to $\mathcal{R}_0$ from $I(t)$ is

$$\beta N\int_0^\infty \mathrm{e}^{-\gamma\tau}P(\tau)\,\mathrm{d}\tau.$$

To find the contribution from $T(t)$, we need to write the equation in the form

$$T(t) = \int_0^t [-S'(t-\tau)]Y(\tau)\,\mathrm{d}\tau,$$

so that the contribution from $T(t)$ would be

$$\delta\beta N\int_0^\infty Y(\tau)\,\mathrm{d}\tau$$

and we would obtain

$$\mathcal{R}_0 = \beta N\Big[\int_0^\infty \mathrm{e}^{-\gamma\tau}P(\tau)\,\mathrm{d}\tau+\delta\int_0^\infty Y(\tau)\,\mathrm{d}\tau\Big].$$

We rewrite $T(t)$ to find $Y(\tau)$. Using (7.3), we obtain

$$T(t) = \int_0^t \gamma I(t-\sigma)Q(\sigma)\,\mathrm{d}\sigma = \int_0^t [-S'(t-\tau)]B(\tau)\,\mathrm{d}\tau$$

with

$$B(\tau) = \gamma \int_0^\tau \mathrm{e}^{-\gamma(\tau-\sigma)} P(\tau-\sigma) Q(\sigma)\,\mathrm{d}\sigma.$$

Now, we have

$$\begin{aligned}
\int_0^\infty B(\tau)\,\mathrm{d}\tau &= \gamma \int_0^\infty \int_0^\tau \mathrm{e}^{-\gamma(\tau-\sigma)} P(\tau-\sigma) Q(\sigma)\,\mathrm{d}\sigma\,\mathrm{d}\tau \\
&= \gamma \int_0^\infty \int_\sigma^\infty \mathrm{e}^{-\gamma(\tau-\sigma)} P(\tau-\sigma)\,\mathrm{d}\tau\, Q(\sigma)\,\mathrm{d}\sigma \\
&= \gamma \int_0^\infty \int_0^\infty \mathrm{e}^{-\gamma\omega} P(\omega)\,\mathrm{d}\omega\, Q(v)\,\mathrm{d}v \\
&= \gamma \int_0^\infty \mathrm{e}^{-\gamma\omega} P(\omega)\,\mathrm{d}\omega \int_0^\infty Q(\sigma)\,\mathrm{d}\sigma. \qquad (7.4)
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathcal{R}_0 &= \beta N \int_0^\infty [A(\tau) + \delta B(\tau)]\,\mathrm{d}\tau \\
&= \beta N \Big[\int_0^\infty \mathrm{e}^{-\gamma\tau} P(\tau)\,\mathrm{d}\tau + \delta\gamma \int_0^\infty \mathrm{e}^{-\gamma\tau} P(\tau)\,\mathrm{d}\tau \int_0^\infty Q(\tau)\,\mathrm{d}\tau \Big] \\
&= \beta N \int_0^\infty \mathrm{e}^{-\gamma\tau} P(\tau)\,\mathrm{d}\tau \Big[1 + \delta\gamma \int_0^\infty Q(\tau)\,\mathrm{d}\tau \Big]. \qquad (7.5)
\end{aligned}$$

With exponentially distributed infective and treatment periods, $P(\tau) = \mathrm{e}^{-\alpha\tau}$, $Q(\tau) = \mathrm{e}^{-\eta\tau}$, (7.5) gives

$$\mathcal{R}_0 = \beta N \int_0^\infty \mathrm{e}^{-(\alpha+\gamma)\tau}\,\mathrm{d}\tau \Big[1 + \delta\gamma \int_0^\infty \mathrm{e}^{-\eta\tau}\,\mathrm{d}\tau \Big] = \frac{\beta N}{\alpha+\gamma} \Big[1 + \frac{\delta\gamma}{\eta} \Big],$$

the same result as (7.2).

An arbitrary choice of treatment period distribution with mean $\frac{1}{\eta}$ does not affect the quantity $\mathcal{R}_0$, but different infective period distributions may have a significant effect. For example, let us take $\gamma = 1$ and assume that the mean infective period is 1. Then, with an exponential distribution, $P(\tau) = \mathrm{e}^{-\tau}$, we have

$$\int_0^\infty \mathrm{e}^{-\tau} P(\tau)\,\mathrm{d}\tau = \int_0^\infty \mathrm{e}^{-2\tau}\,\mathrm{d}\tau = \frac{1}{2}.$$

With an infective period of fixed length 1, we have

$$\int_0^\infty \mathrm{e}^{-\tau} P(\tau)\,\mathrm{d}\tau = \int_0^1 \mathrm{e}^{-\tau}\,\mathrm{d}\tau = (1 - \mathrm{e}^{-1}) = 0.632.$$

Thus a model with an infective period of fixed length would lead to a basic reproduction number more than 25% higher than a model with an exponentially distributed infective period that has the same mean.

8 Multi-stage Treatment Models

Epidemic models may have multiple infective stages and multiple treatment stages. For example, the *SARS* model of [13] has treatment (quarantine) of exposed members and treatment (isolation) of infective members.

We consider a model with n infective stages $I_1, I_2, \cdots, I_n$ and n treatment stages $T_1, T_2, \cdots, T_n$. A fraction γ_i of members of I_i is transferred to T_i in unit time. Treated individuals pass from T_i to T_{i+1}. We assume distributions P_i in I_i and Q_i in T_i, and we assume relative infectivity parameters ε_i in the infective stages and δ_i in the treatment stages. Then

$$S'(t) = -\beta S(t)\varphi(t)$$

with

$$\varphi(t) = \sum_{i=1}^{2}[\varepsilon_i I_i + \delta_i T_i].$$

Then the basic reproduction number for an age-of-infection model with n infective and treatment stages is given by [28] as

$$\mathcal{R}_0 = \beta N \sum_{i=1}^{n} \Big[\int_0^\infty \varepsilon_i A_i(\tau)\,\mathrm{d}\tau + \delta_i \int_0^\infty B_i(\tau)\,\mathrm{d}\tau \Big], \tag{8.1}$$

with

$$\int_0^\infty A_i(\tau)\,\mathrm{d}\tau = \int_0^\infty \mathrm{e}^{-\gamma_i\tau} P_i(\tau)\,\mathrm{d}\tau \Big[1 - \int_0^\infty \gamma_{i-1}\mathrm{e}^{-\gamma_{i-1}\tau} P_{i-1}(\tau)\,\mathrm{d}\tau\Big], \tag{8.2}$$

$$\int_0^\infty B_i(\tau)\,\mathrm{d}\tau = \int_0^\infty Q_i(\tau)\,\mathrm{d}\tau \Big[\int_0^\infty \gamma_i A_i(\tau)\,\mathrm{d}\tau + \int_0^\infty \gamma_{i-1} A_{i-1}(\tau)\,\mathrm{d}\tau\Big], \tag{8.3}$$

and taking $\gamma_0 = 0$.

The contributions to $\mathcal{R}_0$ of treatment stages depend only on the mean periods of the stages, not on the form of the distribution. The contributions of infective stages, beginning with the first stage from which members are removed for treatment, depend on the form of the distributions as well as the mean periods of these stages.

There is good reason to believe that infective periods are usually not exponentially distributed (see [11, 12, 19, 27]). It has been suggested that gamma distributions may be substantially better approximations. Thus the formulae (8.1)–(8.3) are essential for calculating reproduction numbers.

We have established the following result.

Theorem 8.1 *For the general age-of-infection treatment model, the basic reproduction number is given by* (8.1) *with the integrals given recursively by* (8.2) *and* (8.3).

This result simplifies the calculation of the basic reproduction number for an age-of-infection treatment model considerably as it eliminates the need to explicitly calculate the age-of-infection kernel $A(\tau)$.

9 A Multi-group Age of Infection Model

The models we have been considering up to now assume homogeneous mixing of members of the population. It is more realistic to include some heterogeneity, and we describe this by dividing the population into two subgroups with different contact rates. Our description generalizes easily to any finite number of subgroups and even to a continuous distribution of

subgroups (see [4]), but we restrict ourselves to two subgroups for simplicity and clarity. We divide the population into two subpopulations, having sizes $N_1(t)$, $N_2(t)$ respectively at time t, each divided into susceptibles, infectives, and removed members indicated by subscripts. Division into subgroups could be by any kind of heterogeneity. We assume that there is no permanent movement between groups and that there are no disease deaths, so that $N_i(t)$ is a constant function N_i of t for $i = 1, 2$.

In this section, we assume that the number of contacts per member in unit time is a constant a rather than a constant βN in order to simplify the notation for mixing. Suppose that each group i member makes a_i contacts sufficient to transmit infection in unit time. Assume that the fraction of contacts made by a member of group i that is with a member of group j is p_{ij}. Then

$$p_{11} + p_{12} = p_{21} + p_{22} = 1.$$

The total number of contacts made in unit time by members of group 1 with members of group 2 is $a_1 p_{12} N_1$, and because this must equal the total number of contacts by members of group 2 with members of group 1, there is a balance relation

$$a_1 p_{12} N_1 = a_2 p_{21} N_2.$$

It is convenient to use the notations

$$g(\infty) = \lim_{t\to\infty} g(t), \quad \widehat{g} = \int_0^\infty g(t)\mathrm{d}t$$

for any non-negative function g defined on $0 \le t < \infty$.

A two-group age of infection model with general mixing would be

$$\begin{aligned}
S_1'(t) &= -a_1 S_1 \Big[p_{11}\frac{\varphi_1(t)}{N_1} + p_{12}\frac{\varphi_2(t)}{N_2}\Big],\\
\varphi_1(t) &= \varphi_1^{(0)}(t) + \int_0^t [-S_1'(t-\tau)]A_1(\tau)\mathrm{d}\tau,\\
S_2'(t) &= -a_2 S_2 \Big[p_{21}\frac{\varphi_1(t)}{N_1} + p_{22}\frac{\varphi_2(t)}{N_2}\Big],\\
\varphi_2(t) &= \varphi_2^{(0)}(t) + \int_0^t [-S_2'(t-\tau)]A_2(\tau)\mathrm{d}\tau,
\end{aligned} \tag{9.1}$$

where $\varphi_i(t)$ is the infectivity in group i at time t, $\varphi_i^{(0)}(t)$ is the infectivity at time t of members of group i who were infected at time 0, and $A_i(\tau)$ is the average infectivity of members of group i with infection age τ.

The infectivity of an infected member of group 2 with infection age τ towards a susceptible member of group 1 is

$$a_1 p_{12} A_2(\tau).$$

One type of mixing between groups is proportionate mixing, that is, the number of contacts between groups is proportional to the relative activity levels. In other words, mixing is random

but constrained by the activity levels (see [24]). With proportionate mixing p_{ij} is independent of i and we may write

$$p_{ij} = p_j, \quad p_1 + p_2 = 1.$$

The assumption of proportionate mixing implies that the next generation operator, in the sense of [5, 6], is separable and the basic reproduction number is

$$\mathcal{R}_0 = p_1 a_1 \widehat{A}_1 + p_2 a_2 \widehat{A}_2.$$

If the mixing is not proportionate, the next generation operator is not separable and the calculation of $\mathcal{R}_0$ is much more difficult. However, it is still possible to obtain a system of final size relations.

Integration of the equation for $\frac{S_i'(t)}{S_i(t)}$ in (9.1) gives

$$\ln \frac{S_i(0)}{S_i(\infty)} = a_i \int_0^\infty \sum_{j=1}^2 p_{ij} \frac{\varphi_j(t)}{N_j} \mathrm{d}t + a_i \int_0^\infty \sum_{j=1}^2 p_{ij} \frac{\int_0^t [-S_j'(t-\tau)] A_j(\tau) \mathrm{d}\tau}{N_j} \mathrm{d}t. \tag{9.2}$$

If all initial infectives have infection age zero at time $t = 0$,

$$\varphi_j^{(0)}(t) = A_j(t)[N_j - S_j(0)],$$

and we may rewrite (9.2) as the final size system

$$\ln \frac{S_i(0)}{S_i(\infty)} = a_i \sum_{j=1}^2 p_{ij} \Big[1 - \frac{S_j(\infty)}{N_j}\Big] \widehat{A}_j. \tag{9.3}$$

If there are initial infectives with positive infection age, the final size system contains an initial term and takes the form

$$\ln \frac{S_i(0)}{S_i(\infty)} = a_i \sum_{j=1}^2 p_{ij} \Big[1 - \frac{S_j(\infty)}{N_j}\Big] \widehat{A}_j - \Gamma_i \tag{9.4}$$

with

$$\Gamma_i = a_i \sum_{j=1}^2 \frac{p_{ij}}{N_j} \int_0^\infty [A_j(t)(N_j - S_j(0) - \varphi_j^{(0)}(t)] \mathrm{d}t \geq 0.$$

The system of equations (9.4) has a unique solution $(S_1(\infty), S_2(\infty))$.

The final size relation takes a simpler form if the mixing is proportionate. With proportionate mixing, since p_{ij} is independent of i,

$$\frac{1}{a_i} \ln \frac{S_i(0)}{S_i(\infty)} = \frac{1}{a_j} \ln \frac{S_j(0)}{S_j(\infty)}.$$

This enables us to write $S_2(\infty)$ on the right-hand side of the final size relation in terms of $S_1(\infty)$, and gives the final size system as an equation for $S_1(\infty)$.

It is not difficult to extend the results of this section to models with any finite number of groups.

10 Conclusions

The age of infection epidemic model of Kermack and McKendrick [18] contains compartmental models including staged progression, differential infectivity, and treatment. The final size relation, linking the reproduction number and the epidemic final size gives a convenient way of calculating the epidemic size without the necessity of solving the model analytically. It gives an exact solution if there are no disease deaths and a close approximation if the disease death rate is small. The extension to multi-group models provides a simple way to include heterogeneity in an epidemic model without the computational difficulties of more detailed network models, and may lead the way to models more accurate than simple compartmental models but less complicated than full network models.

References

[1] Arino, J., Brauer, F., van den Driessche, et al, A final size relation for epidemic models, *Math. Biosc. Eng.*, **4**, 2007, 159–176.

[2] Arino, J., Brauer, F., van den Driessche, et al, Simple models for containment of a pandemic, *J. Roy. Soc. Interface*, **3**, 2006, 453–457.

[3] Brauer, F., Age-of-infection and the final size relation, *Math. Biosc. Eng.*, **5**, 2008, 681–690.

[4] Brauer, F. and Watmough, J., Age of infection epidemic models with heterogeneous mixing, *J. Biol. Dyn.*, **3**, 2009, 324–330.

[5] Diekmann, O. and Heesterbeek, J. A. P., Mathematical Epidemiology of Infectious Diseases, Wiley, Chichester, 2000.

[6] Diekmann, O., Heesterbeek, J. A. P. and Metz, J. A. J., On the definition and the computation of the basic reproductive ratio $\mathcal{R}_0$ in models for infectious diseases in heterogeneous populations, *J. Math. Biol.*, **28**, 1990, 365–382.

[7] Dietz, K., Overall patterns in the transmission cycle of infectious disease agents, Population Biology of Infectious Diseases, R. M. Anderson and R. M. May (eds.), Life Sciences Research Report, **25**, Springer-Verlag, Berlin, Heidelberg, New York, 1982, 87–102.

[8] Erdös, P. and Rényi, A., On random graphs, *Publicationes Mathematicae*, **6**, 1959, 290–297.

[9] Erdös, P. and Rényi, A., On the evolution of random graphs, *Pub. Math. Inst. Hung. Acad. Science*, **5**, 1960, 17–61.

[10] Erdös, P. and Rényi, A., On the strengths of connectedness of a random graph, *Acta Math. Scientiae Hung.*, **12**, 1961, 261–267.

[11] Feng, Z., Final and peak epidemic sizes for $SEIR$ models with quarantine and isolation, *Math. Biosc. Eng.*, **4**, 2007, 675–686.

[12] Feng, Z., Xu, D. and Zhao, W., Epidemiological models with non-exponentially distributed disease stages and applications to disease control, *Bull. Math. Biol.*, **69**, 2007, 1511–1536.

[13] Gumel, A., Ruan, S., Day, T., et al, Modeling strategies for controlling SARS outbreaks based on Toronto, Hong Kong, Singapore and Beijing experience, *Proc. Roy. Soc. London*, **271**, 2004, 2223–2232.

[14] Heesterbeek, J. A. P. and Metz, J. A. J., The saturating contact rate in marriage and epidemic models, *J. Math. Biol.*, **31**, 1993, 529–539.

[15] Heffernan, J. M. Smith, R. J. and Wahl, L. M., Perspectives on the basic reproductive ratio, *J. Roy. Soc. Interface*, **2**, 2005, 281–293.

[16] Hyman, J. M. and Li, J., Infection-age structured epidemic models with behavior change or treatment, *J. Biol. Dyn.*, **1**, 2007, 109–131.

[17] Hyman, J. M., Li, J. and Stanley, E. A., The differential infectivity and staged progression models for the transmission of HIV, *Math. Biosci.*, **155**, 1999, 77–109.

[18] Kermack, W. O. and GMcKendrick, A. G., A contribution to the mathematical theory of epidemics, *Proc. Royal Soc. London*, **115**, 1927, 700–721.

[19] Lloyd, A. L., Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics, *Theor. Pop. Biol.*, **60**, 2001, 59–71.

[20] Mena-Lorca, J. and Hethcote, H. W., Dynamic models of infectious diseases as regulators of population size, *J. Math. Biol.*, **30**, 1992, 693–716.

[21] Newman, M. E. J., The spread of epidemic disease on networks, *Phys. Rev. E*, **66**, 2002, 016128.

[22] Newman, M. E. J., The structure and function of complex networks, *SIAM Review*, **45**, 2003, 167–256.

[23] Newman, M. E. J., Strogatz, S. H. and Watts, D. J., Random graphs with arbitrary degree distributions and their applications, *Phys. Rev. E*, **64**, 2001, 026118.

[24] Nold, A., Heterogeneity in disease transmission modeling, *Math. Biosc.*, **52**, 1980, 227–240.

[25] Riley, S., Fraser, C., Donnelly, C. A., et al, Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions, *Science*, **300**, 2003, 1961–1966.

[26] Strogatz, S. H., Exploring complex networks, *Nature*, **410**, 2001, 268–276.

[27] Wearing, H. J., Rohani, P. and Keeling, M. J., Appropriate models for the management of infectious diseases, *PLOS Medicine*, **2**, 2005, 621–627.

[28] Yang, C. K. and Brauer, F., Calculation of $\mathcal{R}_0$ for age-of-infection models, *Math. Biosc. Eng.*, **5**, 2008, 585–599.