

Simple models of invasions and epidemics

Fred Brauer

April 26, 2009

Introduction

We think of a pathogen invading a population and causing a disease outbreak. Often, this “invasion” is an infected person coming into the population from outside, but it could also come from a member of the population who has been infected while visiting outside the population and has then returned. There are three different possibilities for the development of the disease: (i) The infection may be transmitted only to a small number of people and may then die out. (ii) The infection may develop into an epidemic affecting a substantial fraction of the population and may then die out. (iii) The infection may become endemic, that is, it may become established in the population so that there is always a positive fraction of the population infected. We will begin by describing models for epidemics and then later we will discuss endemic situations. There is a threshold phenomenon in both epidemic models and models for endemic situations between disappearance of the infection without much effect and spread of the disease to a substantial portion of the population.

These notes are intended as an introduction to mathematical epidemiology. For a more detailed description we refer readers to such sources as [5].

Chapter 1

Epidemic Models

1.1 The simple Kermack-McKendrick epidemic model

An epidemic, which acts on a short temporal scale, is a sudden outbreak of a disease that infects a substantial portion of the population in a region before it disappears. Epidemics usually leave many members untouched. Historically these attacks have sometimes recurred with intervals of several years between outbreaks, possibly diminishing in severity as populations develop some immunity.

Throughout history, epidemics have had major effects on the course of events. For example, the Black Death, now identified as probably having been the bubonic plague which had actually invaded Europe as early as the 6th century, spread from Asia throughout Europe in several waves during the 14th century, beginning in 1346, and is estimated to have caused the death of as much as one third of the population of Europe between 1346 and 1350. The disease recurred regularly in various parts of Europe for more than 300 years, notably as the Great Plague of London of 1665–1666. It then gradually withdrew from Europe.

More than 15% of the population of London died in the Great Plague. (1665–1666). It appeared quite suddenly, grew in intensity, and then disappeared, leaving part of the population untouched. One of the early triumphs of mathematical epidemiology [23] was the formulation of a simple model that predicted behaviour very similar to this behaviour, observed in countless epidemics. 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.

There are many questions of interest to public health physicians confronted with a possible epidemic. For example, how many individuals will be infected and require treatment? What is the maximum number of people needing care at any particular time? How long will the epidemic last? How much good would

quarantine or isolation of victims do in reducing the severity of the epidemic? Since experiments on epidemics are not possible, mathematical models are the only feasible approaches to such questions.

We formulate our descriptions as *compartmental models*, with the population under study being divided into compartments and with assumptions about the nature and time rate of transfer from one compartment to another. Diseases that confer immunity have a different compartmental structure from diseases without immunity. We will use the terminology *SIR* to describe a disease which confers immunity against re-infection, to indicate that the passage of individuals is from the susceptible class *S* to the infective class *I* to the removed class *R*. Epidemics are usually diseases of this type. We would use the terminology *SIS* to describe a disease with no immunity against re-infection, to indicate that the passage of individuals is from the susceptible class to the infective class and then back to the susceptible class. Usually, diseases caused by a virus are of *SIR* type while diseases caused by bacteria are of *SIS* type.

In addition to the basic distinction between diseases for which recovery confers immunity against reinfection and diseases for which recovered members are susceptible to reinfection, and the intermediate possibility of temporary immunity signified by a model of *SIRS* type, more complicated compartmental structure is possible. For example, there are *SEIR* and *SEIS* models, with an exposed period between being infected and becoming infective.

When there are only a few infected members the start of a disease outbreak depends on random contacts between small numbers of individuals. In the next section we will use this to describe an approach to the study of the beginning of a disease outbreak by means of branching processes, but we begin with a description of deterministic compartmental models.

The independent variable in our compartmental models is the time t and the rates of transfer between compartments are expressed mathematically as derivatives with respect to time of the sizes of the compartments, and as a result our models are formulated initially as *differential equations*. In order to model such an epidemic we divide the population being studied into three classes labeled *S*, *I*, and *R*. We let $S(t)$ denote the number of individuals who are susceptible to the disease, that is, who are not (yet) infected at time t . $I(t)$ denotes the number of infected individuals, assumed infectious and able to spread the disease by contact with susceptibles. $R(t)$ denotes the number of individuals who have been infected and then removed from the possibility of being infected again or of spreading infection, either through recovery from the disease with full immunity against reinfection or through death caused by the disease. These characterizations of removed members are different from an epidemiological perspective but may be equivalent from a modeling point of view which takes into account only the state of an individual with respect to the disease.

In formulating models in terms of the derivatives of the sizes of each compartment we are assuming that the number of members in a compartment is a differentiable function of time. This assumption is plausible once a disease outbreak has become established but is not valid at the beginning of a dis-

ease outbreak when there are only a few infectives. In the next section we will describe a different approach for the initial stage of a disease outbreak.

In formulating models as differential equations, we are assuming that the epidemic process is *deterministic*, that is, that the behaviour of a population is determined completely by its history and by the rules which describe the model.

The basic compartmental models to describe the transmission of communicable diseases are contained in a sequence of three papers by W.O. Kermack and A.G. McKendrick in 1927, 1932, and 1933 [23, 24, 25]. The first of these papers described epidemic models. What is often called the Kermack-McKendrick epidemic model is actually a special case of the general model introduced in this paper. The general model included dependence on age of infection, that is, the time since becoming infected. Age of infection models have become important in the study of HIV/AIDS. We will return to them in Section 1.4 in order to provide a unified approach to compartmental epidemic models.

The special case of the model proposed by Kermack and McKendrick in 1927 which is the starting point for our study of epidemic models is

$$\begin{aligned} S' &= -\beta SI \\ I' &= \beta SI - \alpha I \\ R' &= \alpha I . \end{aligned} \tag{1.1}$$

A flow chart is shown in Figure 1.1. It is based on the following assumptions:

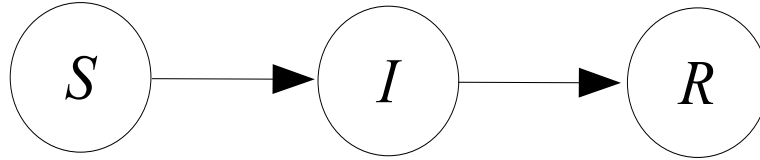


Figure 1.1: Flow chart for the *SIR* model

- (i) An average member of the population makes contact sufficient to transmit infection with βN others per unit time, where N represents total population size (mass action incidence).
- (ii) Infectives leave the infective class at rate αI per unit time.
- (iii) There is no entry into or departure from the population, except possibly through death from the disease.
- (iv) There are no disease deaths, and the total population size is a constant N .

According to (i), since the probability that a random contact by an infective is with a susceptible, who can then transmit infection, is S/N , the number of

new infections in unit time per infective is $(\beta N)(S/N)$, giving a rate of new infections $(\beta N)(S/N)I = \beta SI$. Alternately, we may argue that for a contact by a susceptible the probability that this contact is with an infective is I/N and thus the rate of new infections per susceptible is $(\beta N)(I/N)$, giving a rate of new infections $(\beta N)(I/N)S = \beta SI$. Note that both approaches give the same rate of new infections; in models with more complicated compartmental structure one may be more appropriate than the other.

We need not give an algebraic expression for N since it cancels out of the final model, but we should note that for an *SIR* disease model $N = S + I + R$. Later, we will allow the possibility that some infectives recover while others die of the disease. The hypothesis (iii) really says that the time scale of the disease is much faster than the time scale of births and deaths so that demographic effects on the population may be ignored. An alternative view is that we are only interested in studying the dynamics of a single epidemic outbreak.

The assumption (ii) requires a fuller mathematical explanation, since the assumption of a recovery rate proportional to the number of infectives has no clear epidemiological meaning. We consider the ‘‘cohort’’ of members who were all infected at one time and let $u(s)$ denote the number of these who are still infective s time units after having been infected. If a fraction α of these leave the infective class in unit time then

$$u' = -\alpha u ,$$

and the solution of this elementary differential equation is

$$u(s) = u(0) e^{-\alpha s} .$$

Thus, the fraction of infectives remaining infective s time units after having become infective is $e^{-\alpha s}$, so that the length of the infective period is distributed exponentially with mean $\int_0^\infty e^{-\alpha s} ds = 1/\alpha$, and this is what (ii) really assumes. If we assume, instead of (ii), that the fraction of infectives remaining infective a time τ after having become infective is $P(\tau)$, the second equation of (1.1) would be replaced by the integral equation

$$I(t) = I_0(t) + \int_0^\infty \beta S(t - \tau) I(t - \tau) P(\tau) d\tau,$$

where $I_0(t)$ represents the members of the population who were infective at time $t = 0$ and are still infective at time t .

The assumptions of a rate of contacts proportional to population size N with constant of proportionality β , and of an exponentially distributed recovery rate are unrealistically simple. More general models can be constructed and analyzed, but our goal here is to show what may be deduced from extremely simple models. It will turn out that that many more realistic models exhibit very similar qualitative behaviours.

In our model R is determined once S and I are known, and we can drop the

R equation from our model, leaving the system of two equations

$$\begin{aligned} S' &= -\beta SI \\ I' &= (\beta S - \alpha)I, \end{aligned} \tag{1.2}$$

together with initial conditions

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

We think of introducing a small number of infectives into a population of susceptibles and ask whether there will be an epidemic. We remark that the model makes sense only so long as $S(t)$ and $I(t)$ remain non-negative. Thus if either $S(t)$ or $I(t)$ reaches zero we consider the system to have terminated. We observe that $S' < 0$ for all t and $I' > 0$ if and only if $S > \alpha/\beta$. Thus I increases so long as $S > \alpha/\beta$ but since S decreases for all t , I ultimately decreases and approaches zero. If $S_0 < \alpha/\beta$, I decreases to zero (no epidemic), while if $S_0 > \alpha/\beta$, I first increases to a maximum attained when $S = \alpha/\beta$ and then decreases to zero (epidemic).

The quantity $\beta S_0/\alpha$ is a threshold quantity, called the *basic reproduction number* and denoted by \mathcal{R}_0 , which determines whether there is an epidemic or not. If $\mathcal{R}_0 < 1$ the infection dies out, while if $\mathcal{R}_0 > 1$ there is an epidemic. The definition of the basic reproduction number \mathcal{R}_0 is that the basic reproduction number is the number of secondary infections caused by a single infective introduced into a wholly susceptible population of size $N \approx S_0$ over the course of the infection of this single infective. In this situation, an infective makes βN contacts in unit time, all of which are with susceptibles and thus produce new infections, and the mean infective period is $1/\alpha$; thus the basic reproduction number is actually $\beta N/\alpha$ rather than $\beta S_0/\alpha$.

Since (1.2) is a two-dimensional autonomous system of differential equations, the natural approach would be to find equilibria and linearize about each equilibrium to determine its stability. However, since every point with $I = 0$ is an equilibrium, the system (1.2) has a line of equilibria and this approach is not applicable (the linearization matrix at each equilibrium has a zero eigenvalue).

Fortunately, there is an alternate approach which enables us to analyze the system (1.2). The sum of the two equations of (1.2) is

$$(S + I)' = -\alpha I.$$

Thus $S + I$ is a non-negative smooth decreasing function and therefore tends to a limit as $t \rightarrow \infty$. Also, it is not difficult to prove that the derivative of a smooth decreasing function must tend to zero, and this shows that

$$I_\infty = \lim_{t \rightarrow \infty} I(t) = 0.$$

Thus $S + I$ has limit S_∞ .

Integration of the sum of the two equations of (1.2) from 0 to ∞ gives

$$\alpha \int_0^\infty (S(t) + I(t)) dt = S_0 + I_0 - S_\infty = N - S_\infty.$$

Division of the first equation of (1.2) by S and integration from 0 to ∞ gives

$$\begin{aligned} \ln \frac{S_0}{S_\infty} &= \beta \int_0^\infty I(t) dt \\ &= \frac{\beta}{\alpha} [N - S_\infty] \\ &= \mathcal{R}_0 \left[1 - \frac{S_\infty}{N} \right]. \end{aligned} \tag{1.3}$$

The equation (1.3) is called the *final size relation*. It gives a relation between the basic reproduction number and the size of the epidemic. 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* $(1 - S_\infty/N)$. In addition, since the right side of (1.3) is finite, the left side is also finite, and this shows that $S_\infty > 0$. The final size relation (1.3) is valid for a large variety of epidemic models, as we shall see in later sections.

It is generally difficult to estimate the contact rate β which depends on the particular disease being studied but may also depend on social and behavioural factors. The quantities S_0 and S_∞ may be estimated by serological studies (measurements of immune responses in blood samples) before and after an epidemic, and from these data the basic reproduction number \mathcal{R}_0 may be estimated by using (1.3). This estimate, however, is a retrospective one which can be derived only after the epidemic has run its course.

Initially, the number of infectives grows exponentially because the equation for I may be approximated by

$$I' = (\beta N - \alpha)I$$

and the initial growth rate is

$$r = \beta N - \alpha = \alpha(\mathcal{R}_0 - 1).$$

This initial growth rate r may be determined experimentally when an epidemic begins. Then since N and α may be measured β may be calculated as

$$\beta = \frac{r + \alpha}{N}.$$

However, because of incomplete data and under-reporting of cases this estimate may not be very accurate. This inaccuracy is even more pronounced for an outbreak of a previously unknown disease, where early cases are likely to be misdiagnosed. Because of the final size relation, estimation of β or \mathcal{R}_0 is an important question that has been studied by a variety of approaches.

Before examining compartmental models which go beyond the simple Kermack-McKendrick model, we look at the shortcomings of the model for describing the beginning of a disease outbreak.

1.2 A branching process disease outbreak model

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 and 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 will be 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.

There is an important difference between the behaviour of branching process models and the behaviour of models of Kermack-McKendrick type, namely that for a stochastic disease outbreak model if $\mathcal{R}_0 < 1$ the probability that the infection will die out is 1, but if $\mathcal{R}_0 > 1$ there is a positive probability that the infection will increase initially but will produce only a minor outbreak and will die out before triggering a major epidemic.

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 Erdős and Rényi of the 1950's and 1960's [11, 12, 13]. It has become important in many areas of application, including social contacts and computer networks, as well as the spread of communicable diseases. We will think of networks as bi-directional, with disease transmission possible in either direction along an edge.

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 of [9] and then develops along the lines of [6, 30, 32]. 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.

It is convenient to 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 $0 \leq z \leq 1$, and may be differentiated term by term. Thus

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

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

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

More generally, we define the moments

$$\langle k^j \rangle = \sum_{k=1}^{\infty} k^j p_k, \quad j = 1, 2, \dots, \infty.$$

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 $k p_k$ 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

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

We let $\mathcal{R}_0 = G_1'(1)$, 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 .

Our next goal is to calculate the probability that the infection will die out and will not develop into a major epidemic. We begin by assuming that patient zero is a vertex of degree k of the network. Suppose patient zero transmits infection to a vertex of degree j . We let z_n denote the probability that this infection dies out within the next n generations. For the infection to die out in n generations each of these j secondary infections must die out in $(n-1)$ generations. The probability of this is z_{n-1} for each secondary infection, and the probability that all secondary infections will die out in $(n-1)$ generations is z_{n-1}^j . Now z_n is the sum over j of these probabilities, weighted by the probability q_j of j secondary infections. Thus

$$z_n = \sum_{j=0}^{\infty} q_j z_{n-1}^j = G_1(z_{n-1}).$$

Since $G_1(z)$ is an increasing function, the sequence z_n is an increasing sequence and has a limit z_{∞} , which is the probability that this infection will die out eventually. Then z_{∞} is the limit as $n \rightarrow \infty$ of the solution of the difference equation

$$z_n = G_1(z_{n-1}), \quad z_0 = 0.$$

Thus z_{∞} must be an equilibrium of this difference equation, that is, a solution of $z = G_1(z)$. Let w be the smallest positive solution of $z = G_1(z)$. Then, because $G_1(z)$ is an increasing function of z , $z \leq G_1(z) \leq G_1(w) = w$ for $0 \leq z \leq w$. Since $z_0 = 0 < w$ and $z_{n-1} \leq w$ implies

$$z_n = G_1(z_{n-1}) \leq G_1(w) = w,$$

it follows by induction that

$$z_n \leq w, \quad n = 0, 1, \dots, \infty.$$

From this we deduce that

$$z_{\infty} = w.$$

The equation $G_1(z) = z$ has a root $z = 1$ since $G_1(1) = 1$. Because the function $G_1(z) - z$ has a positive second derivative, its derivative $G_1'(z) - 1$ is increasing

and can have at most one zero. This implies that the equation $G_1(z) = z$ has at most two roots in $0 \leq z \leq 1$. If $\mathcal{R}_0 < 1$ the function $G_1(z) - z$ has a negative first derivative

$$G'_1(z) - 1 \leq G'_1(1) - 1 = \mathcal{R}_0 - 1 < 0$$

and the equation $G_1(z) = z$ has only one root, namely $z = 1$. On the other hand, if $\mathcal{R}_0 > 1$ the function $G_1(z) - z$ is positive for $z = 0$ and negative near $z = 1$ since it is zero at $z = 1$ and its derivative is positive for $z < 1$ and z near 1. Thus in this case the equation $G_1(z) = z$ has a second root $z_\infty < 1$.

This root is the probability that an infection transmitted along one of the k edges at the initial vertex will die out. The probability that the disease outbreak will die out eventually is the sum over k of the probabilities that the initial infection in a vertex of degree k will die out, weighted by the degree distribution $\{p_k\}$ of the original infection, and this is

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

To summarize this analysis, we see that if $\mathcal{R}_0 < 1$ the probability that the infection will die out is 1. 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 an 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*

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

for some constant c [5, pp. 142-3]. The generating function for the Poisson distribution is $e^{c(z-1)}$. Then $G_1(z) = G_0(z)$, and $= c$, so that

$$G_1(z) = G_0(z) = e^{\mathcal{R}_0(z-1)}.$$

The commonly observed situation that most infectives do not pass on infection but there are a few “superspreading events” [33] 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. For example, if $\mathcal{R}_0 = 2.5$ the assumption of a Poisson distribution gives $z_\infty = 0.107$ and $G_0(z_\infty) = 0.107$, so that the probability of an epidemic is 0.893. The assumption that nine out of ten infectives do not transmit infection while the tenth transmits 25 infections gives

$$G_0(z) = (z^{25} + 9)/10, \quad G_1(z) = z^{24}, \quad z_\infty = 0, \quad G_0(z_\infty) = 0.9,$$

from which we see that the probability of an epidemic is 0.1. Another example, possibly more realistic, is to assume that a fraction $(1 - p)$ of the population follows a Poisson distribution with constant r while the remaining fraction p consists of superspreaders each of whom makes L contacts. This would give a generating function

$$\begin{aligned} G_0(z) &= (1 - p)e^{r(z-1)} + pz^L \\ G_1(z) &= \frac{r(1 - p)e^{r(z-1)} + pLz^{L-1}}{r(1 - p) + pL}, \end{aligned}$$

and

$$\mathcal{R}_0 = \frac{r^2(1 - p) + pL(L - 1)}{r(1 - p) + pL}.$$

For example, if $r = 2.2$, $L = 10$, $p = 0.01$, numerical simulation gives

$$\mathcal{R}_0 = 2.5, \quad z_\infty = 0.146,$$

so that the probability of an epidemic is 0.849.

These 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.

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.

These examples indicate that the probability of an epidemic depends strongly on the contact network at the beginning of a disease outbreak. 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 [31, 34].

In the remainder of this chapter, we assume that we are in an epidemic situation following a disease outbreak which has been modeled initially by a branching process. Thus we return to the study of compartmental models.

1.3 More complicated epidemic models

We have established that the simple Kermack-McKendrick epidemic model (1.2) has the basic properties

1. There is a basic reproduction number \mathcal{R}_0 such that if $\mathcal{R}_0 < 1$, the disease dies out while if $\mathcal{R}_0 > 1$ there is an epidemic.
2. The number of infectives always approaches zero and the number of susceptibles always approaches a positive limit as $t \rightarrow \infty$.
3. There is a relation between the reproduction number and the final size of the epidemic which is an equality if there are no disease deaths.

In fact, these properties hold for epidemic models with more complicated compartmental structure. This may be established by considering epidemic models in which infectivity is a function of the time since becoming infected (called the age of infection), which was actually the form originally described by Kermack and McKendrick. Before carrying out this general analysis we will describe some common epidemic models as examples.

1.3.1 Exposed periods

In many infectious diseases there is an exposed period after the transmission of infection from susceptibles to potentially infective members but before these potential infectives develop symptoms and can transmit infection. To incorporate an exposed period with mean exposed period $1/\kappa$ we add an exposed class E and use compartments S, E, I, R and total population size $N = S + E + I + R$ to give a generalization of the epidemic model (1.2)

$$\begin{aligned} S' &= -\beta SI \\ E' &= \beta SI - \kappa E \\ I' &= \kappa E - \alpha I. \end{aligned} \tag{1.4}$$

A flow chart is shown in Figure 1.2.

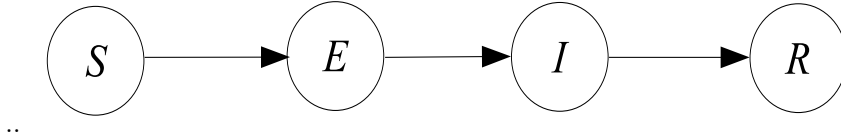


Figure 1.2: Flow chart for the $SEIR$ model

The analysis of this model is the same as the analysis of (1.2), but with I replaced by $E + I$. That is, instead of using the number of infectives as one of the variables we use the total number of infected members, whether or not they are capable of transmitting infection.

In some diseases there is some infectivity during the exposed period. This may be modeled by assuming infectivity reduced by a factor ε during the exposed period. A calculation of the rate of new infections per susceptible leads to a model

$$\begin{aligned} S' &= -\beta S(I + \varepsilon E) \\ E' &= \beta S(I + \varepsilon E) - \kappa E \\ I' &= \kappa E - \alpha I. \end{aligned} \tag{1.5}$$

We take initial conditions

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

For this model

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

Integration of the sum of the equations of (1.4) from 0 to ∞ gives

$$N - S_\infty = \alpha \int_0^\infty I(s) ds.$$

Integration of the third equation of (1.5) gives

$$\kappa \int_0^\infty E(s) ds = \alpha \int_0^\infty I(s) ds - I_0,$$

and division of the first equation of (1.5) by S followed by integration from 0 to ∞ gives

$$\begin{aligned} \ln \frac{S_0}{S_\infty} &= \int_0^\infty \beta [I(s) + \varepsilon E(s)] ds \\ &= \beta \int_0^\infty [I(s) + \varepsilon E(s)] ds \\ &= \beta \left[\varepsilon + \frac{\kappa}{\alpha} \right] \int_0^\infty E(s) ds - \frac{\varepsilon \beta I_0}{\kappa} \\ &= \mathcal{R}_0 \left[1 - \frac{S_\infty}{N} \right] - \frac{\varepsilon \beta I_0}{\kappa}. \end{aligned}$$

In this final size relation there is an initial term $\beta I_0/\alpha$, caused by the assumption that there are individuals infected originally who are beyond the exposed stage in which they would have had some infectivity. In order to obtain a final size relation without such an initial term it is necessary to assume $I(0) = 0$, that initial infectives are in the first stage in which they can transmit infection. If $I(0) = 0$, the final size relation has the form (1.3).

1.3.2 Treatment models

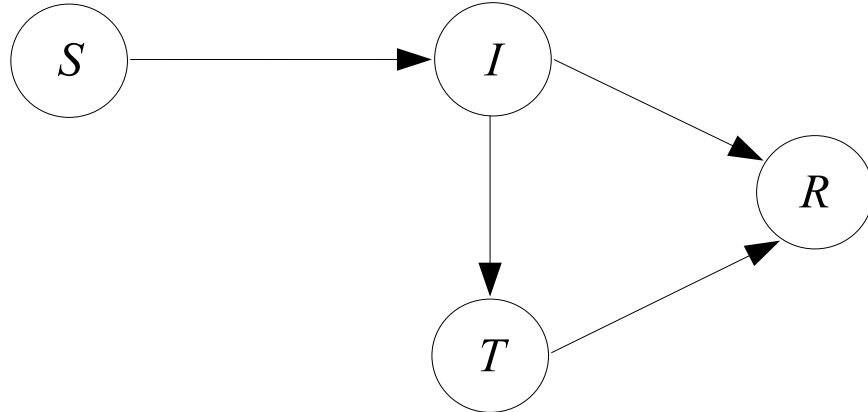
One form of treatment that is possible for some diseases is vaccination to protect against infection before the beginning of an epidemic. For example, this approach is commonly used for protection against annual influenza outbreaks. A simple way to model this would be to reduce the total population size by the fraction of the population protected against infection.

In reality such inoculations are only partly effective, decreasing the rate of infection and also decreasing infectivity if a vaccinated person does become infected. This may be modeled by dividing the population into two groups with different model parameters which would require some assumptions about the mixing between the two groups. This is not difficult but we will not explore this direction here.

If there is a treatment for infection once a person has been infected, this may be modeled this by supposing that a fraction γ per unit time of infectives is selected for treatment, and that treatment reduces infectivity by a fraction δ . Suppose that the rate of removal from the treated class is η . This leads to the *SITR* model, where T is the treatment class, given by

$$\begin{aligned} S' &= -\beta S[I + \delta T] \\ I' &= \beta S[I + \delta T] - (\alpha + \gamma)I \\ T' &= \gamma I - \eta T. \end{aligned} \tag{1.6}$$

A flow chart is shown in Figure 1.3.



..

Figure 1.3: Flow chart for the *SITR* model

It is not difficult to prove, much as was done for the model (1.2) that

$$S_\infty = \lim_{t \rightarrow \infty} S(t) > 0, \quad \lim_{t \rightarrow \infty} I(t) = \lim_{t \rightarrow \infty} T(t) = 0.$$

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 $1/(\alpha + \gamma)$. In addition, a fraction $\gamma/(\alpha + \gamma)$ of infectives are 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 $1/\eta$. Thus \mathcal{R}_0 is

$$\mathcal{R}_0 = \frac{\beta N}{\alpha + \gamma} + \frac{\gamma}{\alpha + \gamma} \frac{\delta\beta N}{\eta} \quad (1.7)$$

It is also possible to establish the final size relation (1.3) by means very similar to those used for the simple model (1.2). We integrate the first equation of (1.6) to obtain

$$\begin{aligned} \ln \frac{S_0}{S_\infty} &= \int_0^\infty \beta [I(t) + \delta T(t)] dt \\ &= \beta \int_0^\infty [I(t) + \delta T(t)] dt. \end{aligned}$$

Integration of the third equation of (1.6) gives

$$\gamma \int_0^\infty I(t) dt = \eta \int_0^\infty T(t) dt.$$

Integration of the sum of the first two equations of (1.6) gives

$$N - S_\infty = (\alpha + \gamma) \int_0^\infty I(t) dt.$$

Combination of these three equations and (1.7) gives (1.3).

1.3.3 An influenza model

In some diseases, such as influenza, at the end of a stage individuals may proceed to one of two stages. 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 $1/\eta$. The influenza model of [2, 3] is

$$\begin{aligned} S' &= -\beta S [I + \delta A] \\ L' &= \beta S [I + \delta A] - \kappa L \\ I' &= p\kappa L - \alpha I \\ A' &= (1 - p)\kappa L - \eta A \end{aligned} \quad (1.8)$$

and

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

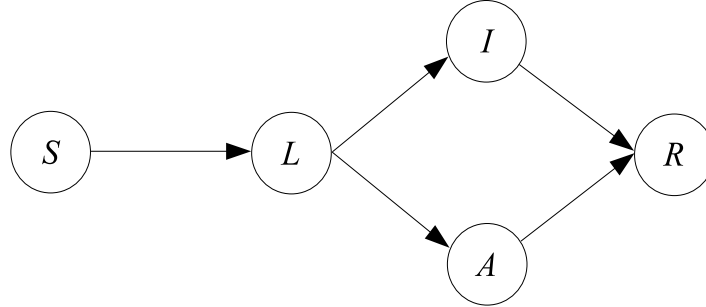


Figure 1.4: Influenza model flowchart

A flow chart is shown in Figure 1.4.

The same approach used in earlier examples leads to the same final size relation (1.3).

The model (1.8) is an example of a differential infectivity model. In such models, also used in the study of HIV/AIDS [22], 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.

1.3.4 A quarantine-isolation model

For an outbreak of a new disease, where no vaccine is available, isolation of diagnosed infectives and quarantine of people who are suspected of having been infected (usually by tracing of contacts of diagnosed infectives) are the only control measures available. We formulate a model to describe the course of an epidemic, originally introduced for modeling the SARS epidemic of 2002-3 [16], when control measures are begun under the assumptions:

1. Exposed members may be infective with infectivity reduced by a factor ε_E , $0 \leq \varepsilon_E < 1$.
2. Exposed members who are not isolated become infective at rate κ_E .
3. We introduce a class Q of quarantined members and a class H of isolated (hospitalized) members and exposed members are quarantined at a proportional rate γ_Q in unit time (in practice, a quarantine will also be applied to many susceptibles, but we ignore this in the model). Quarantine is not perfect, but reduces the contact rate by a factor ε_Q . The effect of this assumption is that some susceptibles make fewer contacts than the model assumes.

4. Infectives are diagnosed at a proportional rate γ_H per unit time and isolated. Isolation is imperfect, and there may be transmission of disease by isolated members, with an infectivity factor of ε_H .
5. Quarantined members are monitored and when they develop symptoms at rate κ_Q they are isolated immediately.
6. Infectives leave the infective class at rate α_I and isolated members leave the isolated class at rate α_H .

These assumptions lead to the *SEQIHR* model [16]

$$\begin{aligned}
S' &= -\beta S[\varepsilon_E E + \varepsilon_E \varepsilon_Q Q + I + \varepsilon_J J] \\
E' &= \beta S[\varepsilon_E E + \varepsilon_E \varepsilon_Q Q + I + \varepsilon_J J] - (\kappa_E + \gamma_Q)E \\
Q' &= \gamma_Q E - \kappa_H Q \\
I' &= \kappa_E E - (\alpha_I + \gamma_H)I \\
H' &= \kappa_Q Q + \gamma_H I - \alpha_H H.
\end{aligned} \tag{1.9}$$

The model before control measures are begun is the special case

$$\gamma_Q = \gamma_H = \kappa_Q = \alpha_H = 0, \quad Q = H = 0$$

of (1.9). It is the same as (1.5).

We define the *control reproduction number* \mathcal{R}_c to be the number of secondary infections caused by a single infective in a population consisting essentially only of susceptibles with the control measures in place. It is analogous to the basic reproduction number but instead of describing the very beginning of the disease outbreak it describes the beginning of the recognition of the epidemic. The basic reproduction number is the value of the control reproduction number with

$$\gamma_Q = \gamma_H = \kappa_Q = \alpha_H = 0.$$

We have already calculated \mathcal{R}_0 for (1.5) and we may calculate \mathcal{R}_c in the same way but using the full model with quarantined and isolated classes. We obtain

$$\mathcal{R}_c = \frac{\varepsilon_E \beta N}{D_1} + \frac{\beta N \kappa_E}{D_1 D_2} + \frac{\varepsilon_Q \varepsilon_E \beta N \gamma_Q}{D_1 \kappa_Q} + \frac{\varepsilon_H \beta N \kappa_E \gamma_H}{\alpha_H D_1 D_2} + \frac{\varepsilon_H \beta N \gamma_Q}{\alpha_H D_1},$$

where $D_1 = \gamma_Q + \kappa_E$, $D_2 = \gamma_H + \alpha_I$.

Each term of \mathcal{R}_c has an epidemiological interpretation. The mean duration in E is $1/D_1$ with contact rate $\varepsilon_E \beta$, giving a contribution to \mathcal{R}_c of $\varepsilon_E \beta N / D_1$. A fraction κ_E / D_1 goes from E to I , with contact rate β and mean duration $1/D_2$, giving a contribution of $\beta N \kappa_E / D_1 D_2$. A fraction γ_Q / D_1 goes from E to Q , with contact rate $\varepsilon_E \varepsilon_Q \beta$ and mean duration $1/\kappa_Q$, giving a contribution of $\varepsilon_E \varepsilon_Q \beta N \gamma_Q / D_1 \kappa_Q$. A fraction $\kappa_E \gamma_H / D_1 D_2$ goes from E to I to H , with a contact rate of $\varepsilon_J \beta$ and a mean duration of $1/\alpha_H$, giving a contribution of $\varepsilon_J \beta N \kappa_E \gamma_H / \alpha_H D_1 D_2$. Finally, a fraction γ_Q / D_1 goes from E to Q to H with

a contact rate of $\varepsilon_J\beta$ and a mean duration of $1/\alpha_H$ giving a contribution of $\varepsilon_J\beta N\gamma_Q/D_1\alpha_H$. The sum of these individual contributions gives \mathcal{R}_c .

In the model (1.9) the parameters γ_Q and γ_H are *control* parameters which may be chosen in the attempt to manage the epidemic. The parameters ε_Q and ε_J depend on the strictness of the quarantine and isolation processes and are thus also control measures in a sense. The other parameters of the model are specific to the disease being studied. While they are not variable, their measurements are subject to experimental error.

The linearization of (1.9) at the disease-free equilibrium $(N, 0, 0, 0, 0)$ has matrix

$$\begin{bmatrix} \varepsilon_E\beta N - (\kappa_E + \gamma_Q) & \varepsilon_E\varepsilon_Q\beta & \beta N & \varepsilon_J\beta N \\ \gamma_Q & -\kappa_Q & 0 & 0 \\ \kappa_E & 0 & -(\alpha_I + \gamma_H) & 0 \\ 0 & \kappa_Q & \gamma_H & -\alpha_H \end{bmatrix}$$

The corresponding characteristic equation is a fourth degree polynomial equation whose leading coefficient is 1 and whose constant term is a positive constant multiple of $1 - \mathcal{R}_c$, thus positive if $\mathcal{R}_c < 1$ and negative if $\mathcal{R}_c > 1$. If $\mathcal{R}_c > 1$ there is a positive eigenvalue, corresponding to an initial exponential growth rate of solutions of (1.9). If $\mathcal{R}_c < 1$ it is possible to show that all eigenvalues of the coefficient matrix have negative real part, and thus solutions of (1.9) die out exponentially [38].

In order to show that analogues of the relation (1.3) and $S_\infty > 0$ derived for the model (1.2) are valid for the management model (1.9), we begin by integrating the equations for $S + E$, Q , I , J , of (1.9) with respect to t from $t = 0$ to $t = \infty$, using the initial conditions

$$S(0) + E(0) = N(0) = N, \quad Q(0) = I(0) = H(0) = 0.$$

We continue by integrating the equation for S and then an argument similar to the one used for (1.2) but technically more complicated may be used to show that $S_\infty > 0$ for the treatment model (1.9) and also to establish the final size relation

$$\ln \frac{S_0}{S_\infty} = \mathcal{R}_c \left[1 - \frac{S_\infty}{N} \right].$$

Thus the asymptotic behaviour of the management model (1.9) is the same as that of the simpler model (1.2).

In the various compartmental models that we have studied, there are significant common features. This suggests that compartmental models can be put into a more general framework. In fact, this general framework is the age of infection epidemic model originally introduced by Kermack and McKendrick in [23].

1.4 The age of infection epidemic model

The general epidemic model described by Kermack and McKendrick [23] 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 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 $0 \leq \pi(\tau) \leq 1$ 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

$$\begin{aligned} S' &= -\beta S\varphi \\ \varphi(t) &= \varphi_0(t) + \int_0^t \beta S(t-\tau)\varphi(t-\tau)A(\tau)d\tau \\ &= \varphi_0(t) + \int_0^t [-S'(t-\tau)]A(\tau)d\tau. \end{aligned} \quad (1.10)$$

The basic reproduction number is

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

We write

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

Integration with respect to t from 0 to ∞ gives

$$\begin{aligned} \ln \frac{S_0}{S_\infty} &= \beta \int_0^\infty \varphi_0(t)dt + \beta \int_0^\infty \int_0^t [-S'(t-\tau)]A(\tau)d\tau dt \\ &= \beta \int_0^\infty \varphi_0(t)dt + \beta \int_0^\infty A(\tau) \int_\tau^\infty [-S'(t-\tau)]dt d\tau \\ &= \beta \int_0^\infty \varphi_0(t)dt + [S_0 - S_\infty] \int_0^\infty A(\tau)d\tau \\ &= \beta[N - S_\infty] \int_0^\infty A(\tau)d\tau + \beta \int_0^\infty [\varphi_0(t) - (N - S_0)A(t)]dt \\ &= \mathcal{R}_0 \left[1 - \frac{S_\infty}{N}\right] - \beta \int_0^\infty [(N - S_0)A(t) - \varphi_0(t)]dt. \end{aligned} \quad (1.11)$$

Here, $\varphi_0(t)$ is the total infectivity of the initial infectives when they reach age of infection t . If all initial infectives have infection age zero at $t = 0$, $\varphi_0(t) = [N - S_0]A(t)$, and

$$\int_0^\infty [\varphi_0(t) - (N - S_0)A(t)]dt = 0.$$

Then (1.11) takes the form

$$\ln \frac{S_0}{S_\infty} = \mathcal{R}_0 \left(1 - \frac{S_\infty}{N} \right), \quad (1.13)$$

and this is the general final size relation. If there are initial infectives with infection age greater than zero, let $u(\tau)$ be the fraction of these individuals with infection age τ , $\int_0^\infty u(\tau) d\tau = 1$. At time t these individuals have infection age $t + \tau$ and mean infectivity $A(t + \tau)$. Thus

$$\varphi_0(t) = (N - S_\infty) \int_0^\infty u(\tau) A(t + \tau) d\tau,$$

and

$$\begin{aligned} \int_0^\infty \varphi_0(t) dt &= (N - S_\infty) \int_0^\infty \int_0^\infty u(\tau) A(t + \tau) d\tau dt \\ &= (N - S_\infty) \int_0^\infty u(\tau) \left[\int_\tau^\infty A(v) dv \right] d\tau \\ &= (N - S_\infty) \int_0^\infty A(v) \left[\int_0^v u(\tau) d\tau \right] dv \\ &\leq (N - S_\infty) \int_0^\infty A(v) dv, \end{aligned}$$

since $\int_0^v u(\tau) d\tau \leq 1$.

Thus, the initial term satisfies

$$\int_0^\infty [(N - S_0) A(t) - \varphi_0(t)] dt \geq 0.$$

The final size relation is sometimes presented in the form

$$\ln \frac{S_0}{S_\infty} = \mathcal{R}_0 \left(1 - \frac{S_\infty}{S_0} \right), \quad (1.14)$$

with an initial term which is assumed small and omitted, see for example [2, 9, 17].

If the reproduction number $\mathcal{R}_0 = \mathcal{R}(\gamma)$, depends on a parameter γ , as might be the case in a treatment model, so that S_∞ is a function of γ , implicit differentiation of the final size relation with respect to γ gives

$$\left(\frac{1}{S_\infty(\gamma)} - \frac{\mathcal{R}(\gamma)}{N} \right) S'(\gamma) = -\mathcal{R}'(\gamma) \left[1 - \frac{S_\infty(\gamma)}{N} \right].$$

Since it is possible to show that $S_\infty(\gamma) < N/\mathcal{R}(\gamma)$, it follows that $\mathcal{R}'(\gamma) < 0$, then $S'_\infty(\gamma) > 0$.

The *SEIR*, *SITR*, and *SEQIHR* models described in previous sections can be formulated as age of infection models, but we will not go into this formulation.

The point is that compartmental epidemic models can be put into the age of infection framework and the general results about epidemic final size apply. This also applies to staged progression models in which there is a sequence of stages of different lengths and infectivities and individuals pass from one stage to the next.

The age of infection formulation also contains another generalization. We have been assuming an exponential distribution of stays in each compartment of a compartmental model, and the age of infection model allows arbitrary distributions. Where a rate of progression α leads to a factor $1/\alpha$ in a simple compartmental model, the assumption that the fraction of members remaining in a stage for at least time τ is a decreasing function $P(\tau)$ leads to a factor $\int_0^\infty P(\tau)d\tau$. As the length of stay in a compartment is often not exponentially distributed, this is an essential generalization [14, 15, 26, 39].

1.5 Models with disease deaths

The assumption in the model (1.2) 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 (1.2) by dropping the assumption (iv) and replacing the assumption (i) by the assumption that an average member of the population makes $C(N)$ contacts in unit time with $C'(N) \geq 0$ [7, 10], 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 that

$$\beta(N) + N\beta'(N) \geq 0. \quad (1.15)$$

Some epidemic models [10] 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 [?] 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 [28] 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

$$\begin{aligned} S' &= -\beta(N)SI \\ I' &= \beta(N)SI - \alpha I \\ N' &= -(1 - f)\alpha I. \end{aligned} \tag{1.16}$$

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 (1.16) reduces to the simpler model (1.2) with β replaced by the constant $\beta(N_0)$.

We wish to show that the model (1.16) has the same qualitative behaviour as the model (1.2), namely that there is a basic reproduction number which distinguishes between disappearance of the disease and an epidemic outbreak, and that some members of the population are left untouched when the epidemic passes. These two properties are the central features of all epidemic models.

For the model (1.16) the basic reproduction number is given by

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

because a single infective introduced into a wholly susceptible population makes $C(N_0) = N_0\beta(N_0)$ contacts in unit time, all of which are with susceptibles and thus produce new infections, and the mean infective period is $1/\alpha$.

We assume that $\beta(0)$ is finite, thus ruling out standard incidence (standard incidence does not appear to be realistic if the total population N approaches zero, and it would be more natural to assume that $C(N)$ grows linearly with N for small N). If we let $t \rightarrow \infty$ in the sum of the first two equations of (1.16) we obtain

$$\alpha \int_0^\infty I(s) ds = S_0 + I_0 - S_\infty = N - S_\infty.$$

The first equation of (1.16) may be written as

$$-\frac{S'(t)}{S(t)} = \beta(N(t))I(t).$$

Since

$$\beta(N) \geq \beta(N_0),$$

integration from 0 to ∞ gives

$$\begin{aligned} \ln \frac{S_0}{S_\infty} &= \int_0^\infty \beta(N(t))I(t)dt \\ &\geq \beta(N_0) \int_0^\infty I(t)dt \\ &= \frac{\beta(N_0)(N_0 - S_\infty)}{\alpha N_0}. \end{aligned}$$

We now obtain a final size inequality

$$\begin{aligned} \ln \frac{S_0}{S_\infty} &= \int_0^\infty \beta(N(t))I(t)dt \\ &\geq \beta(N_0) \int_0^\infty I(t)dt = \mathcal{R}_0 \left[1 - \frac{S_\infty}{N_0} \right]. \end{aligned}$$

If the disease death rate is small, the final size inequality is an approximate equality.

It is not difficult to show that $N(t) \geq fN_0$ and then a similar calculation using the inequality $\beta(N) \leq \beta(fN_0) < \infty$ shows that

$$\ln \frac{S_0}{S_\infty} \leq \beta(fN_0) \int_0^\infty I(t)dt,$$

from which we may deduce that $S_\infty > 0$.

1.6 Some warnings

An actual epidemic differs considerably from the idealized models (1.2) or (1.16). Some notable differences are:

1. When it is realized that an epidemic has begun, individuals are likely to modify their behaviour by avoiding crowds to reduce their contacts and by being more careful about hygiene to reduce the risk that a contact will produce infection.
2. If a vaccine is available for the disease which has broken out, public health measures will include vaccination of part of the population. Various vaccination strategies are possible, including vaccination of health care workers and other first line responders to the epidemic, vaccination of members of the population who have been in contact with diagnosed infectives, or vaccination of members of the population who live in close proximity to diagnosed infectives.
3. Isolation may be imperfect; in-hospital transmission of infection was a major problem in the SARS epidemic.

4. In the SARS epidemic of 2002-3 in-hospital transmission of disease from patients to health care workers or visitors because of imperfect isolation accounted for many of the cases. This points to an essential heterogeneity in disease transmission which must be included whenever there is any risk of such transmission.

Chapter 2

Models for Endemic Diseases

2.1 The *SIR* Model

Epidemics which sweep through a population attract much attention and arouse a great deal of concern. We have omitted births and deaths in our description of epidemic models because the time scale of an epidemic is generally much shorter than the demographic time scale. In effect, we have used a time scale on which the number of births and deaths in unit time is negligible.

However, there are diseases which are endemic in many parts of the world and which cause millions of deaths each year. To model a disease which may be endemic we need to think on a longer time scale and include births and deaths. A reference describing the properties of many endemic diseases is [1]. For diseases that are endemic in some region public health physicians would like to be able to estimate the number of infectives at a given time as well as the rate at which new infections arise. The effects of quarantine or vaccine in reducing the number of victims are of importance, just as in the treatment of epidemics. In addition, the possibility of defeating the endemic nature of the disease and thus controlling or even eradicating the disease in a population is worthy of study.

A model of Kermack and McKendrick [24] includes births in the susceptible class proportional to total population size and a death rate in each class proportional to the number of members in the class. This model allows the total population size to grow exponentially or die out exponentially if the birth and death rates are unequal. It is applicable to such questions as whether a disease will control the size of a population that would otherwise grow exponentially. We shall return to this topic, which is important in the study of many diseases in less developed countries with high birth rates. To formulate a model in which total population size remains bounded we could follow the approach suggested by [18] in which the total population size K is held constant by making birth

and death rates equal. Such a model is

$$\begin{aligned} S' &= -\beta SI + \mu(K - S) \\ I' &= \beta SI - \gamma I - \mu I \\ R' &= \gamma I - \mu R. \end{aligned}$$

Because $S + I + R = K$, we can view R as determined when S and I are known and consider the two-dimensional system

$$\begin{aligned} S' &= -\beta SI + \mu(K - S) \\ I' &= \beta SI - \gamma I - \mu I. \end{aligned}$$

We shall examine a slightly more general *SIR* model with births and deaths for a disease that may be fatal to some infectives. For such a disease the class R of removed members should contain only recovered members, not members removed by death from the disease. It is not possible to assume that the total population size remain constant if there are deaths due to disease; a plausible model for a disease that may be fatal to some infectives must allow the total population to vary in time. The simplest assumption to allow this is a constant birth rate Λ , but in fact the analysis is quite similar if the birth rate is a function $\Lambda(N)$ of total population size N .

Let us analyze the model

$$\begin{aligned} S' &= \Lambda - \beta SI - \mu S \\ I' &= \beta SI - \mu I - \alpha I \\ N' &= \Lambda - (1 - f)\alpha I - \mu N, \end{aligned} \tag{2.1}$$

where $N = S + I + R$, with a mass action contact rate, a constant number of births Λ per unit time, a proportional natural death rate μ in each class, and a rate of recovery or disease death α of infectives with a fraction f of infectives recovering with immunity against reinfection. In this model if $f = 1$ the total population size approaches a limit $K = \Lambda/\mu$. Then K is the carrying capacity of the population. If $f < 1$ the total population size is not constant and K represents a carrying capacity or maximum possible population size, rather than a population size. We view the first two equations as determining S and I , and then consider the third equation as determining N once S and I are known. This is possible because N does not enter into the first two equations. Instead of using N as the third variable in this model we could have used R , and the same reduction would have been possible.

If the birth or recruitment rate $\Lambda(N)$ is a function of total population size then in the absence of disease the total population size N satisfies the differential equation

$$N' = \Lambda(N) - \mu N.$$

The *carrying capacity* of population size is the limiting population size K , satisfying

$$\Lambda(K) = \mu K, \quad \Lambda'(K) < \mu.$$

The condition $\Lambda'(K) < \mu$ assures the asymptotic stability of the equilibrium population size K . It is reasonable to assume that K is the only positive equilibrium, so that

$$\Lambda(N) > \mu N$$

for $0 \leq N \leq K$. For most population models,

$$\Lambda(0) = 0, \quad \Lambda''(N) \leq 0.$$

However, if $\Lambda(N)$ represents recruitment into a behavioural class, as would be natural for models of sexually transmitted diseases, it would be plausible to have $\Lambda(0) > 0$, or even to consider $\Lambda(N)$ to be a constant function. If $\Lambda(0) = 0$, we require $\Lambda'(0) > \mu$ because if this requirement is not satisfied there is no positive equilibrium and the population would die out even in the absence of disease.

We have used a mass action contact rate for simplicity, even though a more general contact rate would give a more accurate model, just as in the epidemics considered in the preceding section. With a general contact rate and a density-dependent birth rate we would have a model

$$\begin{aligned} S' &= \Lambda(N) - \beta(N)SI - \mu S \\ I' &= \beta(N)SI - \mu I - \alpha I \\ N' &= \Lambda(N) - (1-f)\alpha I - \mu N, \end{aligned} \quad (2.2)$$

If $f = 1$, so that there are no disease deaths, the equation for N is

$$N' = \Lambda(N) - \mu N,$$

so that $N(t)$ approaches a limiting population size K . The theory of *asymptotically autonomous systems* [8, 27, 35, 36] implies that if N has a constant limit then the system is equivalent to the system in which N is replaced by this limit. Then the system (2.2) is the same as the system (2.1) with β replaced by the constant $\beta(K)$ and N by K , and $\Lambda(N)$ replaced by the constant $\Lambda(K) = \mu K$.

We shall analyze the model (2.1) qualitatively. In view of the remark above, our analysis will also apply to the more general model (2.2) if there are no disease deaths. Analysis of the system (2.2) with $f < 1$ is much more difficult. We will confine our study of (2.2) to a description without details.

The first stage of the analysis is to note that the model (2.1) is a properly posed problem. That is, since $S' \geq 0$ if $S = 0$ and $I' \geq 0$ if $I = 0$, we have $S \geq 0, I \geq 0$ for $t \geq 0$ and since $N' \leq 0$ if $N = K$ we have $N \leq K$ for $t \geq 0$. Thus the solution always remains in the biologically realistic region $S \geq 0, I \geq 0, 0 \leq N \leq K$ if it starts in this region. By rights, we should verify such conditions whenever we analyze a mathematical model, but in practice this step is frequently overlooked.

Our approach will be to identify equilibria (constant solutions) and then to determine the asymptotic stability of each equilibrium. Asymptotic stability of an equilibrium means that a solution starting sufficiently close to the equilibrium remains close to the equilibrium and approaches the equilibrium as $t \rightarrow \infty$,

while instability of the equilibrium means that there are solutions starting arbitrarily close to the equilibrium which do not approach it. To find equilibria (S_∞, I_∞) we set the right side of each of the two equations equal to zero. The second of the resulting algebraic equations factors, giving two alternatives. The first alternative is $I_\infty = 0$, which will give a disease-free equilibrium, and the second alternative is $\beta S_\infty = \mu + \alpha$, which will give an endemic equilibrium, provided $\beta S_\infty = \mu + \alpha < \beta K$. If $I_\infty = 0$ the other equation gives $S_\infty = K = \Lambda/\mu$. For the endemic equilibrium the first equation gives

$$I_\infty = \frac{\Lambda}{\mu + \alpha} - \frac{\mu}{\beta}. \quad (2.3)$$

We linearize about an equilibrium (S_∞, I_∞) by letting $y = S - S_\infty$, $z = I - I_\infty$, writing the system in terms of the new variables y and z and retaining only the linear terms in a Taylor expansion. We obtain a system of two linear differential equations,

$$\begin{aligned} y' &= -(\beta I_\infty + \mu)y - \beta S_\infty z \\ z' &= \beta I_\infty y + (\beta S_\infty - \mu - \alpha)z. \end{aligned}$$

The coefficient matrix of this linear system is

$$\begin{bmatrix} -\beta I_\infty - \mu & -\beta S_\infty \\ \beta I_\infty & \beta S_\infty - \mu - \alpha \end{bmatrix}$$

We then look for solutions whose components are constant multiples of $e^{\lambda t}$; this means that λ must be an eigenvalue of the coefficient matrix. The condition that all solutions of the linearization at an equilibrium tend to zero as $t \rightarrow \infty$ is that the real part of every eigenvalue of this coefficient matrix is negative. At the disease-free equilibrium the matrix is

$$\begin{bmatrix} -\mu & -\beta K \\ 0 & \beta K - \mu - \alpha \end{bmatrix},$$

which has eigenvalues $-\mu$ and $\beta K - \mu - \alpha$. Thus, the disease-free equilibrium is asymptotically stable if $\beta K < \mu + \alpha$ and unstable if $\beta K > \mu + \alpha$. Note that this condition for instability of the disease-free equilibrium is the same as the condition for the existence of an endemic equilibrium.

In general, the condition that the eigenvalues of a 2×2 matrix have negative real part is that the determinant be positive and the trace (the sum of the diagonal elements) be negative. Since $\beta S_\infty = \mu + \alpha$ at an endemic equilibrium, the matrix of the linearization at an endemic equilibrium is

$$\begin{bmatrix} -\beta I_\infty - \mu & -\beta S_\infty \\ \beta I_\infty & 0 \end{bmatrix} \quad (2.4)$$

and this matrix has positive determinant and negative trace. Thus, the endemic equilibrium, if there is one, is always asymptotically stable. If the quantity

$$R_0 = \frac{\beta K}{\mu + \alpha} = \frac{K}{S_\infty} \quad (2.5)$$

is less than one, the system has only the disease-free equilibrium and this equilibrium is asymptotically stable. In fact, it is not difficult to prove that this asymptotic stability is *global*, that is, that every solution approaches the disease-free equilibrium. If the quantity \mathcal{R}_0 is greater than one then the disease-free equilibrium is unstable, but there is an endemic equilibrium that is asymptotically stable. Again, the quantity \mathcal{R}_0 is the basic reproduction number. It depends on the particular disease (determining the parameter α) and on the rate of contacts, which may depend on the population density in the community being studied. The disease model exhibits a *threshold* behaviour: If the basic reproduction number is less than one the disease will die out, but if the basic reproduction number is greater than one the disease will be endemic. Just as for the epidemic models of Chapter 1, the basic reproduction number is the number of secondary infections caused by a single infective introduced into a wholly susceptible population because the number of contacts per infective in unit time is βK , and the mean infective period (corrected for natural mortality) is $1/(\mu + \alpha)$.

There are two aspects of the analysis of the model (2.2) which are more complicated than the analysis of (2.1). The first is in the study of equilibria. Because of the dependence of $\Lambda(N)$ and $\beta(N)$ on N , it is necessary to use two of the equilibrium conditions to solve for S and I in terms of N and then substitute into the third condition to obtain an equation for N . Then by comparing the two sides of this equation for $N = 0$ and $N = K$ it is possible to show that there must be an endemic equilibrium value of N between 0 and K .

The second complication is in the stability analysis. Since (2.2) is a three dimensional system which can not be reduced to a two dimensional system, the coefficient matrix of its linearization at an equilibrium is a 3×3 matrix and the resulting characteristic equation is a cubic polynomial equation of the form

$$\lambda^3 + a_1\lambda^2 + a_2\lambda + a_3 = 0.$$

The *Routh-Hurwitz* conditions

$$a_1 > 0, \quad a_1a_2 > a_3 > 0$$

are necessary and sufficient conditions for all roots of the characteristic equation to have negative real part. A technically complicated calculation is needed to verify that these conditions are satisfied at an endemic equilibrium for the model (2.2).

The asymptotic stability of the endemic equilibrium means that the compartment sizes approach a steady state. If the equilibrium had been unstable, there would have been a possibility of sustained oscillations. Oscillations in a disease model mean fluctuations in the number of cases to be expected, and if the oscillations have long period could also mean that experimental data for a short period would be quite unreliable as a predictor of the future. Epidemiological models which incorporate additional factors may exhibit oscillations. A variety of such situations is described in [20, 21].

The epidemic models of the first chapter also exhibited a threshold behaviour of a slightly different kind. For these models, which were *SIR* models without births or natural deaths, the threshold distinguished between a dying out of the disease and an epidemic, or short term spread of disease.

From the third equation of (2.1) we obtain

$$N' = \Lambda - \mu N - (1 - f)\alpha I ,$$

where $N = S + I + R$. From this we see that at the endemic equilibrium $N = K - (1 - f)\alpha I/\mu$, and the reduction in the population size from the carrying capacity K is

$$(1 - f)\frac{\alpha}{\mu}I_{\infty} = (1 - f)\left[\frac{\alpha K}{\mu + \alpha} - \frac{\alpha}{\beta}\right] .$$

The parameter α in the *SIR* model may be considered as describing the pathogenicity of the disease. If α is large it is less likely that $\mathcal{R}_0 > 1$. If α is small then the total population size at the endemic equilibrium is close to the carrying capacity K of the population. Thus, the maximum population decrease caused by disease will be for diseases of intermediate pathogenicity.

2.2 The *SIS* Model

In order to describe a model for a disease from which infectives recover with no immunity against reinfection and that includes births and deaths as in the model (2.2), we may modify the model (2.2) by removing the equation for R and moving the term $f\alpha I$ describing the rate of recovery from infection to the equation for S . This gives the model

$$\begin{aligned} S' &= \Lambda(N) - \beta(N)SI - \mu S + f\alpha I \\ I' &= \beta(N)SI - \alpha I - \mu I \end{aligned} \quad (2.6)$$

describing a population with a density-dependent birth rate $\Lambda(N)$ per unit time, a proportional death rate μ in each class, and with a rate α of departure from the infective class through recovery or disease death and with a fraction f of infectives recovering with no immunity against reinfection. In this model, if $f < 1$ the total population size is not constant and K represents a *carrying capacity*, or maximum possible population size, rather than a constant population size.

It is easy to verify that

$$\mathcal{R}_0 = \frac{K\beta(K)}{\mu + \alpha} .$$

If we add the two equations of (2.6), and use $N = S + I$ we obtain

$$N' = \Lambda(N) - \mu N - (1 - f)\alpha I .$$

We will carry out the analysis of the *SIS* model only in the special case $f = 1$, so that N is the constant K . The system (2.6) is asymptotically autonomous and

its asymptotic behaviour is the same as that of the single differential equation

$$I' = \beta(K)I(K - I) - (\alpha + \mu)I, \quad (2.7)$$

where S has been replaced by $K - I$. But (2.7) is a logistic equation which is easily solved analytically by separation of variables or qualitatively by an equilibrium analysis. We find that $I \rightarrow 0$ if $K\beta(K) < (\mu + \alpha)$, or $\mathcal{R}_0 < 1$ and $I \rightarrow I_\infty > 0$ with

$$I_\infty = K - \frac{\mu + \alpha}{\beta(K)} = K\left(1 - \frac{1}{\mathcal{R}_0}\right)$$

if $K\beta(K) > (\mu + \alpha)$ or $\mathcal{R}_0 > 1$.

The endemic equilibrium, which exists if $\mathcal{R}_0 > 1$, is always asymptotically stable. If $\mathcal{R}_0 < 1$ the system has only the disease-free equilibrium and this equilibrium is asymptotically stable. The verification of these properties remains valid if there are no births and deaths. This suggests that a requirement for the existence of an endemic equilibrium is a flow of new susceptibles either through births, as in the *SIR* model or through recovery without immunity against reinfection, as in the *SIS* model with or without births and deaths.

Numerical simulations indicate that the approach to endemic equilibrium for an *SIR* model is like a rapid and severe epidemic if the epidemiological and demographic time scales are very different. The same happens in the *SIS* model. If there are few disease deaths the number of infectives at endemic equilibrium may be substantial, and there may be damped oscillations of large amplitude about the endemic equilibrium. For both the *SIR* and *SIS* models we may write the differential equation for I as

$$I' = I[\beta(N)S - (\mu + \alpha)] = \beta(N)I[S - S_\infty],$$

which implies that whenever S exceeds its endemic equilibrium value S_∞ , I is increasing and epidemic-like behaviour is possible. If $\mathcal{R}_0 < 1$ and $S < K$ it follows that $I' < 0$, and thus I is decreasing. Thus, if $\mathcal{R}_0 < 1$, I cannot increase and no epidemic can occur.

Next, we will turn to some applications of *SIR* and *SIS* models, taken mainly from [4].

2.3 Some Applications

2.3.1 Herd Immunity

In order to prevent a disease from becoming endemic it is necessary to reduce the basic reproduction number \mathcal{R}_0 below one. This may sometimes be achieved by immunization. If a fraction p of the Λ newborn members per unit time of the population is successfully immunized, the effect is to replace K by $K(1 - p)$, and thus to reduce the basic reproduction number to $\mathcal{R}_0(1 - p)$. The requirement $\mathcal{R}_0(1 - p) < 1$ gives $1 - p < 1/\mathcal{R}_0$, or

$$p > 1 - \frac{1}{\mathcal{R}_0}.$$

A population is said to have *herd immunity* if a large enough fraction has been immunized to assure that the disease cannot become endemic. The only disease for which this has actually been achieved worldwide is smallpox for which \mathcal{R}_0 is approximately 5, so that 80 percent immunization does provide herd immunity.

For measles, epidemiological data in the United States indicate that \mathcal{R}_0 for rural populations ranges from 5.4 to 6.3, requiring vaccination of 81.5 percent to 84.1 percent of the population. In urban areas \mathcal{R}_0 ranges from 8.3 to 13.0, requiring vaccination of 88.0 percent to 92.3 percent of the population. In Great Britain, \mathcal{R}_0 ranges from 12.5 to 16.3, requiring vaccination of 92 percent to 94 percent of the population. The measles vaccine is not always effective, and vaccination campaigns are never able to reach everyone. As a result, herd immunity against measles has not been achieved (and probably never can be). Since smallpox is viewed as more serious and requires a lower percentage of the population be immunized, herd immunity was attainable for smallpox. In fact, smallpox has been eliminated; the last known case was in Somalia in 1977, and the virus is maintained now only in laboratories. The eradication of smallpox was actually more difficult than expected because high vaccination rates were achieved in some countries but not everywhere, and the disease persisted in some countries. The eradication of smallpox was possible only after an intensive campaign for worldwide vaccination [19].

2.3.2 Age at Infection.

In order to calculate the basic reproduction number \mathcal{R}_0 for a disease, we need to know the values of the contact rate β and the parameters μ , K , and α . The parameters μ , K , and α can usually be measured experimentally but the contact rate β is difficult to determine directly. There is an indirect means of estimating \mathcal{R}_0 in terms of the life expectancy and the mean age at infection which enables us to avoid having to estimate the contact rate. In this calculation, we will assume that β is constant, but we will also indicate the modifications needed when β is a function of total population size N . The calculation assumes exponentially distributed life spans and infective periods. In fact, the result is valid so long as the life span is exponentially distributed, but if the life span is not exponentially distributed the result could be quite different.

Consider the “age cohort” of members of a population born at some time t_0 and let a be the age of members of this cohort. If $y(a)$ represents the fraction of members of the cohort who survive to age (at least) a , then the assumption that a fraction μ of the population dies per unit time means that $y'(a) = -\mu y(a)$. Since $y(0) = 1$, we may solve this first order initial value problem to obtain $y(a) = e^{-\mu a}$. The fraction dying at (exactly) age a is $-y'(a) = \mu y(a)$. The mean life span is the average age at death, which is $\int_0^\infty a[-y'(a)]da$, and if we integrate by parts we find that this life expectancy is

$$\int_0^\infty [-ay'(a)] da = [-ay(a)]_0^\infty + \int_0^\infty y(a) da = \int_0^\infty y(a) da .$$

Since $y(a) = e^{-\mu a}$, this reduces to $1/\mu$. The life expectancy is often denoted by

L , so that we may write

$$L = \frac{1}{\mu} .$$

The rate at which surviving susceptible members of the population become infected at age a and time $t_0 + a$, is $\beta I(t_0 + a)$. Thus, if $z(a)$ is the fraction of the age cohort alive and still susceptible at age a , $z'(a) = -[\mu + \beta I(t_0 + a)]z(a)$. Solution of this first linear order differential equation gives

$$z(a) = e^{-[\mu a + \int_0^a \beta I(t_0+b) db]} = y(a) e^{-\int_0^a \beta I(t_0+b) db} .$$

The mean length of time in the susceptible class for members who may become infected, as opposed to dying while still susceptible, is

$$\int_0^\infty e^{-\int_0^a \beta I(t_0+b) db} da ,$$

and this is the mean age at which members become infected. If the system is at an equilibrium I_∞ , this integral may be evaluated, and the mean age at infection, denoted by A , is given by

$$A = \int_0^\infty e^{-\beta I_\infty a} da = \frac{1}{\beta I_\infty} .$$

For our model the endemic equilibrium is

$$I_\infty = \frac{\mu K}{\mu + \alpha} - \frac{\mu}{\beta} ,$$

and this implies

$$\frac{L}{A} = \frac{\beta I_\infty}{\mu} = \mathcal{R}_0 - 1 . \quad (2.8)$$

This relation is very useful in estimating basic reproduction numbers. For example, in some urban communities in England and Wales between 1956 and 1969 the average age of contracting measles was 4.8 years. If life expectancy is assumed to be 70 years, this indicates $\mathcal{R}_0 = 15.6$.

If β is a function $\beta(N)$ of total population size the relation (2.8) becomes

$$\mathcal{R}_0 = \frac{\beta(K)}{\beta(N)} \left[1 + \frac{L}{A} \right] .$$

If disease mortality does not have a large effect on total population size, in particular if there is no disease mortality, this relation is very close to (2.8).

The relation between age at infection and basic reproduction number indicates that measures such as inoculations, which reduce \mathcal{R}_0 , will increase the average age at infection. For diseases such as rubella (German measles), whose effects may be much more serious in adults than in children, this indicates a danger that must be taken into account: While inoculation of children will decrease the number of cases of illness, it will tend to increase the danger to those who are not inoculated or for whom the inoculation is not successful. Nevertheless, the number of infections in older people will be reduced, although the fraction of cases which are in older people will increase.

2.3.3 Diseases as Population Control

Many parts of the world experienced very rapid population growth in the 18th century. The population of Europe increased from 118 million in 1700 to 187 million in 1800. In the same time period the population of Great Britain increased from 5.8 million to 9.15 million, and the population of China increased from 150 million to 313 million [29]. The population of English colonies in North America grew much more rapidly than this, aided by substantial immigration from England, but the native population, which had been reduced to one tenth of their previous size by disease following the early encounters with Europeans and European diseases, grew even more rapidly. While some of these population increases may be explained by improvements in agriculture and food production, it appears that an even more important factor was the decrease in the death rate due to diseases. Disease death rates dropped sharply in the 18th century, partly from better understanding of the links between illness and sanitation and partly because the recurring invasions of bubonic plague subsided, perhaps due to reduced susceptibility. One plausible explanation for these population increases is that the bubonic plague invasions served to control the population size, and when this control was removed the population size increased rapidly.

In developing countries it is quite common to have high birth rates and high disease death rates. In fact, when disease death rates are reduced by improvements in health care and sanitation it is common for birth rates to decline as well, as families no longer need to have as many children to ensure that enough children survive to take care of the older generations. Again, it is plausible to assume that population size would grow exponentially in the absence of disease but is controlled by disease mortality.

The *SIR* model with births and deaths of Kermack and McKendrick [24] includes births in the susceptible class proportional to population size and a natural death rate in each class proportional to the size of the class. Let us analyze a model of this type with birth rate r and a natural death rate $\mu < r$. For simplicity we assume the disease is fatal to all infectives with disease death rate α , so that there is no removed class and the total population size is $N = S + I$. Our model is

$$\begin{aligned} S' &= r(S + I) - \beta SI - \mu S \\ I' &= \beta SI - (\mu + \alpha)I. \end{aligned} \tag{2.9}$$

From the second equation we see that equilibria are given by either $I = 0$ or $\beta S = \mu + \alpha$. If $I = 0$ the first equilibrium equation is $rS = \mu S$, which implies $S = 0$ since $r > \mu$. It is easy to see that the equilibrium $(0,0)$ is unstable. What actually would happen if $I = 0$ is that the susceptible population would grow exponentially with exponent $r - \mu > 0$. If $\beta S = \mu + \alpha$ the first equilibrium condition gives

$$r \frac{\mu + \alpha}{\beta} + rI - (\mu + \alpha)I - \frac{\mu(\mu + \alpha)}{\beta} = 0,$$

which leads to

$$(\alpha + \mu - r)I = \frac{(r - \mu)(\mu + \alpha)}{\beta}.$$

Thus, there is an endemic equilibrium provided $r < \alpha + \mu$, and it is possible to show by linearizing about this equilibrium that it is asymptotically stable. On the other hand, if $r > \alpha + \mu$ there is no positive equilibrium value for I . In this case we may add the two differential equations of the model to give

$$N' = (r - \mu)N - \alpha I \geq (r - \mu)N - \alpha N = (r - \mu - \alpha)N$$

and from this we may deduce that N grows exponentially. For this model either we have an asymptotically stable endemic equilibrium or population size grows exponentially. In the case of exponential population growth we may have either vanishing of the infection or an exponentially growing number of infectives.

If only susceptibles contribute to the birth rate, as may be expected if the disease is sufficiently debilitating, the behaviour of the model is quite different. Let us consider the model

$$\begin{aligned} S' &= rS - \beta SI - \mu S = S(r - \mu - \beta I) \\ I' &= \beta SI - (\mu + \alpha)I = I(\beta S - \mu - \alpha) \end{aligned} \quad (2.10)$$

which has the same form as the Lotka-Volterra predator-prey model of population dynamics. This system has two equilibria, obtained by setting the right sides of each of the equations equal to zero, namely $(0,0)$ and an endemic equilibrium $((\mu + \alpha)/\beta, (r - \mu)/\beta)$. It turns out that the qualitative analysis approach we have been using is not helpful as the equilibrium $(0,0)$ is unstable and the eigenvalues of the coefficient matrix at the endemic equilibrium have real part zero. In this case the behaviour of the linearization does not necessarily carry over to the full system. However, we can obtain information about the behaviour of the system by a method that begins with the elementary approach of separation of variables for first order differential equations. We begin by taking the quotient of the two differential equations and using the relation

$$\frac{I'}{S'} = \frac{dI}{dS}$$

to obtain the separable first order differential equation

$$\frac{dI}{dS} = \frac{I(\beta S - \mu - \alpha)}{S(r - \beta I)}.$$

Separation of variables gives

$$\int \left(\frac{r}{I} - \beta \right) dI = \int \left(\beta - \frac{\mu + \alpha}{S} \right) dS.$$

Integration gives the relation

$$\beta(S + I) - r \log I - (\mu + \alpha) \log S = c$$

where c is a constant of integration. This relation shows that the quantity

$$V(S, I) = \beta(S + I) - r \log I - (\mu + \alpha) \log S$$

is constant on each orbit (path of a solution in the (S, I) plane). Each of these orbits is a closed curve corresponding to a periodic solution.

This model is the same as the simple epidemic model of Section 1.2 except for the birth and death terms, and in many examples the time scale of the disease is much faster than the time scale of the demographic process. We may view the model as describing an epidemic initially, leaving a susceptible population small enough that infection cannot establish itself. Then there is a steady population growth until the number of susceptibles is large enough for an epidemic to recur. During this growth stage the infective population is very small and random effects may wipe out the infection, but the immigration of a small number of infectives will eventually restart the process. As a result, we would expect recurrent epidemics. In fact, bubonic plague epidemics did recur in Europe for several hundred years. If we modify the demographic part of the model to assume limited population growth rather than exponential growth in the absence of disease, the effect would be to give behaviour like that of the model studied in the previous section, with an endemic equilibrium that is approached slowly in an oscillatory manner if $\mathcal{R}_0 > 1$.

2.4 Possible extensions

In all the models that we have analyzed, when there is an endemic equilibrium it is unique and is asymptotically stable. There are models for which there is an unstable endemic equilibrium, for example an *SIRS* model with temporary immunity for a fixed length of time following recovery from disease. In such situations there is a transition to an asymptotically stable periodic solution signalled by the passage of the real parts of a pair of eigenvalues of the characteristic equation of the linearization at the equilibrium through zero. Such a change in behaviour is called a *Hopf bifurcation* and the infection still persists but in an oscillatory manner.

We have been assuming homogeneous mixing of members of the population being studied, and this is certainly unrealistically simple. Members of the population may differ, for example, in rate of contact or in location. In the study of sexually transmitted diseases differences in activity levels are important aspects. Contact rates may be age-dependent, and this would suggest the use of age-structured models. Spatial dependence may take two forms, the local diffusion of members of the population, which would lead to partial differential equations of reaction-diffusion types, or travel between communities, which would lead to patch or metapopulation models.

Models incorporating one or more of these kinds of heterogeneity can be developed and analyzed. Inevitably, their analysis involves more structure, equations, and parameters, as well as more sophisticated mathematical methods.

There are other modes of transmission of communicable diseases that can be described by compartmental models. Some infections can be transmitted *vertically*, that is from mother to daughter prior to birth. Another form is transmission by a *vector*. For example, malaria is transmitted back and forth between humans and mosquitoes. Thus an infected mosquito may bite a human and thus infect the human. An uninfected mosquito may bite an infected human and become infected, but infection is not transmitted directly from human to human or from mosquito to mosquito. Sexually transmitted diseases that are transmitted by heterosexual contact are also examples of vector transmission, with male and females playing the role of the two species.

Bibliography

- [1] R.M. Anderson & R.M. May, *Infectious Diseases of Humans*. Oxford Science Publications, Oxford (1991)
- [2] J. Arino, F. Brauer, P. van den Driessche, J. Watmough & J. Wu, A final size relation for epidemic models, *Math. Biosc. & Eng.* **4** (2007), 159–176.
- [3] J. Arino, F. Brauer, P. van den Driessche, J. Watmough & J. Wu, Simple models for containment of a pandemic, *J. Roy. Soc. Interface*, **3** (2006), 453-457.
- [4] F. Brauer & C. Castillo-Chavez, *Mathematical Models in Population Biology and Epidemiology*. Springer, New York (2001)
- [5] F. Brauer, P. van den Driessche & J. Wu, eds., *Mathematical Epidemiology*, *Lecture Notes in Mathematics*, *Mathematical Biosciences subseries 1945*, Springer (2008).
- [6] D.S. Callaway, M.E.J. Newman, S.H. Strogatz, D.J. Watts, Network robustness and fragility: Percolation on random graphs, *Phys. Rev. Letters*, **85**, 5468–5471 (2000).
- [7] C. Castillo-Chavez, K.L. Cooke, W. Huang, S.A. Levin, The role of long incubation periods in the dynamics of HIV/AIDS. Part 1: Single Populations Models. *J. Math. Biol.*, **27**, 373–98 (1989)
- [8] C. Castillo-Chavez, H.R. Thieme, Asymptotically autonomous epidemic models. In: O. Arino, D. Axelrod, M. Kimmel, M. Langlais (eds) *Mathematical Population Dynamics: Analysis of Heterogeneity*, Vol. 1, *Theory of Epidemics*, Wuerz, Winnipeg, pp. 33–50 (1993)
- [9] O. Diekmann & J.A.P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases*. Wiley, Chichester (2000)
- [10] K. Dietz, Overall patterns in the transmission cycle of infectious disease agents. In: R.M. Anderson, R.M. May (eds) *Population Biology of Infectious Diseases*. *Life Sciences Research Report 25*, Springer-Verlag, Berlin-Heidelberg-New York, pp. 87–102 (1982)

- [11] P. Erdős & A. Rényi: On random graphs, *Publicationes Mathematicae* **6**, 290–297 (1959).
- [12] P. Erdős & A. Rényi: On the evolution of random graphs, *Pub. Math. Inst. Hung. Acad. Science* **5**, 17–61 (1960).
- [13] P. Erdős & A. Rényi: On the strengths of connectedness of a random graph, *Acta Math. Scientiae Hung.* **12**, 261–267 (1961).
- [14] Z. Feng, Final and peak epidemic sizes for *SEIR* models with quarantine and isolation, *Math. Biosc. & Eng.* **4** (2007), 675–686.
- [15] Z. Feng, D. Xu & W. Zhao, Epidemiological models with non-exponentially distributed disease stages and applications to disease control, *Bull. Math. Biol.* **69** (2007), 1511–1536,
- [16] A. Gumel, S. Ruan, T. Day, J. Watmough, P. van den Driessche, F. Brauer, D. Gabrielson, C. Bowman, M.E. Alexander, S. Ardal, J. Wu & B.M. Sahai, Modeling strategies for controlling SARS outbreaks based on Toronto, Hong Kong, Singapore and Beijing experience, *Proc. Roy. Soc. London* **271**, 2223–2232 (2004)
- [17] J.M. Heffernan, R.J. Smith? & L.M. Wahl, Perspectives on the basic reproductive ratio, *J. Roy. Soc. Interface* **2** (2005), 281–293.
- [18] H.W. Hethcote, Qualitative analysis for communicable disease models. *Math. Biosc.*, **28**, 335–356 (1976)
- [19] H.W. Hethcote, An immunization model for a heterogeneous population, *Theor. Pop. Biol.*, **14**, 338–349 (1978)
- [20] H.W. Hethcote & S.A. Levin, Periodicity in epidemic models. In : S.A. Levin, T.G. Hallam & L.J. Gross (eds) *Applied Mathematical Ecology. Biomathematics 18*, Springer-Verlag, Berlin-Heidelberg-New York, pp. 193–211 (1989)
- [21] H.W. Hethcote, H.W. Stech & P. van den Driessche, Periodicity and stability in epidemic models: a survey. In: S. Busenberg & K.L. Cooke (eds.) *Differential Equations and Applications in Ecology, Epidemics and Population Problems*, pp. 65–82 (1981)
- [22] J. M. Hyman, J. Li & E. A. Stanley, The differential infectivity and staged progression models for the transmission of HIV, *Math. Biosci.* **155** (1999), 77–109.
- [23] W.O. Kermack & A.G. McKendrick, A contribution to the mathematical theory of epidemics. *Proc. Royal Soc. London.* **115**, 700–721 (1927)
- [24] W.O. Kermack & A.G. McKendrick, Contributions to the mathematical theory of epidemics, part. II. *Proc. Roy. Soc. London*, **138**, 55–83 (1932)

- [25] W.O. Kermack & A.G. McKendrick, Contributions to the mathematical theory of epidemics, part. III. *Proc. Roy. Soc. London*, **141**, 94–112 (1932)
- [26] A. L. Lloyd, Realistic distributions of infectious periods in epidemic models: Changing patterns of persistence and dynamics, *Theor. Pop. Biol.*, **60**(2001), 59–71.
- [27] L. Markus, Asymptotically autonomous differential systems. In: S. Lefschetz (ed) *Contributions to the Theory of Nonlinear Oscillations III*. *Annals of Mathematics Studies* **36**, Princeton University Press, Princeton, N.J., pp. 17–29 (1956)
- [28] J. Mena-Lorca & H.W. Hethcote, Dynamic models of infectious diseases as regulators of population size. *J. Math. Biol.*, **30** (1992), 693–716.
- [29] W.H. McNeill, *Plagues and Peoples*. Doubleday, New York (1976)
- [30] M.E.J. Newman, The spread of epidemic disease on networks, *Phys. Rev. E* **66**, 016128 (2002).
- [31] M.E.J. Newman, The structure and function of complex networks. *SIAM Review*, **45**, 167–256 (2003)
- [32] M.E.J. Newman, S.H. Strogatz & D.J. Watts, Random graphs with arbitrary degree distributions and their applications, *Phys. Rev. E* **64**, 026118 (2001).
- [33] S. Riley, C. Fraser, C.A. Donnelly, A.C. Ghani, L.J. Abu-Raddad, A.J. Hedley, G.M. Leung, L-M Ho, T-H Lam, T.Q. Thach, P. Chau, K-P Chan, S-V Lo, P-Y Leung, T. Tsang, W. Ho, K-H Lee, E.M.C. Lau, N.M. Ferguson, & R.M. Anderson, Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions, *Science* **300**, 1961–1966 (2003).
- [34] S.H. Strogatz, Exploring complex networks, *Nature*, **410**, 268–276 (2001)
- [35] H.R. Thieme, Asymptotically autonomous differential equations in the plane. *Rocky Mountain J. Math.*, **24**, 351–380 (1994)
- [36] H.R. Thieme & C. Castillo-Chavez, On the role of variable infectivity in the dynamics of the human immunodeficiency virus. In: C. Castillo-Chavez, C. (ed) *Mathematical and Statistical Approaches to AIDS Epidemiology*. *Lect. Notes Biomath.* **83**, Springer-Verlag, Berlin-Heidelberg-New York, pp. 200–217 (1989)
- [37] Thieme, H.R., Castillo-Chavez, C.: How may infection-age dependent infectivity affect the dynamics of HIV/AIDS? *SIAM J. Appl. Math.*, **53**, 1447–1479 (1989)

- [38] P. van den Driessche & J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math. Biosc.*, **180** (2002), 29–48.
- [39] H. J. Wearing, P. Rohani & M. J. Keeling, Appropriate models for the management of infectious diseases, *PLOS Medicine*, **2** (2005), 621–627.