describe some of these in this chapter. The modelling literature is now extensive and growing very quickly. Although now quite old, a good introduction to the variety of problems and models for the spread and control of infectious diseases is the book by Bailey (1975). The article by Hethcote (1994) reviews three basic epidemiological models. The book by Diekmann and Heesterbeek (2000) is a good introduction to the field. For example, they discuss how to use biological assumptions in constructing models and present applications; they cover both deterministic and stochastic modelling. Other sources are to be found in the above references and in the papers referred to in the rest of this chapter. Particularly useful sources for the latest information on specific diseases, either globally or for a specific country, are the WHO (http://www.who.org/) and the CDC (http://www.cdc.gov/); their search and information features are very efficient.

In this chapter we discuss several quite different models for very different diseases which incorporate some general aspects of epidemiological modelling of disease transmission, time evolution of epidemics, acquired resistance to infection, vaccination strategies and so on. The use of mathematical modelling in immunology and virology is also growing very quickly. We discuss in some detail models for the dynamics of HIV infections and relate them to patient data. We also discuss a bacterial infection and one involving parasites. In Chapter 13, Volume II we consider the geographic spread of infectious diseases and describe in detail a practical model for the spatial spread of rabies, a possible means of its control and the effect of including immunity. The modelling of infectious diseases involves the concepts of population dynamics which we have dicussed in earlier chapters. Although the detailed forms of the equations are different the essential elements and analysis are very similar.

At the basic level we consider two types of models. In one the total population is taken to be approximately constant with, for example, the population divided into susceptible, infected and immune groups: other groupings are also possible, depending on the disease. We first discuss models in this category. In the other, the population size is affected by the disease via the birth rate, mortality and so on. Host–parasite interacting populations often come into this category. We only discuss deterministic models which are deficient in certain situations—eradication of a disease is one, since here the probability that the last few infected individuals will infect another susceptible is not deterministic. Nevertheless it is perhaps surprising how useful, and quantitatively predictive, deterministic models can often be; the examples below are only a very few examples where this has proven to be the case.

## 10.2 Simple Epidemic Models and Practical Applications

In the classical (but still highly relevant) models we consider here the total population is taken to be constant. If a small group of infected individuals is introduced into a large population, a basic problem is to describe the spread of the infection within the population as a function of time. Of course this depends on a variety of circumstances, including the actual disease involved, but as a first attempt at modelling directly transmitted diseases we make some not unreasonable general assumptions.

Consider a disease which, after recovery, confers immunity which, if lethal, includes deaths: dead individuals are still counted. Suppose the disease is such that the

population can be divided into three distinct classes: the susceptibles, $S$, who can catch the disease; the infectives, $I$, who have the disease and can transmit it; and the removed class, $R$, namely, those who have either had the disease, or are recovered, immune or isolated until recovered. The progress of individuals is schematically represented by

$$S \longrightarrow I \longrightarrow R.$$

Such models are often called *SIR* models. The number of classes depends on the disease. *SI* models, for example, have only susceptible and infected classes while *SEIR* models have a suceptible class, $S$, a class in which the disease is latent, $E$, an infectious class, $I$, and a recovered or dead class, $R$.

The assumptions made about the transmission of the infection and incubation period are crucial in any model; these are reflected in the terms in the equations and the parameters. With $S(t)$, $I(t)$ and $R(t)$ as the number of individuals in each class we assume here that: (i) The gain in the infective class is at a rate proportional to the number of infectives and susceptibles, that is, $rSI$, where $r > 0$ is a constant parameter. The susceptibles are lost at the same rate. (ii) The rate of removal of infectives to the removed class is proportional to the number of infectives, that is, $aI$ where $a > 0$ is a constant; $1/a$ is a measure of the time spent in the infectious state. (iii) The incubation period is short enough to be negligible; that is, a susceptible who contracts the disease is infective right away.

We now consider the various classes as uniformly mixed; that is, every pair of individuals has equal probability of coming into contact with one another. This is a major assumption and in many situations does not hold as in most sexually transmitted diseases (STD's). The model mechanism based on the above assumptions is then

$$\frac{dS}{dt} = -rSI, \tag{10.1}$$

$$\frac{dI}{dt} = rSI - aI, \tag{10.2}$$

$$\frac{dR}{dt} = aI, \tag{10.3}$$

where $r > 0$ is the infection rate and $a > 0$ the removal rate of infectives. This is the classic Kermack–McKendrick (1927) model. We are, of course, only interested in non-negative solutions for $S$, $I$ and $R$. This is a basic model but, even so, we can make some highly relevant general comments about epidemics and, in fact, adequately describe some specific epidemics with such a model.

The constant population size is built into the system (10.1)–(10.3) since, on adding the equations,

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0 \quad \Rightarrow \quad S(t) + I(t) + R(t) = N, \tag{10.4}$$

where $N$ is the total size of the population. Thus, $S$, $I$ and $R$ are all bounded above by $N$. The mathematical formulation of the epidemic problem is completed given initial

conditions such as

$$S(0) = S_0 > 0, \quad I(0) = I_0 > 0, \quad R(0) = 0. \tag{10.5}$$

A key question in any epidemic situation is, given $r$, $a$, $S_0$ and the initial number of infectives $I_0$, whether the infection will spread or not, and if it does how it develops with time, and crucially when it will start to decline. From (10.2),

$$\left[\frac{dI}{dt}\right]_{t=0} = I_0(rS_0 - a) \quad \begin{cases} > 0 \\ < 0 \end{cases} \quad \text{if} \quad S_0 \quad \begin{cases} > \rho \\ < \rho \end{cases}, \quad \rho = \frac{a}{r}. \tag{10.6}$$

Since, from (10.1), $dS/dt \leq 0$, $S \leq S_0$ we have, if $S_0 < a/r$,

$$\frac{dI}{dt} = I(rS - a) \leq 0 \quad \text{for all} \quad t \geq 0, \tag{10.7}$$

in which case $I_0 > I(t) \to 0$ as $t \to \infty$ and so the infection dies out; that is, no epidemic can occur. On the other hand if $S_0 > a/r$ then $I(t)$ initially increases and we have an epidemic. The term 'epidemic' means that $I(t) > I_0$ for some $t > 0$; see Figure 10.1. We thus have a *threshold phenomenon*. If $S_0 > S_c = a/r$ there is an epidemic while if $S_0 < S_c$ there is not. The critical parameter $\rho = a/r$ is sometimes called the *relative removal rate* and its reciprocal $\sigma(= r/a)$ the infection's *contact rate*.

We write

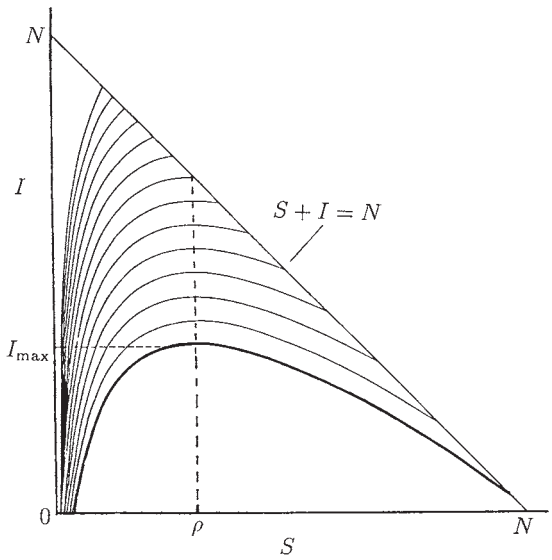


**Figure 10.1.** Phase trajectories in the susceptibles ($S$)-infectives ($I$) phase plane for the *SIR* model epidemic system (10.1)–(10.3). The curves are determined by the initial conditions $I(0) = I_0$ and $S(0) = S_0$. With $R(0) = 0$, all trajectories start on the line $S + I = N$ and remain within the triangle since $0 < S + I < N$ for all time. An epidemic situation formally exists if $I(t) > I_0$ for any time $t > 0$; this always occurs if $S_0 > \rho(= a/r)$ and $I_0 > 0$.

$$R_0 = \frac{rS_0}{a},$$

where $R_0$ is the basic *reproduction rate* of the infection, that is, the number of secondary infections produced by one primary infection in a wholly susceptible population. Here $1/a$ is the average infectious period. If more than one secondary infection is produced from one primary infection, that is, $R_0 > 1$, clearly an epidemic ensues. The whole question of thresholds in epidemics is obviously important. The definition and derivation or computation of the basic reproduction rate is crucial and can be quite complicated. One such example is if the population is heterogeneous (Diekman et al. 1990).

The basic reproduction rate is a crucial parameter grouping for dealing with an epidemic or simply a disease which is currently under control with vaccination, for example. Although the following arguments are based on $R_0$ they are quite general. Clearly one way to reduce the reproduction rate is to reduce the number of sucepti-bles, $S_0$. Vaccination is the common method of doing this and it has been successful in eradicating smallpox. In the U.S.A. it reduced the incidence of measles from 894,134 reported cases in 1941 to 135 in 1997 and for polio from 21,269 in 1952 to the last indigenously acquired case of wild-virus polio reported in 1979 (the Western hempisphere was officially certified polio-free in 1994) with similar reductions in other childhood diseases. Mass vaccination is the cheapest and most effective means of disease control. However, although vaccines are generally extremely safe, no medicine is totally risk-free, however small the risk may be. (There have, however, been a few cases of instant death from diptheria and tetanus vaccines and there is currently much controversy about the vaccine for Anthrax for the military.) As people in the West forget the ravages of polio, measles, diptheria, rubella and so on, many will become less keen to have their children vaccinated because of the risk even if very small. Vaccination not only provides protection for the individual it also provides it for the community at large since it keeps the effective reproduction rate below the level which would allow an epidemic to start. This is the so-called 'herd immunity.' The point is that once the threshold herd immunity level of $R_0$ has been reached and memory of former diseases fades there is the possibility that people will not have their children vaccinated but have a free ride instead; the unvaccinated have effectively the same immunity. In this situation the best, but unethical, strategy for parents is to urge all other parents to have their children vaccinated but free ride with their own. The important point to keep in mind, however, is that an epidemic can start and rise very quickly if the reproduction rate increases beyond the critical value for an epidemic so in the end free-riding is not without its own risks. (This happened with the Conquistadors in Mexico.)

We can derive some other useful analytical results from this simple model. From (10.1) and (10.2)

$$\frac{dI}{dS} = -\frac{(rS - a)I}{rSI} = -1 + \frac{\rho}{S}, \quad \rho = \frac{a}{r}, \quad (I \neq 0).$$

The singularities all lie on the $I = 0$ axis. Integrating the last equation gives the $(I, S)$ phase plane trajectories as

$$I + S - \rho \ln S = \text{constant} = I_0 + S_0 - \rho \ln S_0, \tag{10.8}$$

where we have used the initial conditions (10.5). The phase trajectories are sketched in Figure 10.1. Note that with (10.5), all initial values $S_0$ and $I_0$ satisfy $I_0 + S_0 = N$ since $R(0) = 0$ and so for $t > 0$, $0 \leq S + I < N$.

If an epidemic exists we would like to know how severe it will be. From (10.7) the maximum $I$, $I_{\max}$ , occurs at $S = \rho$ where $dI/dt = 0$. From (10.8), with $S = \rho$,

$$\begin{aligned} I_{\max} &= \rho \ln \rho - \rho + I_0 + S_0 - \rho \ln S_0 \\ &= I_0 + (S_0 - \rho) + \rho \ln \left( \frac{\rho}{S_0} \right) \\ &= N - \rho + \rho \ln \left( \frac{\rho}{S_0} \right). \end{aligned} \tag{10.9}$$

For any initial values $I_0$ and $S_0 > \rho$, the phase trajectory starts with $S > \rho$ and we see that $I$ increases from $I_0$ and hence an epidemic ensues. It may not necessarily be a severe epidemic as is the case if $I_0$ is close to $I_{\max}$. It is also clear from Figure 10.1 that if $S_0 < \rho$ then $I$ decreases from $I_0$ and no epidemic occurs.

Since the axis $I = 0$ is a line of singularities, on all trajectories $I \to 0$ as $t \to \infty$. From (10.1), $S$ decreases since $dS/dt < 0$ for $S \neq 0$, $I \neq 0$. From (10.1) and (10.3),

$$\begin{aligned} &\frac{dS}{dR} = -\frac{S}{\rho} \\ &\Rightarrow \quad S = S_0 e^{-R/\rho} \geq S_0 e^{-N/\rho} > 0 \\ &\Rightarrow \quad 0 < S(\infty) \leq N. \end{aligned} \tag{10.10}$$

In fact from Figure 10.1, $0 < S(\infty) < \rho$. Since $I(\infty) = 0$, (10.4) implies that $R(\infty) = N - S(\infty)$. Thus, from (10.10),

$$S(\infty) = S_0 \exp\left[ -\frac{R(\infty)}{\rho} \right] = S_0 \exp\left[ -\frac{N - S(\infty)}{\rho} \right]$$

and so $S(\infty)$ is the positive root $0 < z < \rho$ of the transcendental equation

$$S_0 \exp\left[ -\frac{N - z}{\rho} \right] = z. \tag{10.11}$$

We then get the total number of susceptibles who catch the disease in the course of the epidemic as

$$I_{\text{total}} = I_0 + S_0 - S(\infty), \tag{10.12}$$

where $S(\infty)$ is the positive solution $z$ of (10.11). An important implication of this analysis, namely, that $I(t) \to 0$ and $S(t) \to S(\infty) > 0$, is that the disease dies out from a lack of infectives and *not* from a lack of susceptibles.

The threshold result for an epidemic is directly related to the relative removal rate, $\rho$: if $S_0 > \rho$ an epidemic ensues whereas it does not if $S_0 < \rho$. For a given disease, the relative removal rate varies with the community and hence determines whether an epidemic may occur in one community and not in another. The number of susceptibles

$S_0$ also plays a major role, of course. For example, if the density of susceptibles is high and the removal rate, $a$, of infectives is low (through ignorance, lack of medical care, inadequate isolation and so on) then an epidemic is likely to occur. Expression (10.9) gives the maximum number of infectives while (10.12) gives the total number who get the infection in terms of $\rho(=a/r)$, $I_0$, $S_0$ and $N$.

In most epidemics it is difficult to determine how many new infectives there are each day since only those that are removed, for medical aid or whatever, can be counted. Public Health records generally give the number of infectives per day, week or month. So, to apply the model to actual epidemic situations, in general we need to know the number removed per unit time, namely, $dR/dt$, as a function of time.

From (10.10), (10.4) and (10.3) we get an equation for $R$ alone; namely,

$$\frac{dR}{dt} = aI = a(N - R - S) = a\left(N - R - S_0 e^{-R/\rho}\right), \quad R(0) = 0, \tag{10.13}$$

which can only be solved analytically in a parametric way: the solution in this form however is not particularly convenient. Of course, if we know $a$, $r$, $S_0$ and $N$ it is a simple matter to compute the solution numerically. Usually we do not know all the parameters and so we have to carry out a best fit procedure assuming, of course, the epidemic is reasonably described by such a model. In practice, however, it is often the case that if the epidemic is not large, $R/\rho$ is small—at least $R/\rho < 1$. Following Kermack and McKendrick (1927) we can then approximate (10.13) by

$$\frac{dR}{dt} = a\left[N - S_0 + \left(\frac{S_0}{\rho} - 1\right)R - \frac{S_0 R^2}{2\rho^2}\right].$$

Factoring the right-hand side quadratic in $R$, we can integrate this equation to get, after some elementary but tedious algebra, the solution

$$\begin{aligned} R(t) &= \frac{r^2}{S_0}\left[\left(\frac{S_0}{\rho} - 1\right) + \alpha \tanh\left(\frac{\alpha a t}{2} - \phi\right)\right] \\ \alpha &= \left[\left(\frac{S_0}{\rho} - 1\right)^2 + \frac{2S_0(N - S_0)}{\rho^2}\right]^{1/2}, \quad \phi = \frac{\tanh^{-1}\left(\frac{S_0}{\rho} - 1\right)}{\alpha}. \end{aligned} \tag{10.14}$$

The removal rate is then given by

$$\frac{dR}{dt} = \frac{a\alpha^2\rho^2}{2S_0} \operatorname{sech}^2\left(\frac{\alpha a t}{2} - \phi\right), \tag{10.15}$$

which involves only 3 parameters, namely, $a\alpha^2\rho^2/(2S_0)$, $\alpha a$ and $\phi$. With epidemics which are not large, it is this function of time which we should fit to the public health records. On the other hand, if the disease is such that we know the actual number of the removed class then it is $R(t)$ in (10.14) we should use. If $R/\rho$ is not small, however, we must use the differential equation (10.13) to determine $R(t)$.

We now apply the model to two very different epidemic situations.

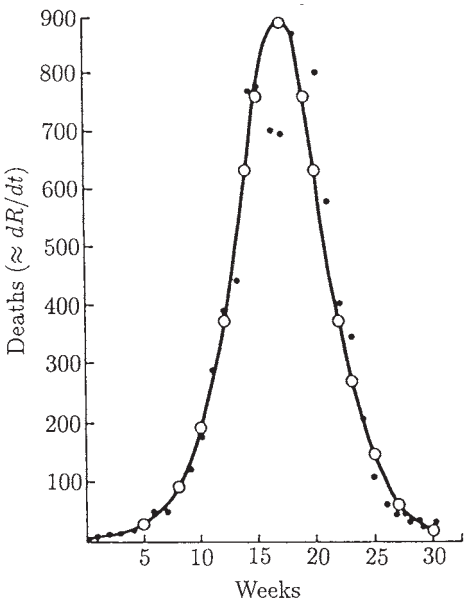


**Figure 10.2.** Bombay plague epidemic of 1905–1906. Comparison between the data (•) and theory (○) from the (small) epidemic model and where the number of deaths is approximately $dR/dt$ given by (10.16). (After Kermack and McKendrick 1927)

*Bombay Plague Epidemic 1905–1906*

This plague epidemic lasted for almost a year. Since most of the victims who got the disease died, the number removed per week, that is, $dR/dt$, was approximately equal to the number of deaths per week. On the basis that the epidemic was not severe (relative to the population size), Kermack and McKendrick (1927) compared the actual data with (10.15) and determined the best fit for the three parameters which resulted in

$$\frac{dR}{dt} = 890 \operatorname{sech}^2 (0.2t - 3.4). \tag{10.16}$$

This is illustrated in Figure 10.2 together with the actual epidemic data.

*Influenza Epidemic in an English Boarding School 1978*

In 1978 in the British medical journal, *The Lancet*, there was a report with detailed statistics of a flu epidemic in a boys' boarding school with a total of 763 boys. Of these 512 were confined to bed during the epidemic, which lasted from 22nd January to 4th February 1978. It seems that one infected boy initiated the epidemic. This situation has many of the requirements assumed in the above model derivation. Here, however, the epidemic was severe and the full system has to be used. Also, when a boy was infected he was put to bed and so we have $I(t)$ directly from the data. Since in this case we have no analytical solution for comparison with the data, a best fit numerical technique was used directly on the equations (10.1)–(10.3) for comparison of the data. Figure 10.3 illustrates the resulting time evolution for the infectives, $I(t)$, together with the epidemic statistics. The $R$-equation (10.3) is uncoupled; the solution for $R(t)$ is simply proportional to the area under the $I(t)$ curve.

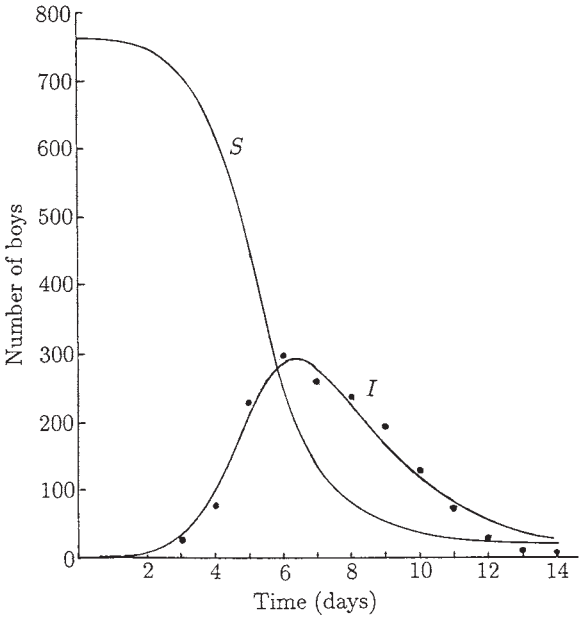


**Figure 10.3.** Influenza epidemic data (•) for a boys' boarding school as reported in the British medical journal, *The Lancet*, 4th March 1978. The continuous curves for the infectives ($I$) and susceptibles ($S$) were obtained from a best fit numerical solution of the *SIR* system (10.1)–(10.3): parameter values $N = 763$, $S_0 = 762$, $I = 1$, $\rho = 202$, $r = 2.18 \times 10^{-3}$/day. The conditions for an epidemic to occur, namely, $S_0 > \rho$, are clearly satisfied and the epidemic is severe since $R/\rho$ is not small.

*Plague in Eyam, England 1665–1666*

There was an outbreak of plague in the village of Eyam in England from 1665 to 1666. In this remarkable altruistic incident, the village sealed itself off when plague was discovered, so as to prevent it spreading to the neighbouring villages, and it was successful. By the end of the epidemic only 83 of the original population of 350 survived. Raggett (1982) applied the *SIR* model (10.1)–(10.3) to this outbreak. Here, $S(\infty) = 83$ out of an initial $S_0 = 350$. This is another example, like the school flu epidemic, where the epidemic was severe. Raggett (1982) shows how to determine the parameters from the available data and knowledge of the etiology of the disease. He reiterates the view that although the initial form was probably bubonic plague, the pneumonic form most likely became prevalent; the latter form can be transmitted from the cough of a victim (see Chapter 13, Volume II for a brief description of the plague and its history). The comparison between the solutions from the deterministic model and the Eyam data is very good. The comparison is much better than that obtained from the corresponding stochastic model, which Raggett (1982) also considered. We discuss a model for the spatial spread of plague in Chapter 13, Volume II.

If a disease is *not* of short duration then (10.1), the equation for the susceptibles, should include birth and death terms. Mortality due to natural causes should also be included in equation (10.2) for the infectives and in (10.3) for the removed class. The

resulting models can be analysed in a similar way to that used here and in Chapter 3 on interacting populations: they are still systems of ordinary differential equations. It is not surprising, therefore, that oscillatory behaviour in disease epidemics is common; these are often referred to as epidemic waves. Here they are *temporal* waves. *Spatial* epidemic waves appear as an epidemic spreads geographically. The latter are also common and we consider them in detail in Chapter 13, Volume II.

Many diseases have a latent or incubation period when a susceptible has become infected but is not yet infectious. Measles, for example, has an 8- to 13-day latent period. The incubation time for AIDS, on the other hand, is anything from a few months to years after the patient has been shown to have antibodies to the human immunodeficiency virus (HIV). We can, for example, incorporate this as a delay effect, or by introducing a new class, $E(t)$ say, in which the susceptible remains for a given length of time before moving into the infective class. Such models give rise to integral equation formulations and they can exhibit oscillatory behaviour as might be expected from the inclusion of delays. Some of these are described by Hoppensteadt (1975, see also 1982). Nonlinear oscillations in such models have been studied by Hethcote et al. (1981); see also Hethcote (1994). Alternative approaches recently used in modelling AIDS are discussed below. Finally age, $a$, is often a crucial factor in disease susceptibility and infectiousness. The models then become partial differential equations with independent variables $(t, a)$; we consider one such model later in this chapter.

There are many modifications and extensions which can and often must be incorporated in epidemic models; these depend critically on the disease and location. In the following sections we discuss a few more general models to illustrate different but important points. The books and references already cited describe numerous models and go into them in considerable detail.

## 10.3 Modelling Venereal Diseases

The incidence of sexually transmitted diseases (STDs), such as gonorrhea (*Neisseria gonorrhoeae*), chlamydia, syphilis and, of course, AIDS, is a major health problem in both developed and developing countries. In the U.S.A., for example, as reported by the Centers for Disease Control (www.cdc.gov), in 1996 there were over 300,000 cases of gonorrhea reported and over 11,000 cases of syphilis and nearly 500,000 cases of chlamydia. Whereas the rate has been decreasing for gonorrhea and syphilis it is growing for chlamydia. We give some of the numbers for HIV incidence in the AIDS sections below.

STDs have certain characteristics which are different from other infections, such as measles or rubella (German measles). One difference is that they are mainly restricted to the sexually active community so the assumption of uniform mixing in the whole population is not really justified. Another is that often the carrier is asymptomatic (that is, the carrier shows no overt symptoms) until quite late on in the development of the infection. A third crucial difference is that STDs induce little or no acquired immunity following an infection. Equally important in virus infections is the lack of present knowledge of some of the parameters which characterise the transmission dynamics.