Chapter Three

Host Heterogeneities

The standard models introduced in Chapter 2 compartmentalize the population only in terms of infection status and history—classifying individuals as susceptible, infected, or recovered—and modeling the number of individuals in each compartment. As such, there is only one degree of subdivision within the population. In this chapter, we introduce a second degree, further dividing the population into classes with similar behavioral characteristics. These characteristics should be chosen such that all members of a class have comparable risk of both contracting and transmitting infection.

Two clear motivating examples dominate the literature of models dealing with risk: (1) age structure for childhood infections, and (2) risk structure for sexually transmitted infections (STIs). Models for STIs frequently subdivide the population into classes dependent upon the risk associated with the behavior in each class. High-risk individuals, for example, have many sexual partners (or for some STIs could partake in intravenous drug use, frequently sharing needles). As a result, individuals in this class have a higher risk of both contracting and transmitting infection. Given the very clear link between the number of partners and the risk of infection, and the heterogeneity between individuals (Johnson et al. 1994), it seems intuitive to include such variation in models if we wish to understand and predict the patterns of sexually transmitted infections. Although ignoring the behavioral heterogeneity and assuming that everyone has the same number of partners is appealing for its simplicity, we will show that such averaging can produce very misleading results.

For many infections, such as measles, mumps, or chickenpox, a high basic reproductive ratio (R_0) means that the average age at first infection is low (see Chapter 2), and therefore they are most commonly encountered during childhood. The modeling of these so-called "childhood diseases" also requires further partitioning of the population, this time in terms of age. For STIs, the subdivisions are determined by the *number* of contacts. For childhood diseases, in contrast, the subdivisions are due to the *nature* of contacts. Because such diseases are common in childhood but rare in adults, those individuals who mix most with children are at the greatest risk. Due to their aggregation in schools, children predominantly mix with other children, and therefore age acts as the major determinant of risk.

Both of these examples show how differential patterns of mixing influence the likelihood of contracting and transmitting an infection. Some subsets of the population are clearly at greater risk, whereas some are relatively isolated. By incorporating such heterogeneities into models, we gain three distinct advantages: The aggregate behavior of models becomes more accurate, we can determine the prevalence of infection within the different classes, and finally we can use this information to determine more efficient targeted measures for disease control.

Including risk heterogeneities inevitably increases the number of equations. Incorporating high-risk and low-risk classes with STI models doubles the number of equations

compared to homogeneous (averaged) models. With age-structured models, which can have tens of age classes, the increase in the number of equations is even more pronounced. However, the equations for each class have a very similar structure, so an increase in their number does not correspond to an increase in the intellectual challenge or computational difficulty, merely an increase in computational time. However, including heterogeneities does increase the number of parameters it is necessary to estimate—this often translates into a need for more biological data. Generally, it is preferable to have case reports (or other information) subdivided into the same classes as used in the model because this simplifies parameterization and the comparison of model results with data.

Finally, in this chapter we consider levels of heterogeneity within the infected class, discriminating between individuals due to the time since infection. Frequently, very accurate data is available to parameterize these forms of models, where the distribution of latent and infectious periods is well known. Although these changes in the shape of the distribution may at first appear trivial, they can often have a profound impact on the infection dynamics.

Ultimately, three applied questions drive the work in this chapter:

1. How does risk structure influence the spread and prevalence of pathogens?
2. Can risk structure be used to more effectively control the spread of infection?
3. How can risk structure be parameterized in realistic scenarios?

To answer these questions, we need to extend the basic models of Chapter 2 to include the various risk groups.

3.1. RISK-STRUCTURE: SEXUALLY TRANSMITTED INFECTIONS

In this section, we introduce the concepts of modeling population heterogeneity with the particular example of sexually transmitted infections and just two groups (high risk and low risk). In practice, there are a vast range of sexual risk groups, from prostitutes to celibate individuals, and each ideally requires its own particular class within the model. However, the two-class model used as the primary example in this chapter is sufficiently complex to demonstrate the necessary tools and techniques, yet simple enough to be intuitively understood.

Sexually transmitted infections are a growing problem in many areas of the world. The AIDS epidemic that began in the 1980s is continuing to increase, and in recent years there has been a rise in the prevalence and incidence of many other less well-known STIs. It is important to understand how our ever-changing sexual practices dictate the prevalence of such infections, and how their increase can be effectively combated. Since HIV and AIDS were first diagnosed in 1983 (Barre-Sinoussi et al. 1983), it has spread worldwide with alarming speed (Mertens and LowBeer 1996). In 2002, this pandemic was estimated to infect around 42 million people, increasing at a rate of 5 million new cases a year, and be responsible for around 3 million deaths per year—the vast majority of which are in Sub-Saharan Africa and the heterosexual community (UNAIDS/WHO 2002). In the United States, over 15 million new cases of sexually transmitted infections are diagnosed every year (Cates et al. 1999). In the United Kingdom, there has been a clear increase in many STIs over the last decade (Figure 3.1), with the greatest growth in syphilis, gonorrhoea,

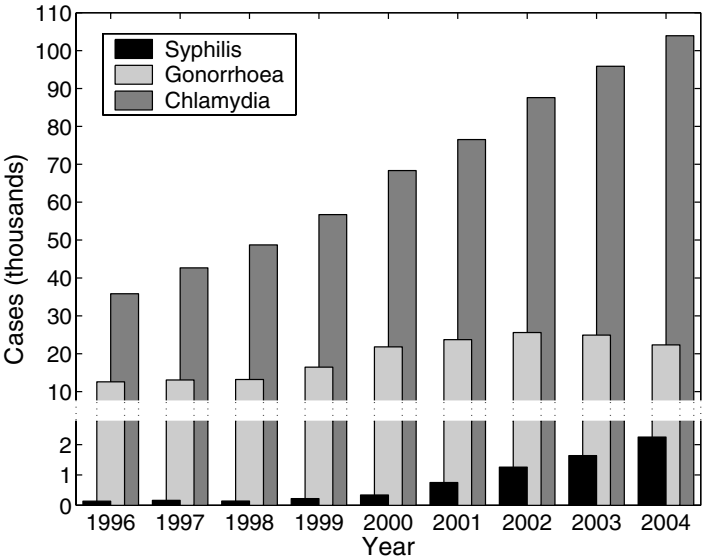


Figure 3.1. The recent trends in three sexually transmitted infections in England, highlighting the dramatic rise in STIs over a nine-year period. The y-axis has been split to show the lower but increasing levels of syphilis. (Data from the Public Health Laboratory Services www.phls.co.uk/infections/topics_az/hiv_and_sti/epidemiology/dataresource.htm).

and chlamydia. This has prompted a great deal of research activity to be focused on modeling these sexually transmitted infections and asserting effective means of control. (syphilis: Cates et al. 1996; Oxman et al. 1996; Chesson et al. 1999; Morris 2001; Pourbohloul et al. 2003. Gonorrhoea: Kretzschmar et al. 1996; Garnett et al. 1999. Chlamydia: Delamaza and Delamaza 1995; Welte et al. 2000; Kretzschmar et al. 2001; Kretzschmar 2002. General: Boily and Masse 1997; Stigum et al. 1997; Garnett and Bowden 2000.)

The majority of sexually transmitted infections conform to the SIS framework (Chapter 2), which leads to one of the simplest of all disease models. For the purposes of epidemiological modeling, the natural history of sexually transmitted infections is relatively simple: Individuals are born susceptible to infection and remain so until they enter sexually active relationships. Transmission of infection, from an infected to a susceptible individual, occurs during sex—the potential transmission routes are, therefore, more clearly defined and determinable than for airborne infections. Recovery from infection generally occurs only following medical treatment, after which the individual becomes susceptible again. HIV is the notable exception to this rule because individuals never recover from the infection. Although the population-level dynamics of STIs with SIS behavior are relatively simple, the within-host dynamics are far more complex. In many cases, the immune response is suppressed so that the infection persists without clinical symptoms for many months or years, thus increasing the number of partners to which it can be spread. As laid out in Chapter 2, for an unstructured (sexually active) population, where everyone is assumed to be at equal risk, the SIS framework of progression leads to

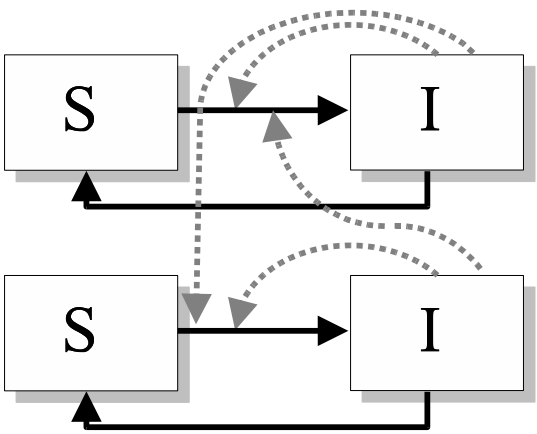
the following equations:

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI + \gamma I \\ \frac{dI}{dt} &= \beta SI - \gamma I, \end{aligned} \tag{3.1}$$

where the effects of individuals entering and leaving the sexually active population have been ignored. As before, the parameter β determines the contact and transmission rates between susceptible and infectious individuals, and γ is the rate at which treatment is sought. If the population size is constant, the rates of change in the fractions susceptible and infectious are always reflections of each other ($\frac{dS}{dt} = -\frac{dI}{dt}$) and only one needs to be calculated. We now consider how this framework can be extended to include multiple interacting groups.

3.1.1. Modeling Risk Structure

From biological and mechanistic principles, we can derive sets of equations for the various risk groups within the population, and from these equations develop a robust, generic framework to explain the interaction between risk and epidemiological dynamics. We start with a simple two-class model, incorporating high-risk and low-risk individuals.



3.1.1.1. High-Risk and Low-Risk Groups

We focus initially on the behavior of the high-risk group, and denote the number of susceptible and infectious individuals within this group by X_H and Y_H, and the total number in the high-risk group by $N_H(= X_H + Y_H)$. Alternatively, it is often simpler to use a frequentist approach, such that S_H and I_H refer to the proportion of the entire population that are susceptible or infectious and also in the high-risk group, in which case n_H is the proportion of the population in the high-risk group: $S_H = X_H/N$, $I_H = Y_H/N$, $n_H = N_H/N$. This is the approach that will be adopted throughout this chapter. With this formulation, a disease-free population has $S_H = n_H < 1$, which is crucial when calculating R_0 or invasion criteria. (A third formulation is to model the proportion of each group that

are infected or susceptible, such that $\widehat{S}_H = X_H/N_H$, $\widehat{I}_H = Y_H/N_H$, $\widehat{S}_H + \widehat{I}_H = 1$; however, this approach does not correspond intuitively with the earlier unstructured models.)

The dynamics of either group is derived from two basic events, infection and recovery. In this simple formulation we explicitly do not allow the movement of individuals between risk groups; individuals are "born" into a risk group and remain so for life. We initially focus on the dynamics of the high-risk group. Recovery, or the loss of infectious cases, can occur only through treatment and, following the unstructured formulation, we assume this occurs at a constant rate γ. For more generality, we could let the treatment rate be specific to the group (i.e., use γ_H), but at this stage such complexity is unnecessary. New infectious cases within the high-risk group occur when a high-risk susceptible is infected by someone in either the high- or low-risk group. These two distinct transmission types require different transmission parameters: We let β_{HH} denote transmission to high risk from high-risk and β_{HL} represent transmission to high risk from low risk. (Note throughout this book we use the same ordering of subscripts such that transmission is always $\beta_{\text{to from}}$.) Putting these elements together, we arrive at the following differential equation:

$$\frac{dI_H}{dt} = \beta_{HH} S_H I_H + \beta_{HL} S_H I_L - \gamma I_H. \tag{3.2}$$

By a similar argument, we can derive an expression for the low-risk individuals:

$$\frac{dI_L}{dt} = \beta_{LH} S_L I_H + \beta_{LL} S_L I_L - \gamma I_L. \tag{3.3}$$

This is online program 3.1

As demonstrated in Chapter 2, the susceptible equations can be ignored because their dynamics are determined by the number of infectious individuals: $S_H = n_H - I_H$ and similarly for the low-risk group (assuming proportions in each group do not change).

There are now four transmission parameters, and the simplest way to encapsulate this information is in a matrix $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{HH} & \beta_{HL} \\ \beta_{LH} & \beta_{LL} \end{pmatrix}.$$

These transmission matrices are often termed WAIFW (Who Acquires Infection From Whom) matrices, and provide a convenient means of capturing the mixing between different social groups. This matrix $\boldsymbol{\beta}$ plays a similar role to the scalar parameter β of the unstructured model. We now seek to make that relationship more transparent. The first consideration is the relative magnitudes of the four terms. The individuals in the high-risk group should be at a higher risk of infection, therefore $\beta_{HH} + \beta_{HL}$ should be larger than $\beta_{LH} + \beta_{LL}$. We would also expect *assortative* mixing, where individuals from the high-risk group are more likely to partner with other high-risk individuals, and low-risk individuals are more likely to be in long-term relationships with another member of the low-risk group (see Section 3.1.6 and Box 3.4). This means that the diagonal elements of the matrix dominate, with β_{HH} being the largest term. Finally, we insist that interactions between the groups are symmetric such that the number of interactions between high- and low-risk groups is the same as interactions between low and high; this implies that $\beta_{HL} = \beta_{LH}$, or more generally that the matrix is symmetric. This assumption equates with the fact that individuals in both groups have an equal response to infectious challenge—if one group was, for some reason, more susceptible, then the symmetry property may break down (see Section 3.1.2.2 for an example of such asymmetry).

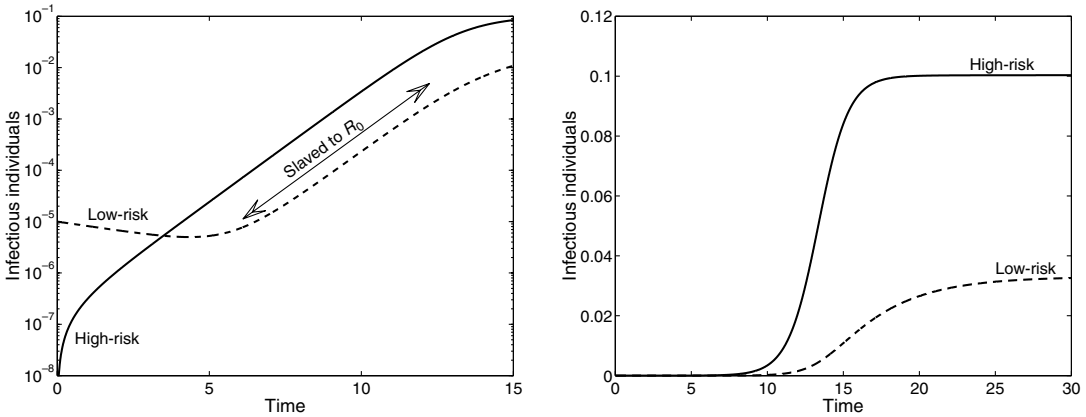


Figure 3.2. The dynamics of an STI introduced into the low-risk class at a small level, showing the prevalence of infection in the high- and low-risk groups relative to the entire population size (I_H and I_L, respectively). Three distinct phases of dynamics can be observed; a transient phase that is determined by the initial distribution of infection, a phase dominated by the value of R_0 when the prevalence of infection increases (or decreases) exponentially, and finally an asymptotic phase when density-dependent factors begin to operate and the equilibrium levels are achieved.

The single transmission parameter in the unstructured models is replaced by a matrix of values for structured models. In general, when all groups have an equal response to infection challenge, this matrix is symmetric.

From the above arguments, a plausible transmission matrix is therefore:

$$\boldsymbol{\beta} = \begin{pmatrix} 10 & 0.1 \\ 0.1 & 1 \end{pmatrix},$$

and we shall use this particular example in the illustrative calculations that follow. We suppose that high-risk individuals make up 20% of the population, with the remaining 80% being low risk. Finally, we make the convenient assumption that $\gamma = 1$.

3.1.1.2. Initial Dynamics

For unstructured models, the single parameter β was vital in determining the basic reproductive ratio, R_0, and hence the rate of increase in infection following invasion. A naive proposition might be to assume that there is a specific R_0 for each class, and this can be calculated from the expected number of secondary cases a primary case (in a particular group) would cause. So, using the above matrix, the R_0 for the high-risk group is 2.08 ($R_0^H = 10 \times 0.2 + 0.1 \times 0.8 = [\beta_{HH} n_H + \beta_{LH} n_L]/\gamma$), whereas the R_0 for the low-risk group is 0.82 ($R_0^L = 0.1 \times 0.2 + 1 \times 0.8$). For the disease-free state $S_H = n_H < 1$ and $S_L = n_L < 1$, and so the size of the risk groups enters the calculation.

Although such formulations provide some insight into the dynamics (generally allowing us to bound the possible rates of increase following invasion), soon both risk groups contain infected individuals and their dynamics become *slaved*, increasing at the same exponential rate (Figure 3.2). The rate of increase in this slaved region is independent of the initial

seed of infection and determines whether or not a infection will successfully invade and, as such, it is related to our basic intuition about R_0. R_0 for the entire population lies between the values calculated for members of each group, and is generally greater than the weighted average if the WAIFW matrix is assortative. To calculate the actual value of R_0, an eigenvalue value approach is required to deal with the recursive nature of transmission (Diekmann et al. 1990; Heesterbeek 2002); Box 3.1 outlines this approach in more detail.

To keep the sense of the original verbal definition of R_0 (Chapter 2), for structured models we may wish to consider R_0 as:

the average number of secondary cases arising from an average infected individual in an entirely susceptible population, once initial transients have decayed

We therefore calculate R_0 from the distribution of infection across risk groups in the region of slaved dynamics, where the behavior is independent of the initial conditions. This slaved distribution provides a natural weighting for the number of secondary cases generated by a primary case in each group.

Box 3.1 Eigenvalue Approach

Let us consider infection dynamics in both classes following invasion. The standard mathematical way of doing this is to linearize the equations about the disease-free state. Simply put, during the initial invasion phase, the relative change to the number of susceptibles is small, and therefore we can fix their values at the disease-free equilibrium ($S_H = n_H$ and $S_L = n_L$). This leads to:

$$\frac{dI_H}{dt} \approx (\beta_{HH} n_H - \gamma_H) I_H + (\beta_{HL} \gamma_H) I_L$$
$$\frac{dI_L}{dt} \approx (\beta_{LH} n_L) I_H + (\beta_{LL} n_L - \gamma_L) I_L.$$

Note that for generality, different recovery rates have been allowed. Such a linear system of differential equations is understood by looking at the matrix of coefficients, $\boldsymbol{J}$:

$$\boldsymbol{J} = \begin{pmatrix} \beta_{HH} n_H - \gamma_H & \beta_{HL} n_H \\ \beta_{LH} n_L & \beta_{LL} n_L - \gamma_L \end{pmatrix} = \begin{pmatrix} 1 & 0.02 \\ 0.08 & -0.2 \end{pmatrix}$$

and its dominant eigenvalue $\lambda_1 \approx 1.0013$ (taking $\gamma_L = \gamma_H = 1$). This eigenvalue then determines the dynamics in the slaved phase:

$$I_H \propto \exp(\lambda_1 t) \qquad \text{and} \qquad I_L \propto \exp(\lambda_1 t).$$

Thus it is clear that when $\lambda_1 > 0$ the infection can successfully invade (c.f. $R_0 > 1$), and when $\lambda_1 < 0$ the infection will always die out (c.f. $R_0 < 1$).

The ratio of I_H to I_L is determined by the eigenvector ($\underline{e}_1$) associated with the maximum eigenvalue. We let I_H^s and I_L^s be the distribution of *infection* in the slaved region as determined by the eigenvector, specifying that $I_H^s + I_L^s = 1$. For the particular matrix under consideration $I_H^s = 0.9376$ and $I_L^s = 0.0624$, such that in the slaved region the high-risk group has about 15 times more infection than the low risk group.

We can now use this slaved distribution to weight our average of R_0 for each of the classes:

$$\begin{aligned} R_0 &= R_0^H I_H^s + R_0^L I_L^s, \\ &= \left(\frac{\beta_{HH} n_H + \beta_{LH} n_L}{\gamma_H}\right) I_H^s + \left(\frac{\beta_{HL} n_H + \beta_{LL} n_L}{\gamma_L}\right) I_L^s, \\ &\approx 2.08 \times 0.9376 + 0.82 \times 0.0624, \\ &\approx 2.0013. \end{aligned} \tag{3.4}$$

Equal Recovery Rates

Finally, for the simpler situation where the recovery rates within all classes are equal, the calculation of the basic reproductive ratio can be simplified and a more intuitive relationship to the single-class results can be obtained. If we define a matrix of the number of secondary cases produced in each class:

$$\boldsymbol{R} = \begin{pmatrix} \beta_{HH} n_H/\gamma & \beta_{HL} n_H/\gamma \\ \beta_{LH} n_L/\gamma_H & \beta_{LL} n_L/\gamma \end{pmatrix} = \begin{pmatrix} 2 & 0.02 \\ 0.08 & 0.8 \end{pmatrix},$$

then the dominant eigenvalue of this matrix is simply R_0. In addition, for the case of equal recovery rates, in the slaved region the growth rate is given by:

$$I \propto \exp([R_0 - 1]\gamma t),$$

which is the same relationship as found in the nonstructured models.

The basic reproductive ratio for the entire population is bounded by values for individuals in each group.

The basic reproductive ratio from structured models is generally larger than if the structures were ignored and all individuals had the same average transmission rates.

The basic reproductive ratio is found using an eigenvalue approach.

Figure 3.2 shows the dynamics for our simple example, starting with a few infectious individuals in the low-risk group. Initially, prevalence drops; this is because the majority of infecteds are in the low-risk group and its "reproductive ratio" in isolation is less than one. The initially steep rise of cases in the high-risk group is due to infection spreading from the low-risk class, governed by β_{HL}. The exact behavior in this transient period is determined by the initial conditions. Only by assuming that infection starts in the low-risk group do we obtain this counterintuitive decreasing-then-increasing behavior, the more common assumption that infection starts in the high-risk group leads to an increasing number of cases from the beginning. The converse is also true; it is possible to have a situation in which the total number infected increases before dying away to zero. With the same 20 : 80 ratio of high-to low-risk groups, the WAIFW matrix:

$$\boldsymbol{\beta} = \begin{pmatrix} 1 & 1.5 \\ 1.5 & 0.5 \end{pmatrix}$$

possesses this surprising behavior. If infection starts in the high-risk group, then this on average causes 0.2 ($= \beta_H H \times n_H$) cases in the high-risk group and 1.2 ($= \beta_L H \times n_L$) in

the low-risk group; a net increase to 1.4 cases. However, the eventual dynamics is given by the basic reproductive ratio, and for this matrix $R_0 = 0.9083$—which is less than one—hence in the long-term the disease fails to persist.

The dynamics of Figure 3.2 exemplify the three major phases observed in all structured models. During the initial transient phase, the dynamics are often complex and frequently related to the reproductive ratios for each group in isolation. During this early phase, the relative disease prevalence rapidly approaches the distribution predicted by the right eigenvector of the equilibrium—this is analogous to the "stable age distribution" of matrix models used in ecology (Caswell 2000). Once this distribution is achieved, the prevalence of infection in all groups grows at the rate determined by the basic reproductive ratio R_0, until the number of susceptibles is sufficiently depleted and density-dependent effects arise. We then enter the final asymptotic phase where prevalence levels tend to their equilibrium value.

The initial behavior of a structure model depends on the initial conditions, not just the basic reproductive ratio, R_0.

3.1.1.3. Equilibrium Prevalence

Calculation of the prevalence of infection at equilibrium is far from trivial. Mathematically, we need to find where the rates of change are zero; hence, remembering that $S_H = n_H - I_H$, we need to solve:

$$\begin{aligned} 0 &= \beta_{HH}(n_H - I_H)I_H + \beta_{HL}(n_H - I_H)I_L - \gamma I_H, \\ 0 &= \beta_{LH}(n_L - I_L)I_H + \beta_{LL}(n_L - I_L)I_L - \gamma I_L. \end{aligned}$$

These equations both contain quadratic terms, and therefore finding simple analytical solutions is generally impossible. We therefore have to rely on either numerical solutions of the above equilibrium equations, or more frequently we iterate the model forward to find the equilibrium levels.

For simple unstructured models, a clear relationship exists between the initial growth rate and the equilibrium density, $S(\infty) = 1/R_0$. In our example, we find (looking at Figure 3.2) that $I_H(\infty) \approx 0.1$ and $I_L(\infty) \approx 0.033$; the total proportion of the population that is susceptible is therefore $0.867 (= [0.2 - 0.1] + [0.8 - 0.033])$. However, our calculations and the observed rate of increase suggest that $R_0 \approx 2$. It is therefore clear that the simple relationship between equilibrium and invasion dynamics no longer holds in structured populations. In our example, the small high-risk group is responsible for the value of R_0, but due to the large low-risk group, the prevalence of infection is much smaller than unstructured theory would predict. This also implies that density-dependent saturation effects occur far earlier in risk-structured models compared to their homogeneous counterparts.

In many structured models that have high associativity, the equilibrium fraction of susceptibles is much higher, and density-dependent saturation effects occur earlier than in unstructured models (where $S(\infty) = 1/R_0$). Hence, although the equilibrium prevalence of infection is low, the disease may be difficult to eradicate because R_0 is still large.

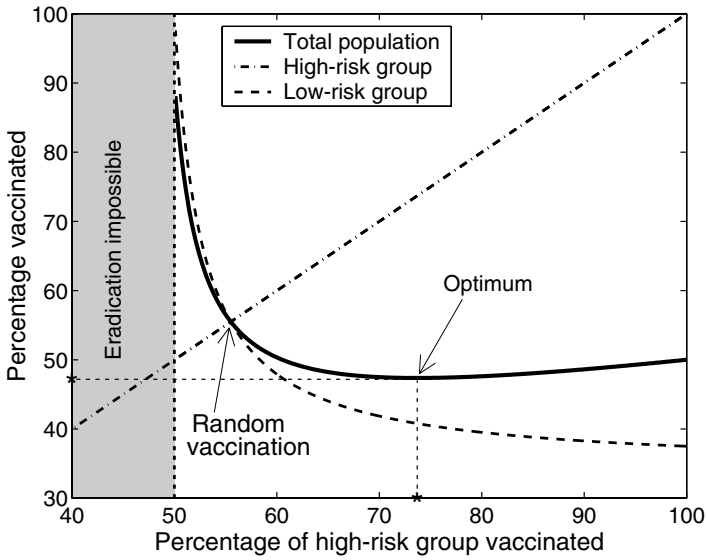


Figure 3.3. The critical level of vaccination needed to eradicate a STI, as a percentage of the entire population, for a range of coverage in the high-risk group. In this example, the transmission matrix is $\boldsymbol{\beta} = \begin{pmatrix} 10 & 1 \\ 1 & 2 \end{pmatrix}$, with $n_H = 0.2$, $n_L = 0.8$, and $\gamma = 1$ as before. For each level of vaccination within the high-risk group, we have searched for the corresponding level of vaccination in the low-risk group that sets R_0 equal to one. Optimal control, which minimizes the amount of vaccine used, is achieved when vaccinating about 75% and 40% of the high- and low-risk groups, respectively.

3.1.1.4. Targeted Control

In unstructured models, a further simple relationship exists between the basic reproductive ratio (or the equilibrium distribution) and the level of control, such as vaccination, required for eradication of the disease. Standard models predict that the critical level of vaccination p_C needed to eradicate infection is given by:

$$p_C \;=\; 1 - S(\infty) \;=\; 1 - 1/R_0.$$

We now wish to investigate whether this basic formula holds and if it can be improved by targeting the control measures at those individuals who are most at risk. We again want a specific example to illustrate the basic concepts behind targeted control, and this is given in Figure 3.3. We have adjusted the matrix from the previous example to better demonstrate the trade-offs. In this graph, we have taken a series of vaccination levels within the high-risk group (p_H, dot-dash line) and found the level of vaccination in the low-risk group (p_L, dashed line), which forces the effective reproductive ratio R (calculated in a similar manner to R_0) to be one. The total percentage of the population that needs to be vaccinated (solid line) is then a measure of the cost associated with eradication and the prevention of successful re-invasion.

Only for the very simplest of models can the calculation of optimal targeting be performed analytically. In general, for a given level of control (e.g., vaccination, screening, or quarantining), we must search through all possible deployments to find which produces

the more desirable results. In Figure 3.3, we explore the minimum level of vaccination that could set R_0 below one and hence eradicate infection. Alternatively, control measures may be optimally targetted to reduce the number of cases for a given control effort, or minimize the duration of an epidemic for given logistical limitations. Whatever the particular scenario, the quantity to be optimized (and any constraints) must be clearly specified by policy makers and planners. More examples of optimal control are discussed in Chapter 8.

For the transmission matrix in Figure 3.3, the basic reproductive ratio is around 2.25. After invasion, each case in the high-risk and low-risk groups causes 2.8 and 1.8 further cases, respectively. If we vaccinate at random (ignoring which group an individual belongs to), then the vaccination threshold is $1 - 1/R_0 \approx 55\%$, so we can always do as well as this standard result. However, as can be seen in Figure 3.3, the optimal strategy is to vaccinate nearly 75% of the high-risk group and only 40% of the low-risk class. A second point is that although excessive targeting of the high-risk group has only moderate adverse effects, failure to meet vaccination targets within this group has severe penalties. In many real scenarios, the difference between high- and low-risk groups is far more extreme, in which case the incentive for optimal targeting is far greater.

In structured models, the critical level of vaccination that eradicates infection is the same as in unstructured models, $1 - 1/R_0$, if vaccination is applied at random.

Targeting vaccination or other control measures generally works far better than random control. It is generally better to over-target the most at-risk groups rather than under-target.

3.1.1.5. Generalizing the Model

The matrix formulation for $\boldsymbol{\beta}$ can be readily adapted to model the interaction of multiple groups (e.g., high-, medium- and low-risk groups). Those infected individuals in group i obey the following differential equation:

$$\frac{dI_i}{dt} = \sum_j \boldsymbol{\beta}_{ij} S_i I_j - \gamma_i I_i, \tag{3.5}$$

where the matrix form of $\boldsymbol{\beta}$ is again used to parameterize transmission between the groups. The above set of equations can be recast into vector notation (as shown in Box 3.2), which can simplify the formulation and in some cases speed the computational solution. The value of R_0, and therefore whether the infection can successfully invade or not, is again determined by an eigenvalue approach as illustrated in the Box 3.1. When there are a large number of classes, this technique is particularly valuable.

This is online program 3.2

3.1.1.6. Parameterization

The main key to applying such structured models is successful parameterization. For unstructured models there were few parameters to estimate. The infectious period (and latent period, if necessary) can usually be estimated from careful observation of infected

Box 3.2 Vector Notation

We have already stated that the contact rates are now most naturally specified as a matrix of values. To correspond to this notational change, we specify the number of infected (or susceptible) individuals in each class as a vector. In this notation the full set of equations becomes:

$$\frac{d\underline{I}}{dt} = \underline{S} \otimes (\boldsymbol{\beta}\underline{I}) - \underline{\gamma} \otimes \underline{I},$$

where $\otimes$ refers to the piecewise multiplication of two vectors.

hosts; this leaves a single transmission parameter β. The traditional means of finding this value is through the use of the relationship $S^* = \frac{\gamma}{\beta}$, where the equilibrium level of susceptibles ($S^* = 1 - I^* - R^*$) can be estimated from either serological surveys or long-term case records.

For structured models, the parameterization is far less straightforward. Whereas the low number of parameters needed for unstructured models allows us to sweep through all possible configurations and develop an intuitive understanding, the complexity and variety of structured models means that they are more reliant on good data with which they can be parameterized. The first step is to determine the appropriate risk groups and the proportions within them. The divisional structure of the model will come from good epidemiological evidence and should hopefully correspond with the structure of any sampled data. The infectious/latent period can once again be found from observing infected hosts. However, we now need a matrix of transmission rates and yet we have only a vector of serological or other information. So, for a general structured model with n distinct classes, we require n^2 transmission terms, but we have at most one piece of information for each class (e.g., we might know S_i^*). The usual way to deal with this lack of specificity is to assume a simplified structure for the transmission matrix; this approach is dealt with in more detail in the next section which considers age-structured models.

Because the transmission matrix generally has more terms than the structured data, simplifications are needed to overcome this deficit in information.

For sexually transmitted infections, there is a very natural and parsimonious means of subdividing the population: using the number of sexual partners. We let class i refer to those individuals who have exactly i partners within a given period. If we then assume that individuals form partnerships at random, but in proportion to their expected number of partners, then the form of the matrix $\boldsymbol{\beta}$ is given, with only a scaling parameter to be determined:

$$\boldsymbol{\beta}_{ij} = \beta \frac{ij}{\sum_k k n_k}, \tag{3.6}$$

where, as usual, n_k is the proportion of individuals in class k. This approach has proved extremely popular and highly successful, as it has the huge advantage that only a single parameter (the scalar β) needs to be estimated. We can now examine what effect this form of heterogeneity has on the basic reproductive rate R_0. Because of the random formation

Box 3.3 Random Partnership Model

Given the situation where partnerships are formed at random but in proportion to the expected number of partners, the contact matrix is given by:

$$\boldsymbol{\beta}_{ij} = \beta \frac{ij}{\sum_k kn_k}.$$

To calculate R_0 for this epidemic model, it is easier and more informative to go back to first principles rather than use the eigenvalue approach. Irrespective of who is initially infected, due to the random nature of the mixing, the distribution of infection across the classes soon asymptotes:

$$\frac{I_i}{I} \rightarrow \frac{in_i}{\sum_k kn_k}, \qquad \text{where } I = \sum_i I_i$$

and remains in this distribution throughout the slaved period. This means that during this early growth phase the level of infection within a class (I_i/n_i) is proportional to the number of partners. We can thus find R_0 as the number of new cases we expect to be caused by infection with this distribution:

$$R_0 = \text{Number of secondary cases produced per primary case,}$$

$$R_0 = \frac{1}{\gamma}\sum_{i,j} S_i \beta_{ij} I_j / I = \frac{1}{\gamma}\sum_{i,j} n_i \beta \frac{ij}{\sum_k kn_k}\frac{jn_j}{\sum_k kn_k}$$

$$= \frac{\beta}{\gamma}\frac{\sum_i in_i \sum_j j^2 n_j}{\left[\sum_k kn_k\right]^2} = \frac{\beta}{\gamma}\frac{M(M^2+V)}{M^2} = \frac{\beta}{\gamma}\frac{M^2+V}{M},$$

where M and V are the mean and variance of the number of sexual partners.

of partnerships, the eigenvalue approach is no longer necessary, and we find:

$$R_0 = \frac{\beta}{\gamma}\frac{M^2+V}{M},$$

where M and V are the mean and variance of the number of sexual partners (greater detail is given in Box 3.3). If the heterogeneities had been ignored, and everyone assumed to have the same average number of partners, then the basic reproductive ratio is reduced to $\frac{\beta M}{\gamma}$. Hence even though we assume that partnerships are formed at random, heterogeneities can play a major role as the infection "focuses" on those high-risk individuals who are both more likely to catch the infection, and also more likely to transmit it.

Assuming individuals form contacts at random and proportional to their expected number of partners provides a very natural means of specifying the matrix $\boldsymbol{\beta}$. We find that R_0 is increased due to the variance in the number of partners.

We can make this approach more explicit by returning to our simple model with high-risk and low-risk groups. If we assume that individuals in the high-risk group have an average of 5 new partners per year, whereas those in the low-risk group average 1 only,

then the matrix becomes:

$$\boldsymbol{\beta} = \beta \begin{pmatrix} \dfrac{5\times 5}{0.2\times 5+0.8\times 1} & \dfrac{5\times 1}{1.8} \\ \dfrac{1\times 5}{1.8} & \dfrac{1\times 1}{1.8} \end{pmatrix} \approx \beta \begin{pmatrix} 13.9 & 2.78 \\ 2.78 & 0.556 \end{pmatrix},$$

for which the mean and variance are $M = 1.8$, $V = 2.56$, and calculating the basic reproductive ratio by either the eigenvalue or the method outlined above gives $R_0 = 3.22\beta/\gamma$.

Although this random partnership assumption captures many aspects of the transmission of STIs, it loses the property of assortative mixing where individuals are more likely to form partnerships with others in the same or similar classes. Two mechanisms can be used to overcome this problem. First, given a small amount of extra information on disease incidence, we could add a second parameter, α, to the transmission matrix that scales the proportion of within-group mixing. (More precisely, in the example given below, α is the fraction of partnerships that are forced to be made within the same class, with the remainder still being formed at random.) For example,

$$\boldsymbol{\beta}_{ij} = \beta(1-\alpha)\frac{ij}{\sum_k kn_k} \qquad \boldsymbol{\beta}_{ii} = \beta\left[\alpha\frac{i}{n_i} + (1-\alpha)\frac{i^2}{\sum_k kn_k}\right],$$

where although the distribution of transmission from an individual may change, the total amount of transmission remains constant.

In this model, as α increases, the matrix becomes closer to diagonal and the degree of assortative mixing becomes larger. When α is close to one, the eigenvalue of the matrix is dominated by the largest diagonal element; hence

$$R_0 \approx \frac{\beta}{\gamma}\max_i(i),$$

which is always larger than the random partnership formation example. Figure 3.4 illustrates this change in R_0 for the simple high-risk/low-risk example given above. We always expect more assortative mixing to lead to larger values of R_0; intuitively this is because as assortativity increases, spread is focused with the high-risk with less being "wasted" by infecting low-risk individuals. Interestingly, as assortativity increases, the targeting of vaccination may also change. For the first scenario considered in Figure 3.4, it is never worthwhile to vaccinate the low-risk group because they cannot maintain a chain of transmission because each infected individual can infect at most only one susceptible person. In contrast, the second example shows that although greater assortative mixing focuses more of the early growth within the high-risk class, the optimal vaccination strategy is increasingly spread between the two groups. This surprising result can be understood by realizing that when the mixing is random, it is always more efficient to concentrate vaccination on the high-risk individuals; however, when the mixing is completely assortative ($\alpha = 1$), then the risk groups act independently and the vaccination level in each must be sufficient to control transmission. More formally, the degree of assortative mixing for any matrix can be measured (Gupta et al. 1989), the details of which are given in Box 3.4.

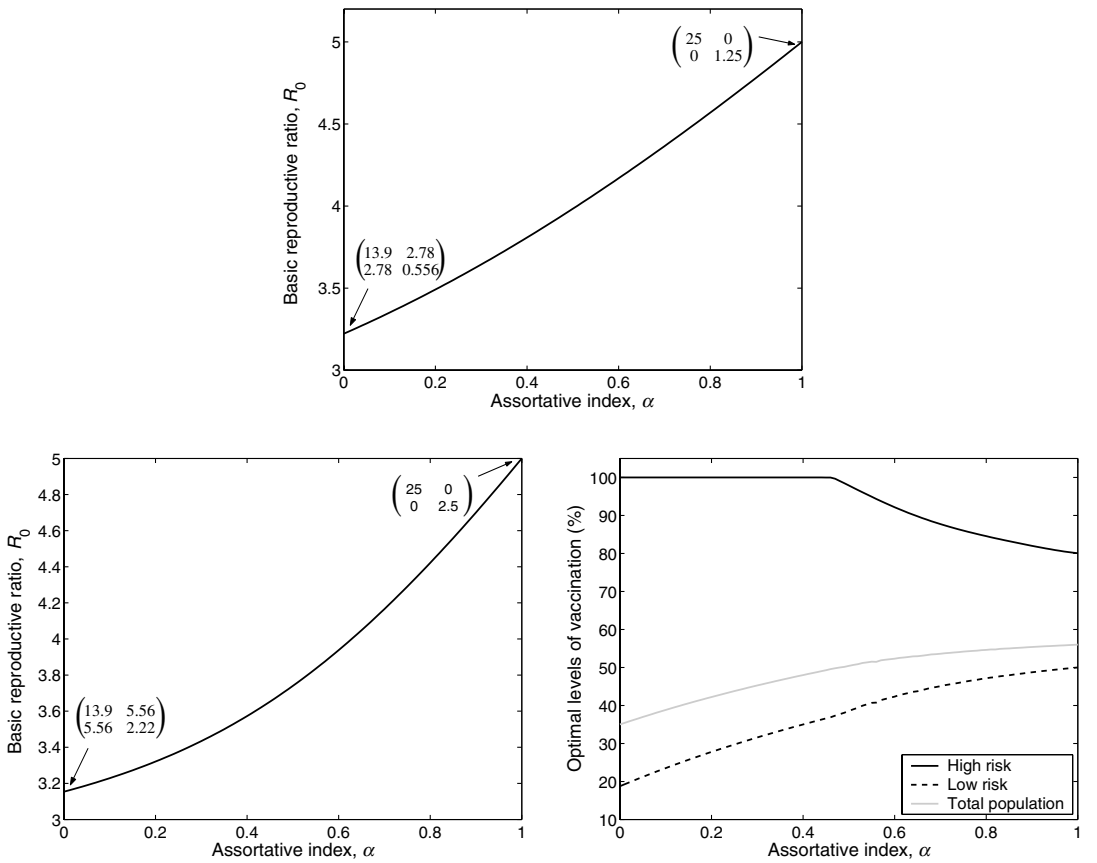


Figure 3.4. The effect of assortativity on the epidemic behavior of a population split into two classes, a high-risk group that comprises 20% of the population and a low-risk group that makes up the remaining 80%. In the top graph, individuals in the high-risk group have an average of five partners per year, whereas the low-risk group averages only one per year. In this situation it is always optimal to focus control exclusively on the high-risk group. For the bottom graphs, members of the high-risk group again average five partners per year, but now the low-risk group averages two partners per year. In this latter case, the optimal control strategy is a complex mix between the two groups. (For simplicity of presentation we have assumed that $\beta = \gamma = 1$.)

Increased assortative mixing increases R_0, but may mean that the optimal vaccination strategy is less targeted toward the high-risk groups.

A second approach to fully parameterize the transmission matrix is to utilize the very detailed data that has been collected on sexual partnership networks. In a few isolated examples this method has allowed researchers to reconstruct the full network of sexual partners within a population (Klovdahl 1985; Potterat et al. 2002). Although this type of network reconstruction may be highly sensitive to the occasional missing connection (which can vastly alter the topology of the network), the transmission matrix that emerges is far more robust, although some elements of the network structure are lost (see Chapter 7).

Box 3.4 Degree of Assortative Mixing

Given a contact matrix $\boldsymbol{\beta}$ and a vector of the proportion of individuals in each risk-group $\underline{n}$, we first define a new matrix $\widehat{\beta}_{ij} = \beta_{ij} n_i$. We next define the matrix $\boldsymbol{B}$, which is the relative proportion of transmission from one group to all others:

$$B_{ij} = \frac{\widehat{\beta}_{ij}}{\sum_{i=1}^{M} \widehat{\beta}_{ij}},$$

where M is the number of classes being modeled. The degree of assortative mixing, Q, can be determined by comparing the relative amounts of within-group transmission to what is expected from the random formation of partnerships (Gupta et al. 1989):

$$Q = \frac{\sum_{i=1}^{M} B_{ii} - 1}{M - 1}.$$

As such, a matrix that comes from random formation of partnerships will have degree $Q = 0$, whereas one with complete assortative mixing (all diagonal elements) will have $Q = 1$. However, this weights all risk groups equally, therefore a more appropriate measure might be to utilize the eigenvalues of the matrix.

The values of the minor eigenvalues $(\widehat{\lambda}_2 \ldots \widehat{\lambda}_M)$ relative to the dominant eigenvalue $(\widehat{\lambda}_1)$ of $\widehat{\beta}$ provides a more natural description of the spread of infection between classes. In particular,

$$q = \frac{\widehat{\lambda}_2}{\widehat{\lambda}_1}$$

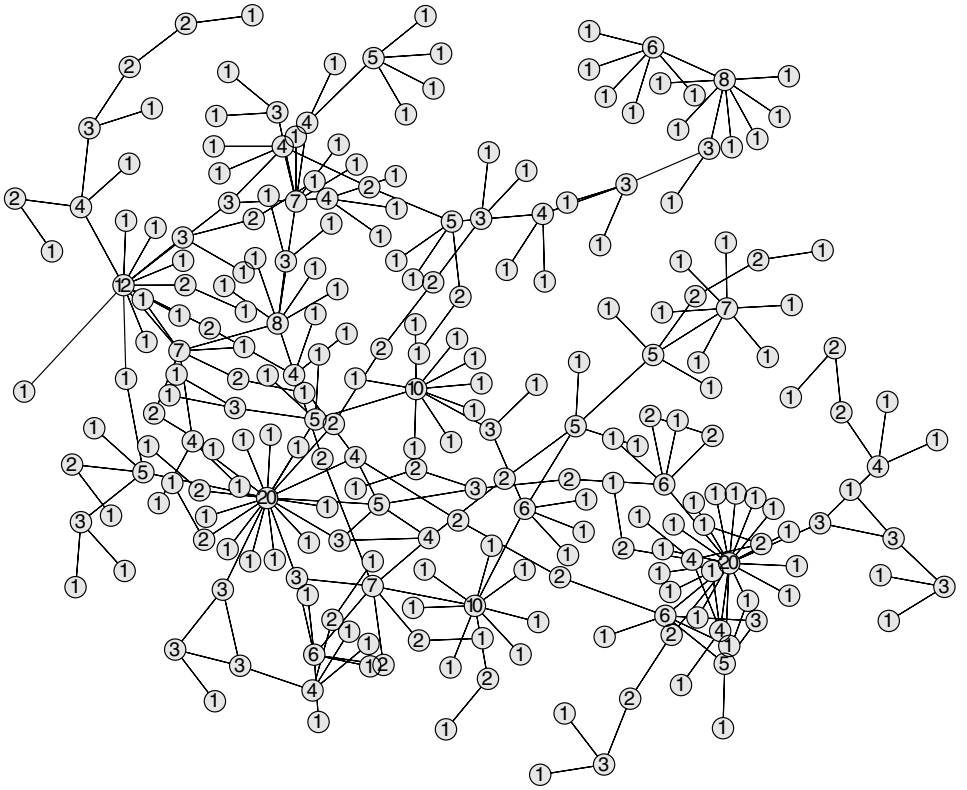
is an alternative measure of assortative mixing, which effectively weights mixing between similar classes.

Figure 3.5 shows an example of a detailed traced network from Colorado Springs together with the associated transmission matrix that arises. For this matrix, the degree of assortative mixing can be calculated as $Q = -0.0776$ or $q = 0.2452$ (see Box 3.4), hence showing that the matrix is either almost random, or only moderately assortative. This type of approach has been advocated by Garnett and Anderson (1993b), where individuals are classified into three basic groups (core, adjacent, and peripheral), and where the partnerships within and between the group are determined by detailed interviews (Rothenberg 1983).

Detailed information on the exact network of sexual partners can be used to parameterize the transmission matrix.

3.1.2. Two Applications of Risk Structure

We now consider two applications from the literature that complement the methodology detailed above. The first comes from the vast body of work done on the early spread of the HIV epidemic in developed countries, and using detailed epidemiological and sociological data illustrates many of the points already made. The second example of an STI comes from the work on chlamydia in koalas, because this provides a radical departure from the standard models and demonstrates some more unusual attributes associated with wildlife diseases.



$$\boldsymbol{\beta} \approx \beta \begin{pmatrix}
0 & 0.8 & 1.6 & 2.5 & 3.2 & 4.8 & 4.4 & 8.1 & 12.1 & 9.7 & 17.7 \\
0.8 & 3.1 & 1.9 & 4.8 & 4.3 & 6.9 & 8.7 & 3.5 & 3.5 & 13.9 & 13.9 \\
1.6 & 1.9 & 5.7 & 6.1 & 7.1 & 6.8 & 5.7 & 11.4 & 5.7 & 22.7 & 28.4 \\
2.5 & 4.8 & 6.1 & 5.9 & 9.6 & 0 & 19.2 & 0 & 0 & 19.2 & 28.8 \\
3.2 & 4.3 & 7.1 & 9.6 & 7.8 & 12.5 & 7.8 & 0 & 15.6 & 0 & 62.5 \\
4.8 & 6.9 & 6.8 & 0 & 12.5 & 0 & 12.5 & 25 & 250 & 0 & 50 \\
4.4 & 8.7 & 5.7 & 19.2 & 7.8 & 12.5 & 0 & 62.5 & 31.3 & 62.5 & 0 \\
8.1 & 3.5 & 11.4 & 0 & 0 & 25 & 62.5 & 0 & 0 & 0 & 0 \\
12.1 & 3.5 & 5.7 & 0 & 15.6 & 25 & 31.3 & 0 & 0 & 0 & 0 \\
9.7 & 13.9 & 22.7 & 19.2 & 0 & 0 & 62.5 & 0 & 0 & 0 & 0 \\
17.7 & 13.9 & 28.4 & 28.8 & 62.5 & 50 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}$$

Figure 3.5. An example of a sexual contact network, taken from the study of HIV transmission in Colorado Springs (Potterat et al. 2002). The matrix below the network, is the associated mixing matrix for individuals with 1, 2, 3, 4, 5, 6, 7, 8, 10, 12, and 20 partners, and is calculated as $\boldsymbol{\beta}_{ij} \propto$ (number of i–j partnerships)$/(n_i n_j)$.

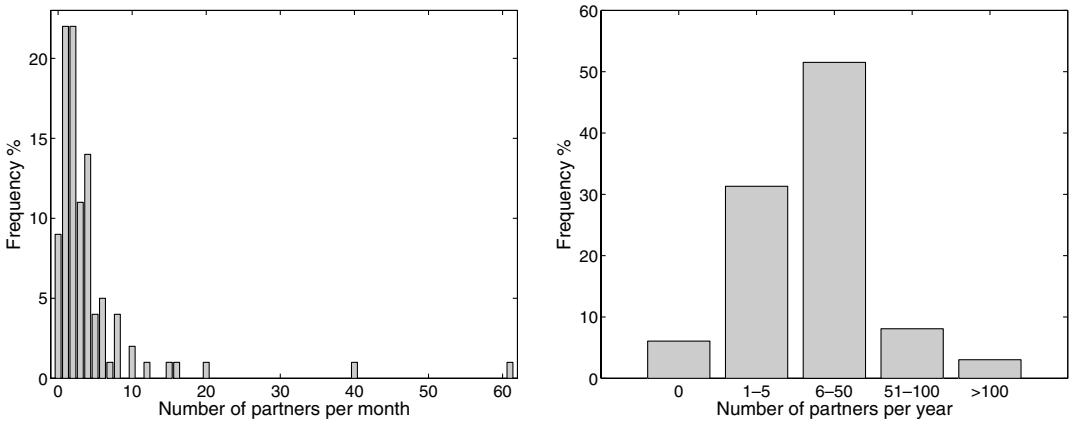


Figure 3.6. Frequency distribution of the number of sexual partners of male homosexuals from two studies in and around London in the mid-1980s. (Data from Anderson and May 1991, taken from original surveys by Carne and Weller (unpublished), and McManus and McEvoy (1987).)

3.1.2.1. Early Dynamics of HIV

To illustrate the use of the methodology already described, we focus on the early work that attempted to predict the spread of HIV in male homosexual communities (Anderson et al. 1986; Kolata 1987; May and Anderson 1987; Jacquez et al. 1988; May and Anderson 1989; Bongaarts 1989). This narrow focus has the distinct advantage that the models used at the time had a small number of risk classes and so the parameterization and dynamics can be shown explicitly.

Recent models, which focus on the more immediate problem of HIV spread in sub-Saharan Africa and elsewhere, are often much more complex either with a large number of risk classes or explicitly model the transmission network (Anderson et al. 1992; Arca et al. 1992; Garnett and Anderson 1993a; Downs et al. 1997; Koopman et al. 1997; Artzrouni et al. 2002; Koopman 2004). Therefore, although such models essentially follow the same principles outlined above, the degree of complexity and number of parameters required makes them infeasible as illustrative examples. A complete description of the spread of an STI would involve multiple overlapping classes, such as gender, age, level of sexual activity, sexual preference, drug use, condom use, visits to sex workers, and so forth, as well as transmission parameters that depend on time since infection (Longini et al. 1989). As such, for any complete model, there could be thousands of classes and at least as many parameters. Here, we will concentrate on the work of Anderson et al. (1986) and May and Anderson (1987), who used data on the number of partners per year together with the assumption of random partnership formation to develop simple mechanistic models. Many model assumptions are clearly vast simplifications of the underlying system, but these models help to demonstrate how an element of risk may be included.

The underlying mixing matrix comes from detailed social studies of the distribution of the number of sexual partners over a given period. The data we were recollected from homosexual males attending STI clinics in and around London, and shows a significantly skewed distribution, with a significant proportion of those questioned having a very large number of partners (Figure 3.6). Clearly these data will be biased by the sampling methods

and are probably focused toward the high-risk groups that are more likely to be infected with an STI and therefore more likely to attend a clinic. However, similar patterns of partners have been recorded in other studies (McKusick et al. 1985). Since then, more detailed and large-scale studies have been performed, most notably by Johnson et al. (1994), who attempted to document the range and distribution of sexual practises in the United Kingdom—such data could be used to study the spread of STIs in the population as a whole, rather than a restricted class of high-risk individuals. The presence of a small core group of high-risk individuals makes the use of risk-structured models necessary for understanding the spread and persistence of sexually transmitted infections in general and HIV in particular.

We are now in a position to develop a model of this epidemic. For simplicity and brevity we adopt the data from the right-hand graph of Figure 3.6, and assume that there are five classes with 0, 3, 10, 60, and 100 partners per year in each class. This gives a matrix and population distribution

$$\boldsymbol{\beta} = \beta \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0.65 & 2.15 & 12.9 & 21.5 \\ 0 & 2.15 & 7.17 & 43.1 & 71.8 \\ 0 & 12.9 & 43.1 & 258 & 431 \\ 0 & 21.5 & 71.8 & 431 & 718 \end{pmatrix} \qquad \underline{n} = \begin{pmatrix} 0.06 \\ 0.31 \\ 0.52 \\ 0.08 \\ 0.03 \end{pmatrix}.$$

The full set of equations is therefore:

$$\begin{aligned} \frac{dS_i}{dt} &= \nu_i - \textstyle\sum_j \boldsymbol{\beta}_{ij} S_i I_j - \mu S_i, \\ \frac{dI_i}{dt} &= \textstyle\sum_j \boldsymbol{\beta}_{ij} S_i I_j - \mu I_i - \gamma I_i, \\ \frac{dA_i}{dt} &= \gamma d I_i - \mu A_i - m A_i. \end{aligned} \tag{3.7}$$

For greater realism, the class A of infected individuals that have developed AIDS is modeled separately, with d being the proportion of infected individuals who develop AIDS. It is assumed that individuals with AIDS curb their sexual behavior and therefore no longer contribute to the spread of infection. "Births" or recruitment, (ν_i), have been added to the susceptible equation, and natural deaths, μ, and AIDS-induced mortality, m, have also been included. This precise form of the equations is specific to HIV, a disease that does not obey the assumptions of the standard SIS framework typical of sexually transmitted infections. Anderson et al. (1986) quote a value for R_0 of around 5; this comes from examination of the doubling time of the epidemic during the early stages and allows us to fix the multiplicative parameter which is part of the mixing matrix. Using the theory developed earlier (Box 3.3), we can calculate the value of R_0 from the given matrix:

$$R_0 = 5 \approx 46.14 \frac{\beta}{\gamma},$$

where 46.14 is the largest eigenvalue of the associate matrix. Other parameters for the model are $\mu = 0.0312$, $\nu_i = \mu n_i$, $\gamma = 0.2$, $m = 1$, and $d = 0.3$, with all rates in years (Anderson et al. 1986). These high birth and natural death rates merely reflect the rate at which individuals enter and leave the sexually active class. From these parameters, we find that $\beta \approx 0.0217$.

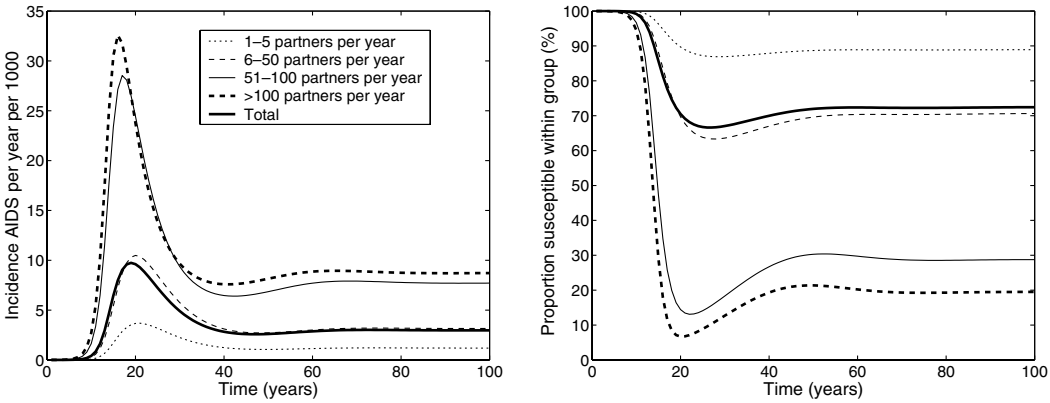


Figure 3.7. The dynamics of HIV infection from the structured model, equation (3.7), showing the behavior within the four classes (the zero class is ignored) and the average behavior. The simulations are begun with 1 infected individual in a population of 100,000. The left-hand graph shows the incidence of AIDS (those entering the A_i classes) per year—this can be substantially different from the prevalence of the infection but corresponds more closely to the information that is recorded. The right-hand graph shows the percentage of each class that are susceptible, S_i/n_i, which provides some indication of the force of infection experienced.

Figure 3.7 shows the dynamics of this system; the results for the class with no partners have not been plotted because there is no risk of these individuals becoming infected. From the left-hand graph, the slaved exponential growth phase lasts for about 10–15 years. The peak in incidence occurs first in the highest-risk group, and because the low-risk groups are primarily infected from more high-risk groups, the other peaks lag by about 3 years. As expected, the high-risk groups show a far greater incidence of HIV. After the peak, the incidence drops, showing damped oscillations toward an equilibrium solution. This type of dynamics is uncharacteristic of sexually transmitted infections and is because the mortality (removal) of infected individuals and recruitment of new susceptibles effectively makes the dynamics SIR (see Chapter 2). This oscillatory nature makes it difficult to assess the impact of control strategies from case report data; the natural decline predicted in years 20 to 40 could easily mask any control efforts. Switching attention to the proportion of susceptibles within each class, we find that although only 20% of the high-risk group are susceptible at equilibrium, for the population as a whole the susceptibility is more than 70%. In a homogeneous model, this would be in direct conflict with the R_0 value of 5, whereas in this structured population R_0 is driven by the small high-risk group.

The parameters used in this model are taken from a select subset of the population (male homosexuals), whose behavior puts them at greater risk than the norm. Therefore, all results pertain to this small subset and not the population as a whole. However, two interesting and robust conclusions can be ascertained from such models: (1) the expected peak numbers of AIDS cases will be much higher (double for the parameters used here) than the equilibrium level, and (2) the time-scale of the epidemic dynamics is long, taking 40–50 years before equilibrium is reached. Over such long time scales social changes may have a significant impact on the dynamics.

Although inherently simple and easy to parameterize from available data, this model demonstrates the power of modeling as a tool for understanding disease dynamics and predicting long-term trends. Obviously, more accurate predictions require more complex

models with more classes; however, although such models are relatively easy to formulate, their parameterization is more awkward and obtaining the necessary social and sexual-behavior data is a difficult and time-consuming task (Johnson et al. 1994). However, several key features could be included to make this a more realistic model of HIV transmission. Infection risk for HIV is known to be highly dependent of time since infection, which can be modeled by subdividing the infectious class (see Section 3 of this chapter). Transmission through shared needles by intravenous drug users is another important route and one that could be modeled by a similar risk-structured approach. Finally, greater realism of sexual behavior could be included, capturing both the assortative nature of contacts, as well as both the heterosexual and homosexual communities.

3.1.2.2. Chlamydia Infections in Koalas

Localized koala (*Phascolarctos cinereus*) populations have undergone significant declines in recent years (Phillips 2000). One possible cause of this decline is infectious disease, which has been demonstrated to regulate other natural populations (Hudson et al. 1998). Chlamydia infections in koalas are known to lead to sterility of females and increased mortality. It therefore seems likely that chlamydia will have an impact on koala population dynamics. Here, we follow the work of Augustine (1998), although we consider a much reduced model to highlight the factors we have already discussed. The original paper used a discrete-time model to mimic the seasonal reproduction of koalas (see Chapter 5), but also included age-structure (see Section 3.2) and stochasticity (see Chapter 6) to achieve greater realism.

This example considers the interaction between epidemiological and ecological interactions. For human populations, there are little or no density-dependent effects; in contrast, most wildlife populations experience considerable density-dependence, limiting their numbers (e.g., Clutton-Brock et al. 1997; Gaillard et al. 2000). The interplay between the epidemiological and ecological factors can substantially complicate the infection dynamics and negate some of our understanding based upon human disease models (Hudson et al. 2001). This issue is further discussed in Chapter 8.

For this sexually transmitted infection, we do not structure the population by the number of sexual partners, but instead males and females are treated as the two risk groups. This gender-based structuring is necessary due to the very different behavior of males and females, both in terms of transmission and in terms of response to infection. A similar argument could be made for models of chlamydia transmission in humans. Finally, the assumption of heterosexuality within the koala population means that the transmission matrix $\boldsymbol{\beta}$ takes a relatively simple form.

We first assume that the underlying population dynamics of koalas can be described by a continuous-time logistic growth model, with density-dependent mortality, and that the carrying capacity is normalized to one. We now structure the population into males and females, such that in the absence of infection the number of susceptible males and females obeys the following differential equations:

$$\frac{dX_F}{dt} = rX_F - rX_F N, \qquad \frac{dX_M}{dt} = rX_F - rX_M N, \qquad N = X_F + X_M.$$

Thus, in an uninfected population, the koalas have a carry-capacity of 1 ($N = 1$) and an equal ratio of males and females. Infection with chlamydia is assumed to be lifelong, and transmission is modeled as frequency dependent (mass-action), which agrees with observations of koala having a fixed number of mates per year irrespective of population

size or density. We initially consider only horizontal transmission (between male and female koala) during intercourse, which leads to the following set of equations:

$$\begin{aligned}
\frac{dX_F}{dt} &= r(X_F + \alpha Y_F) - rX_F N - \beta_{FM} X_F Y_M / N, \\
\frac{dX_M}{dt} &= r(X_F + \alpha Y_F) - rX_M N - \beta_{MF} X_M Y_F / N, \\
\frac{dY_F}{dt} &= \beta_{FM} X_F Y_M / N - rY_F N - mY_F, \\
\frac{dY_M}{dt} &= \beta_{MF} X_M Y_F / N - rY_M N - mY_M, \\
N &= X_F + X_M + Y_F + Y_M.
\end{aligned}$$

where α is the reduction in fertility due to disease, and m is the additional disease-induced mortality rate. There is negligible vertical transmission, so a mother does not pass infection to her offspring and all newborns are susceptible. It is interesting to consider the structure of the transmission matrix $\boldsymbol{\beta}$ in more detail. Based on the work of Augustine (1998), a plausible matrix is:

$$\boldsymbol{\beta} = \begin{pmatrix} 0 & 1.0 \\ 1.2 & 0 \end{pmatrix}.$$

First, we note that the diagonal terms β_{MM} and β_{FF} are zero, which breaks the common assumption of assortative mixing (in fact $Q = -1$, because there are only two classes)—this is because we are interested in transmission between males and females instead of "risk groups," per se. The second is that the matrix is nonsymmetric because females are more likely to catch chlamydia from an infected male than vice versa, although this is outweighed by the fact that breeding males are likely to have more mates. When the koala population is at its disease-free equilibrium, $X_F = X_M = \frac{1}{2}$; this simple, but unconventional, transmission matrix leads to a basic reproductive ratio of:

$$R_0 = \frac{\sqrt{1.0\frac{1}{2} \times 1.2\frac{1}{2}}}{r+m} = \frac{0.5477}{r+m}.$$

Thus, although high disease mortality (large m) may be problematic for the koala population, it can be devastating for chlamydia because it can easily reduce R_0 below one.

Figure 3.8 illustrates the effect the disease can have on the breeding population of koalas, by plotting $X_F + \alpha Y_F$ as a function of both α and m. We observe that if infected females have sufficiently reduced fecundity (as in humans, chlamydia infection is likely to cause sterility), then the entire population can be driven extinct. In contrast, the behavior with respect to disease mortality, m, is more complex and intermediate levels of mortality have the most detrimental effects on the total population size. This is because with very high mortality the infection has a low R_0 and cannot spread. At the other extreme, although low mortality allows the pathogen to spread further it does little to reduce the population size. These results echo the findings of Augustine (1998) and illustrate why detailed mathematical models are often necessary and insightful.

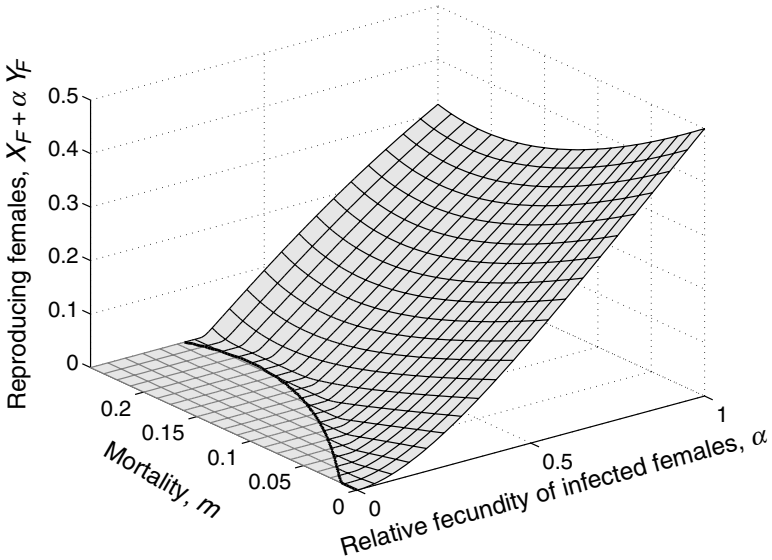


Figure 3.8. The level of reproducing females, $X_F + \alpha Y_F$, as a function of two infection attributes: disease-induced mortality rate, m, and reduced fertility, α. When infected females have low fertility it is possible for the infection to drive the population to extinction. We take $r = 0.2$; all parameters are within the ranges given by Augustine (1998).

3.1.3. Other Types of Risk Structure

The formulation of risk-structured models and the WAIFW matrices provide a useful distinction between two types of important individuals in airborne infections: *super-shedders* and *super-spreaders* (Austin and Anderson 1999b; Riley et al. 2003). Although rare, both of these types of individuals can be responsible for a disproportionately large number of secondary cases and, therefore, can severely hamper control efforts. Let us consider a population in which only a small proportion are either super-shedders or super-spreaders and consider how the transmission matrix $\boldsymbol{\beta}$ reflects their intrinsic differences.

Here, we define super-shedders as individuals who once infected excrete large amounts of the infectious agent. This super-shedding is often due to genetic attributes of the host or a compromised immune system. Although able to produce many more secondary cases, super-shedders are not necessarily at any greater risk of coming into contact with the infections disease. Therefore, labeling super-shedders with a subscript S and the rest of the population with a subscript R, a suitable transmission matrix would be:

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{SS} & \beta_{SR} \\ \beta_{RS} & \beta_{RR} \end{pmatrix} = \begin{pmatrix} f\beta & \beta \\ f\beta & \beta \end{pmatrix},$$

where $f > 1$ reflects the greater transmission *from* super-shedders in comparison with the rest of the population. Because super-shedders are assumed to have different epidemiological responses to infection compared to the rest of the population, the WAIFW matrix is no longer symmetric.

In contrast, we define super-spreaders as individuals with a very high number of contacts, often due to their occupation. Hence these individuals could generate many

secondary cases but are also at a much higher risk of being infected. Therefore, using subscript S to now represent super-spreaders, a suitable transmission matrix would be:

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{SS} & \beta_{SR} \\ \beta_{RS} & \beta_{RR} \end{pmatrix} = \begin{pmatrix} f^2\beta & f\beta \\ f\beta & \beta \end{pmatrix},$$

where $f > 1$ reflects greater transmission from and to super-spreaders. It should be clear from this formulation that super-spreaders generate a much greater risk than super-shedders, because their behavior means that they are much more likely to be infected and subsequently much more likely to transmit.

A recent approach to capture heterogeneities in transmission has been proposed by Lloyd-Smith et al. (2005). They examined contact tracing data for eight directly transmitted diseases, including measles, smallpox, pneumonic plague, and SARS, and found the distribution of individual infectiousness around R_0 to be often rather skewed. To study the epidemiological implications of this, Lloyd-Smith et al. (2005) used results from branching process theory (see Chapter 6) by assuming that the number of secondary cases resulting from each infected is described by an "offspring distribution" given by $Z \sim$ negative binomial (R_0, k), where the parameter k captures the skew in the transmission distribution (note: $k \to \infty$ gives the Poisson distribution and $k = 1$ yields the geometric). The values of k estimated from data ranged from 0.01 to approximately 0.1, highlighting the large variance in individual infectiousness. These results suggest the 20/80 "rule" (whereby 20% of individuals are responsible for 80% of transmission)—which was previously thought to apply to STIs and vector-borne diseases—may also apply to directly transmitted infectious diseases.

The epidemiological implications of these observations are interesting and important. For example, compared to models that assume little individual variation in transmission (k very large), this documented heterogeneity implies an increased disease extinction risk, and a reduced likelihood of epidemics, though outbreaks are more severe when they do occur. Additionally, the public health lessons highlight the importance of individual-specific control measures, rather than population-wide approaches. Let us assume control effort c is imposed ($c = 0$ represents no control and $c = 1$ reflects the full blockage of transmission). Under population-wide control, the infectiousness of every individual in the population is reduced by c—control is homogeneous. Alternatively, with individual-specific control, a proportion c of infecteds (chosen at random) are traced and isolated completely such that they cause zero infections (also see Chapter 8)—control is heterogeneous. Individual-specific control raises the degree of heterogeneity in the outbreak as measured by the variance-to-mean ratio of Z, whereas population-wide control reduces heterogeneity. Both approaches yield the effective reproductive number $R = (1 - c)R_0$, so the threshold control effort for guaranteed disease extinction is $c \geq 1 - 1/R_0$, as in conventional models (compare Chapter 8). For intermediate values of c, however, the individual-specific approach always works better because higher variation favors disease extinction.

3.2. AGE-STRUCTURE: CHILDHOOD INFECTIONS

The dynamics of childhood infections, such as measles, mumps, chickenpox, rubella, or whooping cough, have been extensively studied by epidemiologists and mathematical modelers (Hamer 1906; Bartlett 1957; Black 1966; Yorke and London 1973; Cliff et al.

1981; Fine and Clarkson 1982; Schenzle 1984; Schwartz 1983; McLean and Anderson 1988a,b; Olsen and Schaffer 1990; Rand and Wilson 1991; Grenfell 1992; Grenfell et al. 1994; Bolker and Grenfell 1995; Ferguson et al. 1996a,b; Keeling and Grenfell 1997a; Finkenstädt and Grenfell 1998; Earn et al. 2000; Grenfell et al. 2001; Bjørnstadt et al. 2002; Rohani et al. 2002). From a theoretical perspective, the study of such diseases is facilitated by the relative simplicity of the natural history of the infection, the large number of cases, the high quality of the recorded data, and the interesting dynamics that emerge. From an epidemiological and public health standpoint, these diseases are important to understand because they have high R_0 and can therefore produce large epidemic outbreaks and, although usually fairly benign, complications and mortality can occur (Rohani et al. 2003).

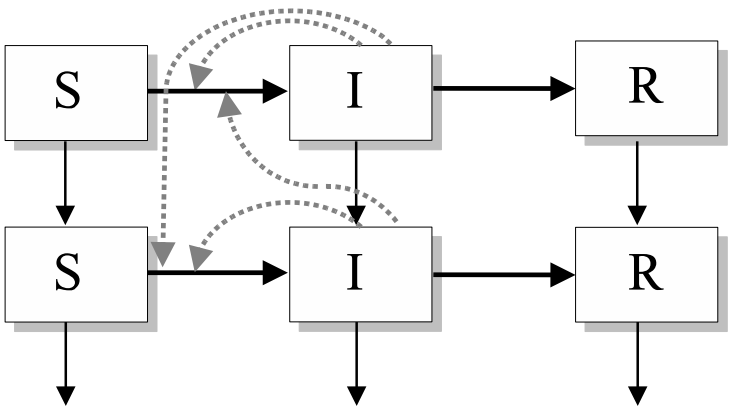
Although the standard theory for unstructured models provides many insights into the dynamics of such infections—especially when seasonal forcing is included (see Chapter 5)—these diseases predominantly affect a subset of the population (children) and hence any models developed should reflect this. The formalism of such age-structured models is superficially the same as the risk-structured models described above, but with the addition of individuals aging. Thus, whereas with risk-structured models hosts are generally assumed to stay in the same risk class for their entire lives, with age-structured models hosts move sequentially through the age classes.

In many countries, these childhood diseases are notifiable with local and national statistics being compiled. The primary reason for collecting these data is public health motivated, enabling health agencies to determine the effects of control measures and providing an insight into the future epidemic behavior. However, these statistics are also a rich source of data for model parameterization and time-series analysis. In England and Wales, the number of cases reported weekly in over 1,400 locations has been collected since 1944, providing the largest such epidemiological data set in existence (Cliff et al. 1981; Finkenstädt and Grenfell 1998; Grenfell et al. 2001). To date, most interest has been focused on measles, which has been recorded with an efficiency of around 60% (Finkenstädt and Grenfell 1998), however data on other childhood diseases is now being examined (Rohani et al. 1999, 2000). In addition, the use of age-structured models is particularly useful when considering pandemic infections, because a significant amount of transmission could occur within schools, and closing schools may be a relatively easy way of restricting disease spread (Ferguson et al. 2006).

3.2.1. Basic Methodology

The standard approach is again to subdivide the population into a number of discrete compartments, classified by the age of the host. Although for sexually transmitted infections there was some arbitrary choice to the qualities used to divide the population, here the situation is much more clear cut and only the age range within each compartment has to be determined. In principle, age is a continuous variable (which would suggest a partial-differential-equation approach, see Box 3.5), but because children are generally grouped into school classes of a given age cohort, the compartmental approach is often more realistic.

Although a continuous parameter, age-structured models usually group individuals into a limited number of classes—often representing school years.



Box 3.5 PDE Model

This type of model considers the variables S, I, and R to be functions of both age and time. As such, $S(a,t)$ refers to the probability-density of susceptibles of age a at time t. The age-dependent SIR equations can now be written as a PDE:

$$\frac{\partial S(a)}{\partial t} = \nu\delta(a) - S(a)\int_0^\infty \beta(a,a')I(a',t)da' - \mu S(a) - \frac{\partial S}{\partial a},$$

$$\frac{\partial I(a)}{\partial t} = S(a)\int_0^\infty \beta(a,a')I(a',t)da' - \mu I(a) - \gamma I(a) - \frac{\partial I}{\partial a},$$

where the delta-function forces all births to be age zero, and the last partial derivative term in each equation accounts for the population getting older. Although this formalism has its mathematical elegance, it is computationally difficult to implement and would be solved numerically by discretization into multiple age classes. Finally, the contact rate β, which is now a function of two continuous ages, would be very difficult to parameterize from real data.

We again start by considering a model that subdivides the population into two classes, in this case children and adults identified by subscripts C and A, respectively. Although most childhood diseases have a clearly defined latent-period leading to $SEIR$-type dynamics, for clarity we begin with an age-structured version of the SIR equations:

$$\frac{dS_C}{dt} = \nu - S_C(\beta_{CC}I_C + \beta_{CA}I_A) - \mu_C S_C - l_C S_C,$$

$$\frac{dI_C}{dt} = S_C(\beta_{CC}I_C + \beta_{CA}I_A) - \gamma I_C - \mu_C I_C - l_C I_C,$$

$$\frac{dS_A}{dt} = l_C S_C - S_A(\beta_{AC}I_C + \beta_{AA}I_A) - \mu_A S_A,$$

$$\frac{dI_A}{dt} = l_C I_C + S_A(\beta_{AC}I_C + \beta_{AA}I_A) - \gamma I_A - \mu_A I_A.$$

This is online program 3.3

The parameters μ_C and μ_A are the age-specific death rates, l_C is the rate that individuals mature (leave the childhood class and move to the adult class), and ν is the birth rate (assuming no maternally derived protection). In this model, the interaction between the two classes comes from transmission between them (as captured by the off-diagonal elements of the matrix $\boldsymbol{\beta}$) and due to the slow trickle of individuals moving from the childhood to the adult class. Again, we are dealing with parameters that represent proportions of the entire population, such that $S_C = X_C/N$ is the proportion of the population that is susceptible children. We now seek to compare the general dynamics observed for this type of model with those of the risk-structured approach, contrasting with the behavior of nonstructured models.

Age-structured models differ from the earlier risk-structured models due to the regular progression of individuals into increasingly older age classes.

3.2.1.1. Initial Dynamics

The first observation is that when the transmission dynamics are rapid in comparison with both host demography and the aging process, the initial behavior and in particular the basic reproductive ratio are largely unchanged from the risk-structured results given earlier, where there was no movement between risk classes. Although this assumption is generally true for common childhood infections of humans, if the infection dynamics are particularly slow relative to the life expectancy (for example BSE in cattle, see Section 3.2.2), we need to use the eigenvalues approach for the full system of equations, which includes the effects of movement between age classes while infected. This can be done in a similar manner to described in Box 3.1, finding the eigenvector distribution of infecteds in the slaved growth phase and using this to weight the individual number of secondary cases expected for each age class.

3.2.1.2. Equilibrium Prevalence

When contrasting risk-structured models of STIs and age-structured models of childhood infections, two distinct elements lead to differences in the equilibrium distribution of infection: (1) the nature of the infectious disease and the inherent differences between SIS- and SIR-type models, and (2) the sequential progression through the age classes, which means that at equilibrium, the proportion of susceptible individuals in the older age classes must be lower,

$$\frac{S_i^*}{n_i} \geq \frac{S_j^*}{n_j} \qquad \forall i < j.$$

For diseases conferring life-long immunity, this must be true because hosts in older age classes must have been subjected to at the very least the same risks of infection as hosts in younger classes. An extreme version of this is the case where the force of infection experienced is independent of age, in which case the fraction susceptible decreases exponentially with age. We note, however, that if immunity to infection can wane, then this result no longer holds, because older individuals may have had sufficient time to recover.

For SIR-type infections at equilibrium, the proportion recovered (measures by seroprevalence) must increase with age.

From this understanding, the equilibrium serological profile (which measures the proportion of the population that have been exposed to the infection) with respect to age can be used to estimate both the incidence of infection and the force of infection acting within that age class. Let us consider the ith age class, and assume that this class lasts for an average of T_i years. Individuals who enter the ith age class are susceptible with probability S_{i-1}/n_{i-1}, whereas when individuals leave the ith age class the probability of being susceptible has dropped to S_i/n_i. Thus, ignoring mortality, during the T_i years a fraction $S_{i-1}/n_{i-1} - S_i/n_i$ must have been infected; this sets the force of infection in this age class to be:

$$\lambda_i = \frac{1}{T_i}\left(\frac{S_{i-1}}{n_{i-1}} - \frac{S_i}{n_i}\right)\frac{n_i}{S_i} = \frac{1}{T_i}\left(\frac{S_{i-1}n_i}{S_i n_{i-1}} - 1\right)$$

and the equilibrium disease prevalence is:

$$I_i^* = \frac{1}{\gamma T_i}\left(\frac{S_{i-1}n_i - S_i n_{i-1}}{n_{i-1}}\right). \tag{3.8}$$

As such, the gathering of age-structured seroprevalence information is a vital step in understanding the dynamics and relative transmission strengths of childhood infections (Figure 3.10).

3.2.1.3. Control by Vaccination

For risk-structured models, it was clear that targeting controls toward the high-risk groups was the most efficient means of combating infection. However, this was because individuals in the high-risk group were assumed to remain there indefinitely. For age-structured models, most individuals make it through all the age classes and therefore experience, and contribute to, the full range of dynamic behaviors. In such situations, and assuming the vaccine provides lifelong protection, it is always best to vaccinate as early as possible so that immunity covers the greatest proportion of the host's lifespan.

For age-structured models, when vaccination offers lifelong protection, it is always best to target the youngest age groups.

Figure 3.9 shows the equilibrium results from an age-structured model with just two classes, children and adults. To allow comparisons to previous work, we take the contact matrix to be a multiple of the risk-structured matrix seen previously:

$$\boldsymbol{\beta} = \begin{pmatrix} 100 & 10 \\ 10 & 20 \end{pmatrix},$$

with children having the higher transmission rate, and an infectious period of length 0.1 years. (This set of parameters gives dynamics very similar to those in Figure 3.2, but operating ten times faster). Finally, we also assume that the childhood class is from ages 0 to 15 years, and that life expectancy is 75 years. Setting $\mu_C = 0$, this implies that $n_C = 0.2$, $n_A = 0.8$, $l_C = 0.0667$, and $\mu_A = 0.0167$ per year. Clearly, epidemiological dynamics are on a much faster time scale than the demography—and so the birth and death rates have an insignificant impact on $R_0 \approx 2$.

Figure 3.9 also illustrates a few basic points that were made earlier. First, a smaller proportion of adults are susceptible compared to children, although obviously the two

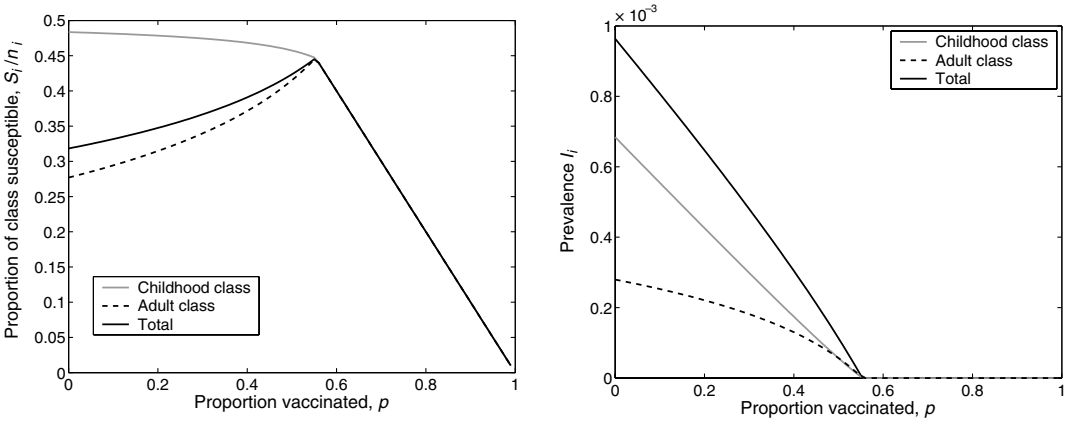


Figure 3.9. Equilibrium results from an age-structured model with vaccination at birth, such that the birth rate of susceptible children is reduced to $\nu(1-p)$. The left-hand graph shows the proportion of each age class that are susceptible (S_i^*/n_i). The right-hand graph shows the absolute prevalence of infection in each class as well as the total.

levels are equal once the disease has been eradicated ($p > 0.55$). However, notice that vaccination actually increases the proportion of the adult class who are susceptible to infection (S_A/n_A) and hence leads to an increase in the proportion of the entire population that is susceptible ($S_C + S_A$). These results can be explained as follows: With increasing vaccination at birth, the decrease in prevalence is greater than the decrease in newborn individuals who are susceptible—although there are fewer susceptible births and even fewer infectious cases and hence the total fraction of susceptibles increases. Considering disease prevalence (right-hand graph), although the adult class contains four times as many individuals as the childhood class, it contains fewer infecteds. Vaccination can also be seen to have a stronger effect on prevalence among children. This is comparable with earlier results (see Chapter 2)—that a lower reproductive ratio (in this case $(1-p)R_0$) corresponds to a higher average age of infection, which means that proportionally fewer cases occur in children. Finally, we find that vaccinating 55.35% of children at birth is sufficient to control the epidemic, which is just slightly less than the standard $1 - 1/R_0$. Pragmatically, it is reassuring that despite substantial heterogeneity and assortative mixing in the transmission matrix, the simple eradication threshold remains a surprisingly accurate approximation when vaccinating newborns. An intuitive explanation for this close agreement is that, as the newborns age, eventually the entire population is vaccinated at a proportion p, and hence this is asymptotically equivalent to random vaccination. We note, however, that for vaccination at birth, the dynamics just after the start of a vaccination campaign may be very different from the standard random-vaccination models (see Chapter 8).

3.2.1.4. Parameterization

Although risk-structured models for STIs often have a natural parameterization of the contact matrix, age-structure models are more difficult to parameterize (Grenfell and Anderson 1985). We return to the problem stated earlier of having an $n \times n$ matrix of values to find, and only a vector of n observed values to use. The standard way to cope

with this sort of problem is to reduce the number of terms needed in the matrix, usually focusing on the age classes that are most responsible for infection transmission.

In our simple two-class example, we may have data on the proportion of susceptibles in each class. Thus, given two pieces of information, we need to specify the matrix with just two parameters. The most natural way to do this is to isolate the childhood interactions, such that the contact matrix becomes

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 & \beta_2 \\ \beta_2 & \beta_2 \end{pmatrix}.$$

Clearly we have lost some of the model structure, but it is now possible to estimate this reduced matrix. As shown in equation (3.8), it is possible to estimate the equilibrium prevalence of infection from the fraction of susceptibles:

$$I_C^* = \frac{1}{\gamma y_C}\left(n_C - S_C^*\right) \qquad I_A^* = \frac{d_A}{\gamma}\left(\frac{S_C^* n_A - S_A^* n_C}{n_C}\right),$$

where y_C and $1/d_A$ are the average times spent in the childhood and adult classes. For this solution to be at equilibrium, we require the transmission matrix parameters to satisfy

$$\frac{dI_C}{dt} = 0 = S_C^*\left(\beta_1 I_C^* + \beta_2 I_A^*\right) - \gamma I_C^* - m_C I_C^*,$$

$$\frac{dI_A}{dt} = 0 = S_A^*\left(\beta_2 I_C^* + \beta_2 I_A^*\right) + m_C I_C^* - \gamma I_A^* - \mu_A I_A^*.$$

After some rearranging, this gives:

$$\beta_2 = \frac{\gamma I_A^* + \mu_A I_A^* - m_C I_C^*}{S_A^*(I_C^* + I_A^*)}, \qquad \beta_1 = \frac{\gamma + m_C}{S_C^*} - \frac{I_A^* \beta_2}{I_C^*}.$$

We can test this result using the transmission matrix defined earlier. The results in Figure 3.9 show that in the absence of vaccination, $S_C^* \approx 0.097$ ($S_C^*/n_C \approx 0.485$) and $S_A^* \approx 0.22$ ($S_A^*/n_A \approx 0.275$). This translates into infection levels of $I_C^* \approx 6.8 \times 10^{-4}$ and $I_A^* \approx 2.8 \times 10^{-4}$, which agrees well with the full numerical calculations, and produces a reduced matrix with $\beta_1 \approx 98.8$ and $\beta_2 \approx 12.9$, which again is in line with the actual values.

The choice of matrix is a question of good epidemiological judgment. Any matrix that contains two free parameters can be made to fit the serological data, but may poorly reflect the general understanding of how transmission operates. In the above example, we chose to focus our attention on transmission between children as distinct from all other transmission events, however we could have equally made the assumption that all transmission involving children was equal and that only transmission between adults was different. Finally, there are a range of bizarre and unlikely matrices:

$$\boldsymbol{\beta}_{unlikely} = \begin{pmatrix} \beta_1 & \beta_2 \\ \beta_2 & \beta_1 \end{pmatrix}, \begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{pmatrix}, \begin{pmatrix} 0 & \beta_1 \\ \beta_2 & 0 \end{pmatrix}, \ldots$$

all of which could be made to fit the available data, but would not agree with our intuition about the transmission of childhood diseases.

This method can, of course, be extended to deal with higher dimensional models with more age classes, although the algebra gets increasingly ugly (Grenfell and Anderson 1985). However, one particular case is worth describing in some detail. The most common

formulation for age-structured models of childhood diseases is to divide the population into four age-classes: pre-school (0–5 years), primary school (5–11 years), secondary school (11–16 years), and adults (16 or over). Thus, the division of the population reflects the standard epidemiological knowledge that it is the mixing of school children (mainly at primary schools) that is responsible for driving the progression of most childhood diseases (Fine and Clarkson 1982; Schenzle 1984; Finkenstädt and Grenfell 2000). There is some degree of freedom in the way that the reduced transmission matrix $\boldsymbol{\beta}$ is chosen; clearly, primarily school contacts must play a dominant role, but there is no clear general format. Two possible formulations are

$$\boldsymbol{\beta} = \begin{pmatrix} b_2 & b_2 & b_3 & b_4 \\ b_2 & b_1 & b_3 & b_4 \\ b_2 & b_3 & b_3 & b_4 \\ b_4 & b_4 & b_4 & b_4 \end{pmatrix} \qquad \boldsymbol{\beta} = \begin{pmatrix} b_2 & b_4 & b_4 & b_4 \\ b_4 & b_1 & b_4 & b_4 \\ b_4 & b_4 & b_3 & b_4 \\ b_4 & b_4 & b_4 & b_3 \end{pmatrix},$$

where $b_1 < b_2 < b_3 < b_4$. In the first formulation, the interaction strength is determined by the weakest member, and it is assumed that primary-school children followed by preschool and secondary school children are most responsible for transmitting infection. In the second formulation, the degree of assortative mixing is high, and transmission to outside the age class is considered to be less important. When seasonal forcing is included to mimic the effects of opening and closing schools, generally only the b_1 parameter is forced (Chapter 5).

Age-structured data on the proportion of seropositives can be used to determine n terms of the $n \times n$ transmission matrix. The way these n terms are used to define the entire matrix should match the underlying biology.

3.2.2. Applications of Age Structure

We again consider two examples from the recent literature. The first is an amalgamation of the huge body of research into the dynamics of measles. This is one of the best understood of all childhood diseases, and as we will show, the implications of age structure are vital if we are to model the observed dynamics. The second example is the control of BSE (*Bovine Spongiform Encephalopathy*), where age-structured models are again necessary if we are to understand the dynamics of this infection and curtail its spread.

3.2.2.1. Dynamics of Measles

As noted in Chapter 2, the distinguishing feature of measles and other childhood diseases is their large basic reproductive ratio, R_0. For a non-age-structured population of constant size, this translates into an average age at first infection of (life expectancy)/$(R_0 - 1)$, hence a large basic reproductive ratio coupled with life-long immunity is synonymous with childhood infections. As with all infections, the value of R_0, estimated from age-structured seroprevalence data, varies between locations. For measles in modern era England and Wales, several studies have estimated R_0 to be around 17 (Anderson and May 1982; Grenfell 1982; Schenzle 1984; Grenfell and Anderson 1985), and this value is used throughout this section. Age structure is important in the modeling of childhood

diseases because of the greater mixing between susceptible and infected children that occurs at schools. In general, although preschool children are the most likely age group to be susceptible, they generally mix only with a small number of other children so the potential to catch and transmit infection is low. However, as soon as children enter school, the number of potential contacts increases to at least class size (20–30 children), and hence the risk of transmission also rises.

The seminal work of Schenzle (1984) modeled measles dynamics using an $SEIR$ framework (Chapter 2) and 21 age classes, where the members of each of the four school groups (preschool school $0 \rightarrow 6$, primary school $6 \rightarrow 10$, secondary school $10 \rightarrow 20$, and adults) share a common mixing pattern. Thus, the mixing matrix is comprised of 16 blocks of identical transmission terms. This early work into realistic age-structured (RAS) models has been greatly extended and expanded by Grenfell and coworkers (Bolker and Grenfell 1993; Bolker 1993; Grenfell et al. 1994; Keeling and Grenfell 1997a; Finkenstädt and Grenfell 1998) using the extensive spatio-temporal record of measles cases available for England and Wales. The values given in Bolker and Grenfell (1993), for a RAS model similar to Schenzle's, are:

$$\boldsymbol{\beta} = \gamma \left(\begin{array}{c|c|c|c} 1.875 \ldots 1.875 & 2.175 \ldots 2.175 & 0.975 \ldots 0.975 & 0.6 \\ \vdots & \vdots & \vdots & \vdots \\ 1.875 \ldots 1.875 & 2.175 \ldots 2.175 & 0.975 \ldots 0.975 & 0.6 \\ \hline 2.175 \ldots 2.175 & 10.74 \pm 8.56 \ldots 10.74 \pm 8.56 & 0.975 \ldots 0.975 & 0.6 \\ \vdots & \vdots & \vdots & \vdots \\ 2.175 \ldots 2.175 & 10.74 \pm 8.56 \ldots 10.74 \pm 8.56 & 0.975 \ldots 0.975 & 0.6 \\ \hline 0.975 \ldots 0.975 & 0.975 \ldots 0.975 & 0.975 \ldots 0.975 & 0.6 \\ \vdots & \vdots & \vdots & \vdots \\ 0.975 \ldots 0.975 & 0.975 \ldots 0.975 & 0.975 \ldots 0.975 & 0.6 \\ \hline 3 \ldots 0.6 & 0.6 \ldots 0.6 & 0.6 \ldots 0.6 & 0.6 \end{array}\right),$$

where individuals are assumed to be exposed (but noninfectious) for an average of eight days and infectious for an average of five days—leading to $SEIR$-type dynamics. Rather than maturation being a continual process, to better mimic the school structure, all individuals except adults are assumed to move up an age class at the start of each school year. Finally, the ± 8.56 terms in the primary-school interactions refer to the changes in mixing that occurs between school terms and holidays (Chapter 5). In Schenzle's original formulation, Sundays are treated as school holidays, although this high-frequency variation has little effect on the dynamics.

Figure 3.10 shows model results together with actual data for measles. The results shown, using Schenzle's formulation, ignore temporal forcing and instead show the equilibrium level of seropositives (effectively the proportion recovered within each age class R_i/n_i) for the term time and holiday matrices separately. The boxes show the maximum and minimum levels over each year, and hence correspond to the seropositive levels just before and just after everyone moves up into the next age class. The high degree of mixing between primary school children during school terms means that

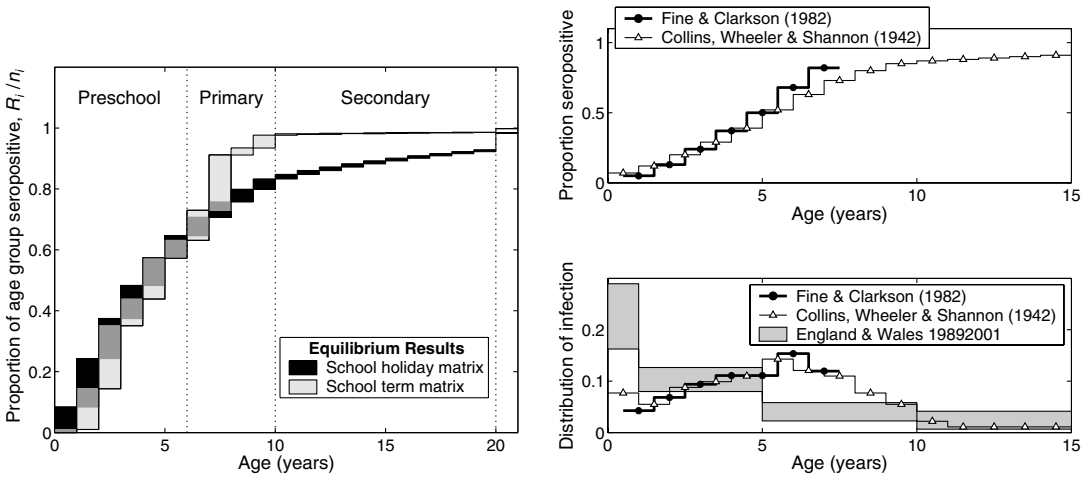


Figure 3.10. Left-hand graph: results from an age-structured model of the same form as Schenzle (1984), showing the level of seropositives across age classes. The boxes show the maximum and minimum levels across the yearly cycle of aging. The results ignore the complexities of temporal forcing and are given for the equilibrium values using the term-time and holiday matrices separately. The right-hand graphs show the measured level of seropositives and the derived distribution of cases from two studies of measles in Europe before vaccination: Collins et al. (1942) and Fine and Clarkson (1982). The lower graph also shows the recent age-structured distribution of cases in England and Wales from the Public Health Laboratory Service, giving some insight into the role of vaccination.

the risk of infection is greatest in these age classes, and therefore the proportion of seropositives increases dramatically. The difference between the school holiday and school term distributions reflects the impact of the extra mixing that occurs within primary schools.

The two right-hand graphs of Figure 3.10 present data from England and Wales (Fine and Clarkson 1982; PHLS 1989–2001) and Germany (Collins et al. 1942). The top graph gives the fraction of seropositives in each age class (R_i/n_i) before the onset of mass vaccination against measles; both data sets show a larger increase in seropositives during the primary school years, with the increase in England and Wales being more pronounced. This information can be translated into prevalence levels ($I_i \propto R_{i+1}/n_{i+1} - R_i/n_i$), and the lower graph shows the distribution of infection ($I_i / \sum_k I_k$) across all age classes for the two prevaccination samples. Prevalence clearly peaks in children around age 5–6. Also plotted on the same axes are the maximum and minimum of the modern case-report data for England and Wales (taken from `www.phls.co.uk`); mass vaccination has clearly driven a substantial shift in the age structure of those individuals who become infected. In recent years, infection has been concentrated in children under one year of age, and presumably occurs in the window between the loss of maternally derived immunity and the age at which the child is vaccinated.

The age-structured nature of the population can be simplified even further, by just dealing with four distinct groups. This was the approach used by Keeling and Grenfell (1997a), with the four classes: preschool (0–5), primary school (6–9), secondary school (10–19) and adult (20+). In this formulation, each age class contains several yearly cohorts; therefore, only a fraction of each class moves to the class above each school year. The

$SEIR$-type model is therefore:

$$\begin{aligned}
\frac{dS_i}{dt} &= \nu_i n_4 - \textstyle\sum_j \boldsymbol{\beta}_{ij} I_j S_i - \mu_i S_i, \\
\frac{dE_i}{dt} &= \textstyle\sum_j \boldsymbol{\beta}_{ij} I_j S_i - \sigma E_i - \mu_i E_i, \\
\frac{dI_i}{dt} &= \sigma E_i - \gamma I_i - \mu_i I_i, \\
\frac{dR_i}{dt} &= \gamma I_i - \mu_i R_i,
\end{aligned} \tag{3.9}$$

This is online program 3.4

and at the start of the school year, moving up an age group is controlled by:

$$\begin{aligned}
Q_1 &= Q_1 - Q_1/6 \\
Q_2 &= Q_2 + Q_1/6 - Q_2/4 \\
Q_3 &= Q_3 + Q_2/4 - Q_3/10 \\
Q_4 &= Q_4 + Q_3/10
\end{aligned} \qquad \text{where } Q \in \{S, E, I, R\}.$$

Here $\nu_1 = \mu_4 = (365 \times 55)^{-1} = 4.98 \times 10^{-5}$ per day and all other ν and μ terms are zero, such that all individuals survive until adulthood (aged 20), the average life expectancy is 75 years, and birth and death are equal. We also set $1/\sigma = 8$ days and $1/\gamma = 5$ days to capture the known latent and infectious periods.

Although the use of just four age classes has conceptual and computational appeal, this simplicity comes at a price. In Schenzle's model with 21 age classes, individuals remain with a particular age cohort throughout their school lives, and each individual spends an exact number of years in each school group (i.e., exactly four years in the preschool group). In contrast, when just four age classes are used, the time spent in each age class and therefore in each school grouping is exponentially distributed. This in turn leads to effectively extra mixing between the age cohorts, and a smoother age-related serology pattern. (This difference between the 4 and 21 age-class models can be compared to the distinction between exponential and constant distributions discussed in Section 3.3 of this chapter). Clearly the 21 age-class model is a more realistic representation of the actual movement of children through the school system, however this requires the use of an extra 17 equations. It should also be noted that although the Schenzle model contains 21 age classes, only four distinct transmission values are still used in the transmission matrix $\boldsymbol{\beta}$. The ideal model would use the full 21 age classes and a transmission matrix with (at least) 21 independent transmission values—however, the historical data on the incidence of measles infection are not sufficiently detailed to support such a parameterization with so many degrees of freedom. We are therefore left with a choice between simplicity or greater realism.

For the reduced four age-class models (equation (3.9)), the matrix of parameters was estimated by fitting the deterministically predicted biennial epidemic curve from the temporally forced model to the available prevaccination (1948–1968) data from England

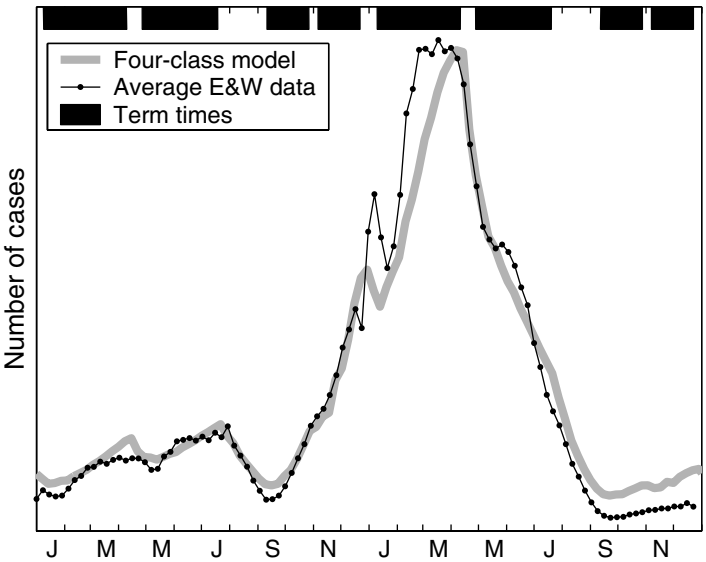


Figure 3.11. Results from a reduced age-structured model of a similar form to Keeling and Grenfell (1997a), where seasonal forcing is included to mimic the opening and closing of schools (Chapter 5). The dots show the average data from England and Wales over the duration of biennial epidemics, the thick gray line gives the deterministic attractor, and the black rectangles show the timing of school terms when the mixing matrix is increased.

and Wales. The best-fit matrix was found to be

$$\boldsymbol{\beta} = \gamma \begin{pmatrix} 2.089 & 2.089 & 2.086 & 2.037 \\ 2.089 & 9.336 \pm 4.571 & 2.086 & 2.037 \\ 2.086 & 2.086 & 2.086 & 2.037 \\ 2.037 & 2.037 & 2.037 & 2.037 \end{pmatrix},$$

such that only the primary-school class shows any significant difference from the norm. It was assumed that only the primary-school class experiences temporal forcing due to the opening and closing of schools (see Chapter 5). Aggregate results ($I = \sum_i I_i$) from this model are compared to the average biennial cycle in Figure 3.11. There is reasonable agreement between the deterministic model and the available data. We note, however, that the observed average epidemic peak is far broader than predicted by the deterministic simulations. In part, this is because different communities and different biennial segments of the data experience peak infection at slightly different times, and the averaging process smears the sharp peaks. We would need to include both stochasticity (Chapter 6) and spatial effects (Chapter 7) to fully correct this discrepancy.

This approach of fitting to the aggregate data is only successful for measles due to the high degree of synchrony between all the communities in England and Wales. For other childhood diseases (e.g., whooping cough), the spatial dynamics in the prevaccination era are far less synchronized (Rohani et al. 1999) and hence the average dynamics are an amalgamation of a variety of epidemic curves. In such cases, we would need to fit

simultaneously to the entire time-series of all communities taking into account the stochastic nature of transmission (Keeling and Grenfell 2002). A more practical approach that is generally taken is to match to some aggregate quality of the epidemic data, such as the power spectrum, so that model and data have a similar frequency of epidemics (Bolker 1993). However, it is probable that the transmission matrix for measles could act as a template for other childhood infections because similar routes of transmission are involved in their spread (Chapter 5).

Although more recent advances have shown that in some situations the age structure may be effectively replaced by more complex seasonality (Earn et al. 2000; Finkenstädt and Grenfell 2000; Bjørnstadt et al. 2000), these models lack the mechanistic framework of the more traditional RAS models. The complex seasonality in these models mimics changes in the average transmission rate that are due to changes in the distribution of infection across age classes over the biennial cycle. RAS models will therefore always be an essential tool in predicting the number of cases of childhood diseases, and the effects of vaccine uptake. The extra age-structured information is often particularly useful in targeting vaccination campaigns toward those age classes that are most at risk and modeling the short-term dynamics of control measures.

3.2.2.2. Spread and Control of BSE

In 1986, a disease known colloquially as "mad-cow-disease" was observed spreading through the cattle farms of the United Kingdom. Later identified as *bovine spongiform encephalopathy*, or BSE for short, this disease is caused by the transmission of a prion and is thought to have entered the cattle population via supplementary food stuffs containing meat and bonemeal from other cows. This feedback loop enabled the pathogenic agent to spread through the food chain. As well as a problem to the cattle industry, BSE also has important, but as yet unquantified (Ghani et al. 2000; Ghani et al. 2003a,b; Hilton et al. 2004) implications for human health, as consumption of infected meat can result in new variant *Creutzfeldt-Jacob Disease* (CJD) (Caughey and Chesebro 1997; Almond 1998; Narang 2001).

Mathematical models, explicitly including the age structure of the cattle population, were developed and played a major role in determining policy (Donnelly et al. 1997; Ferguson et al. 1997b). Here, we consider a simplified version of these models that illustrates the main mechanisms and results. The original equations took the form of complex integro-differential equations (see Box 3.5) and accounted for both the age of cattle, the time since infection, and farming practices in the United Kingdom. For greater clarity, the model presented here ignores the time since infection, and aggregates cattle into discrete age groups.

There are three main routes of infection for each cow: maternal (vertical) transmission, horizontal (cow-to-cow) transmission, and consumption of infected feed—cattle were routinely fed on high-protein food supplements containing meat and bonemeal from other cows. We again use an age-structured framework to study this infection, modeling the three transmission routes separately. The dynamics of infection is best captured as SEI, with infected cattle being destroyed as soon as clinical symptoms emerge. Due to the changing population size, we formulate the equations in terms of the number of cattle in each state, and model transmission as mass-action (frequency dependent) due to the density-independent nature of contacts within the farming industry. Separating out newborn and

TABLE 3.1.

Parameter	*Value*
ν_a	1 if $a > 2$ years, zero otherwise.
μ_a	0 if $a < 2$ years, 0.25 otherwise.
m	3
σ	0.2
l_a	4, using 3-month cohorts
β^M	0.082
$s_a^H = s_a^F$	0.3×0.55^a
τ_a^H	12
τ_a^F	160

older cattle, the equations are:

$$\begin{aligned}
\frac{dX_0}{dt} &= \textstyle\sum_b \nu_b(X_b + W_b + (1-\beta^M)Y_b) - \sum_b(\beta_{0b}^H Y_b + \beta_{0b}^F \mu_b(W_b + Y_b))X_0/N \\
&\quad - \mu_0 X_0 - l_0 X_0, \\
\frac{dW_0}{dt} &= \textstyle\sum_b \nu_b \beta^M Y_b + \sum_b(\beta_{0b}^H Y_b + \beta_{0b}^F \mu_b(W_b + Y_b))X_0/N - \mu_0 W_0 - l_0 W_0 - \sigma W_0, \\
\frac{dY_0}{dt} &= \sigma W_0 - \mu_0 Y_0 - l_0 Y_0 - m Y_0, \\
\frac{dX_a}{dt} &= l_{a-1} X_{a-1} - \textstyle\sum_b(\beta_{ab}^H Y_b + \beta_{ab}^F \mu_b(W_b + Y_b))X_a/N - \mu_a X_a - l_a X_a, \\
\frac{dW_a}{dt} &= l_{a-1} W_{a-1} + \textstyle\sum_b(\beta_{ab}^H Y_b + \beta_{ab}^F \mu_b(W_b + Y_b))X_a/N - \mu_a W_a - l_a W_a - \sigma W_a, \\
\frac{dY_a}{dt} &= l_{a-1} Y_{a-1} + \sigma W_a - \mu_a Y_a - l_a Y_a - m Y_a.
\end{aligned}$$

where, at age a, ν_a is the rate at which a cow gives birth, μ_a is the rate of slaughter, and l_a is the rate that an individual cow leaves an age class and matures into the next. The latent period of BSE is $1/\sigma$; m is the death rate due to infection; and β^M, β^H, and β^F refer to maternal transmission, horizontal transmission, and transmission via food supplements. The total population size is $N(= \sum_b X_b + W_b + Y_b)$. It is assumed that each transmission matrix (for horizontal and food supplements) β_{ab} is the product of an age-dependent susceptibility vector s_a and an age-dependent transmission vector τ_b. Parameter values are given in Table 3.1, where rates are all measured in months. Note that this is a highly infectious disease, with R_0 for the food supplement route alone estimated to be around 10 (Ferguson et al. 1999b).

We emphasize that the three transmission routes occur at very different stages. Maternal transmission can pass only from a mature cow (older that 2 years) to a newborn calf. Horizontal transmission can occur throughout the lifetime of a cow, but susceptibility declines with age. Finally, transmission via food supplements only occurs once a cow dies; we assume only cattle that die of "natural" causes (i.e., those slaughtered) are processed into food supplements, those that obviously die of BSE are not fed back into the food

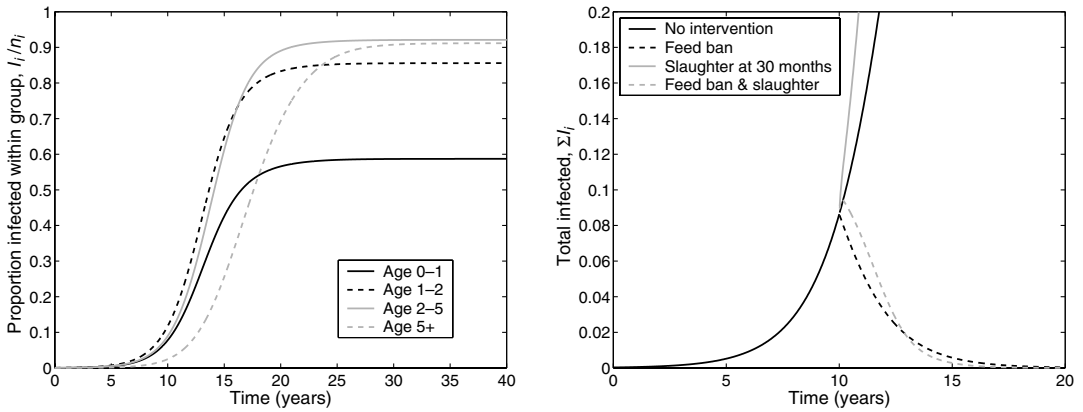


Figure 3.12. Results from a simplified age-structured model of BSE in cattle populations, showing the proportion infected (latent +infectious). The left-hand graph shows the distribution of infection across age-classes measured in years; the infection takes off sooner in young cattle, but reaches higher prevalence in older ones. The right-hand graph shows the effects of controls on feed and age-structured culling.

chain. This age-based heterogeneity in transmission and susceptibility makes the use of age-structured models vital in understanding this infection.

Figure 3.12 shows examples of the infection dynamics predicted by this comparatively simple model. These dynamics display the characteristics that are expected from any such structured model. There is an early short phase (not discernible on the graphs) that depends on the initial distribution of infecteds; during this phase prevalence in young cows increases most rapidly due to their greater susceptibility. In the second phase, the prevalence of infection within the age classes becomes "slaved", increasing at the same exponential rate. Finally, density-dependent forces begin to act and the infection levels settle to equilibrium values. Interestingly, although young cows dominate the early stages of the epidemic, at equilibrium older cattle are most likely to be infected because they will have had a longer exposure. Due to the very high transmission by infected feed and the long duration of the exposed period, the equilibrium prevalence of infecteds is very high. This clearly poses both a considerable hazard to human health, as well as devastating consequences for the farming industry. Control measures therefore needed to be implemented to wipe out the epidemic as quickly as possible, while not creating too large an economic burden. The costs associated with BSE control in Europe during 2001 were estimated at 7 billion euros. Therefore, determining a cost-effective and efficient policy is vital.

Three different control measures are investigated for this model (right-hand graph); these are the slaughter of all animals over 30 months of age ($\mu_{10} = \infty$), the ban of cattle meat and bonemeal in supplementary food ($\tau^F = 0$), and the adopted policy which is both of these measures. Surprisingly, although older animals are the most likely to be infected, a policy that involves just culling these animals actually leads to an increased epidemic. This is for two major reasons: (1) young animals are most susceptible to the infection, so the cull of older animals raises the average susceptibility of the population; and (2) by slaughtering animals much earlier there has been a significant shortening of the generation time for this infection, so infected animals are quickly converted into food supplements which speeds up the epidemic rise.

The major transmission route for BSE is via infected feed, therefore, it is intuitive that a ban on the inclusion of meat and bonemeal, effectively setting $\tau_F = 0$, should rapidly bring the epidemic under control. However, it is more surprising that a combination of early slaughter and a feed ban results in a faster eradication of the disease. This is again attributable to the fact that early slaughter of cattle reduces the infection generation time and thus the controls operate at a faster rate. This faster route to eradication is offset by a less rapid initial decrease, which is again due to the rise in average susceptibility caused by the slaughter of older less-susceptible animals.

Several other features were included in the predictive models used at the time (Donnelly et al. 1997; Ferguson et al. 1997b). The seasonal and demographic trends for cattle herds in the United Kingdom were incorporated. Also, most notably, infectious status was modeled in terms of time-since-infection (see Section 3.3 of this chapter), rather than the standard compartmental (SIR or $SEIR$) approach. This adds an extra dimension to the calculations, because each cow is now indexed by two variables (age and time-since-infection) rather than simply by age. Hence the full model is both age-and stage-structured. This complication allows a great deal more flexibility and realism compared to the simple model illustrated here. The distribution of incubation and latent periods can be modeled explicitly, and the transmission rate can be modified as the infection progresses within each animal. These more realistic features mean that the speed of infection reduction and timing of eradication can be predicted with far greater accuracy.

It is interesting and important to question what benefits can be obtained from such modeling approaches. The ban on meat and bonemeal in feed is an intuitive measure given the strengths of the various transmission routes. So at this level, models of any kind are clearly unnecessary. However, the real power of models comes when investigating issues such as the culling of older cattle, where feedback from nonlinear processes can produce counterintuitive results. Models are also useful in determining whether a control measure is likely to be sufficient; although it was clear that a feed ban was necessary, whether it would control the epidemic, in either the short or long term, could be addressed only with predictive models. Similarly, more or less drastic culls on cattle of different ages could be simulated to assess the benefits they produced. Implicit in all the calculations is a desire to minimize the risks to human health and eradicate infection to re-open export markets, while not placing unnecessary burdens on the farming industry. This type of modeling approach ideally needs to tie into economic models to produce a detailed cost-benefit analysis of various control options.

Finally, we raise the question of how complex the models need to be, what are the advantages of moving from the simple age-structured model developed here to the full (age-and stage-structured) model developed by Ferguson and coworkers (Ferguson et al. 1997b; Ferguson et al. 1999b). Although both models produce qualitatively similar results, the inclusion of time-since-infection allows far more accurate predictions of the future course of the epidemic—this degree of accuracy is required if economic factors are to be assessed. We can be convinced of the accuracy of our simple model only because it agrees with the more complex models that are believed to be a better representation of reality. In conclusion, given the economic and human health importance of the epidemic, it is vital that the predictive models used are believed to be the most accurate description available, while still maintaining sufficient transparency that their results can be interpreted. This trade-off between accuracy and transparency is one that permeates the entire field of mathematical modeling.

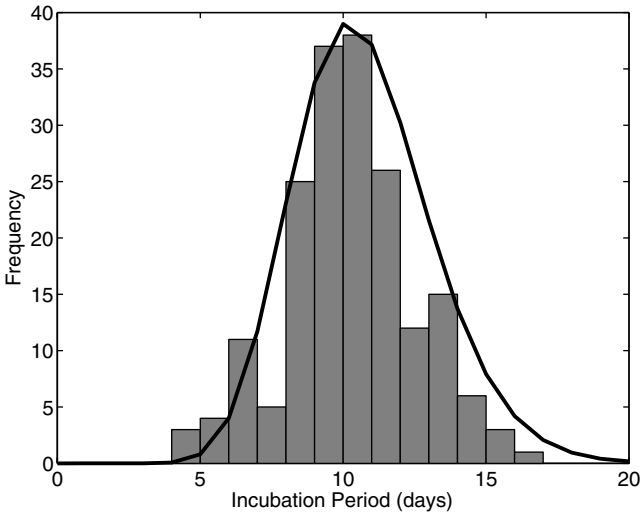


Figure 3.13. Frequency distribution of incubation periods for measles from Hope Simpson's (1952) classic household study. The data (gray bars) demonstrate that on average, the incubation period following measles infection is around 10 days, with some individual variation around this figure. The solid line represents the maximum likelihood fit to a gamma distribution, with the shape parameter (n) given by 20.

3.3. DEPENDENCE ON TIME SINCE INFECTION

In this section, we introduce another aspect of "risk structure" into the standard SIR model. Specifically, we are concerned with an individual's probability of remaining infectious as a function of time since infection, as well as the possibility that the risk of transmission may vary with the time since infection. To motivate this discussion, consider the standard SIR framework (Chapter 2) which assumes that the recovery rate is constant, independent of the time since infection. Thus, if we consider only the recovery and death processes of individuals infected at time $T = 0$, we have the classic equation:

$$\frac{dI}{dT} = -\gamma I - \mu I.$$

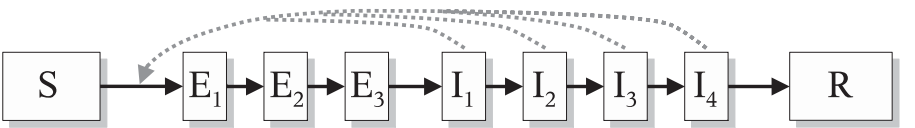
Upon integrating this equation, we see that the assumption of a constant recovery rate leads to the infectious period being exponentially distributed:

$$\mathbb{P}(\text{Infectious after time T}) = \exp(-(\gamma + \mu)T). \tag{3.10}$$

Biologically, this means that some individuals are infectious for only a very short period of time and contribute little to transmission, whereas others may be infectious for much longer. This assumption contradicts many empirical observations for a number of infectious diseases, especially those affecting the respiratory system. Classic work by Sartwell (1950) and Hope Simpson (1952) demonstrated that incubation period distributions for infections such as measles, chicken pox, polio, and the common cold typically show a strong central tendency. As demonstrated for measles in Figure 3.13, the incubation period has a pronounced mode, with some individual variation; the infectious period shows a similar type of distribution. As we show in Section 3.3.1, the precise assumptions made

concerning the distribution of latent and infectious periods have profound epidemiological implications, especially in the early stages of an epidemic when by chance an individual having a long infectious period will promote invasion (Keeling and Grenfell 1999).

3.3.1. *SEIR* and Multi-Compartment Models



To begin this analysis, we start by comparing and contrasting the dynamics of SIR and $SEIR$ models (Chapter 2). We can view the $SEIR$ model as a multi-compartment version of the standard SIR model, in which the infected class has been subdivided into an exposed and an infectious class. As described in Chapter 2, the $SEIR$ model is given by:

$$\begin{aligned} \frac{dS}{dt} &= \nu - \beta SI - \mu S, \\ \frac{dE}{dt} &= \beta SI - \sigma E - \mu E, \\ \frac{dI}{dt} &= \sigma E - \gamma I - \mu I. \end{aligned} \tag{3.11}$$

For this system of equations, the probability of being infectious at time T after becoming infected is given by:

$$\mathbb{P}(\text{infectious after time } T) = \sigma \exp(-\mu T)\frac{\exp(-\gamma T) - \exp(-\sigma T)}{\sigma - \gamma}. \tag{3.12}$$

This accounts for the exponential distribution of the exposed period, the exponential distribution of the infectious period, and the risk of natural mortality. For the SIR and $SEIR$ models, Figure 3.14 shows the two infectious probability distributions (equations (3.10) and (3.12)) and the epidemic curves generated assuming equal R_0 and equal infected (infectious + exposed) periods in both cases. This highlights an important issue; these two models have comparable equilibria (equilibrium traits in terms of the prevalence of infection and the level of susceptibles), however the $SEIR$ has a much slower growth rate. This slower growth is attributable to the delay that infected individuals wait before they can start transmitting caused by the exposed period. Chapter 2 provides a full description of this phenomenon.

We can extend this concept of compartmentalization still further, by subdividing the infected class:

$$\begin{aligned} \frac{dS}{dt} &= \nu - \beta S \sum_{i=m+1}^{n} I_i - \mu S, \\ \frac{dI_1}{dt} &= \beta S \sum_{i=m+1}^{n} I_i - \gamma n I_1 - \mu I_1, \\ \frac{dI_i}{dt} &= \gamma n I_{i-1} - \gamma n I_i - \mu I_i \qquad \forall i = 2, \ldots, n, \end{aligned} \tag{3.13}$$

This is online program 3.5

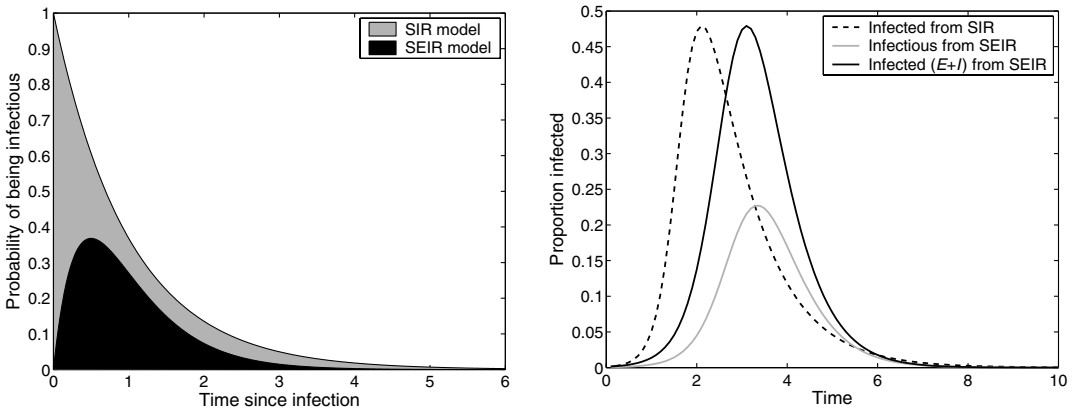


Figure 3.14. Comparison of SIR and $SEIR$ models. The left-hand graph shows the probability of being infectious at time T after being first infected, as given in equations (3.10) and (3.12). The right-hand graph shows the epidemic curve associated with these two distributions. Here we have assumed that the average time from infection to recovery is equal in both models $\frac{1}{\gamma_{SIR}} = \frac{1}{\gamma_{SEIR}} + \frac{1}{\sigma_{SEIR}}$, and $R_0 = 5$ for both models. (SIR model: $B = \mu = 5.5 \times 10^{-5}$, $\beta = 5$, $\gamma = 1$. $SEIR$ model: $B = \mu = 5.5 \times 10^{-5}$, $\beta \approx 10$, $\gamma = 2$, $\sigma = 2$.)

where it is assumed that only individuals in $I_{m+1}, \ldots, I_n$ are infectious, with the rest being in an exposed state. These equations are explicitly formulated to ensure that the average time between infection and recovery remains constant, so as the number of subclasses increases so does the rate at which individuals move between them. When $n = 1$, we return to the standard SIR model and when $n = 2, m = 1$, we obtain the $SEIR$ model with equal exposed and infectious periods. For general n (Figure 3.15), the exposed period, infectious period, and infected period (exposed plus infectious periods) are all gamma distributed (Lloyd 2001):

$$\mathbb{P}(\text{infected after time } T) = \int_T^\infty \frac{(\gamma n)^n}{(n-1)!} \tau^{n-1} \exp(-\gamma n \tau) d\tau.$$

This probability distribution means that the variance in the length of the infected period decreases to zero as n increases, such that in the limit when $n \to \infty$ all individuals spend exactly the same amount of time in the infected class.

Subdividing the infected period allows some control over the distribution of this period, scaling from exponential (when there is just one class) to constant (when there are many classes)

Associated with this change in the distribution of the infectious period at the individual level is a change in the epidemic profile at the population level (Figure 3.15). Despite having identical values of R_0 and the same average infectious period, changing the number of subdivisions (n) has a profound impact on the dynamics. As shown by Wearing et al. (2005), when n is large, the epidemic takes off more quickly and ends more abruptly. This difference is due to the different generation times in the two approaches. When $n = 1$, those individuals with long infectious periods produce the majority of the secondary cases and a high proportion of these are produced when the individual has been infectious

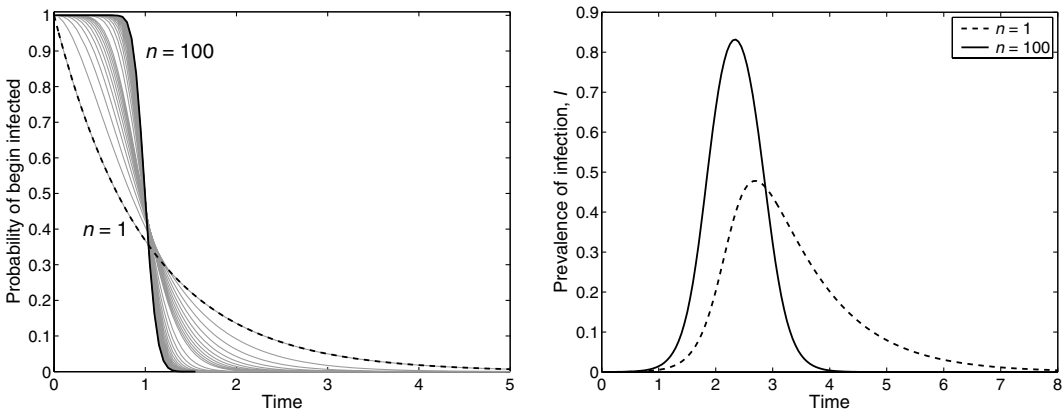


Figure 3.15. The change in the probability of infection and the epidemic profile as the number of subdivisions within the infected class increases from $n = 1$ to $n = 100$. Throughout we assume that all individuals in the infected class are infectious. Left-hand graph: for $n = 1$ the distribution of infected periods is exponential, whereas as n increases the infectious period becomes closer to a constant length ($\gamma = 1$). The right-hand graph shows the consequences of this change for the SIR-type epidemic without births or deaths. For the same basic reproductive ratio, $R_0 = 5$, and the same average infectious period, $\gamma = 1$, more subdivisions of the infectious period lead to a steeper rate of increase and an epidemic of shorter duration. Here we have ignored the exposed class by assuming that $m = 0$.

for a long time. Thus, when n is small, the rate of increase is much slower than when n is large (Kermack and McKendrick 1927; Metz 1978; Keeling and Grenfell 1997b). Conversely, during the latter stages of the epidemic when $n = 1$, the decrease in prevalence cannot be faster than the rate at which individuals recover, so $I(t) \sim \exp(-\gamma t)$. In contrast, when n is large, individuals infected at the peak of the epidemic always recover quickly, rapidly decreasing the prevalence of infection. Such changes to the dynamics may require a different parameterization (in terms of R_0 and average infectious period) for different n values to fit the same epidemic profile (Keeling and Grenfell 2002; Wearing et al. 2005), in particular a longer average infectious period may be required as n increases.

A greater subdividing of the infected population leads to more rapid growth rate and a shorter epidemic, necessitating different parameters for different models.

To demonstrate this point more fully, the growth rate of an outbreak in a totally susceptible population can be explored by determining the dominant eigenvalue of the (unstable) disease-free equilibrium for equations (3.13). This is the same basic methodology that was used to determine the growth rate in Box 3.1. As shown by Anderson and Palmer (1980), one can find an exact expression for the characteristic equation (see Chapter 2), where λ specifies the growth rate:

$$\lambda(\lambda+\gamma n)^n\left[\lambda(\lambda+\sigma m)^m - R_0\gamma(\sigma m)^m\left(1-\left(\frac{\lambda}{\gamma n}+1\right)^{-n}\right)\right]=0.$$

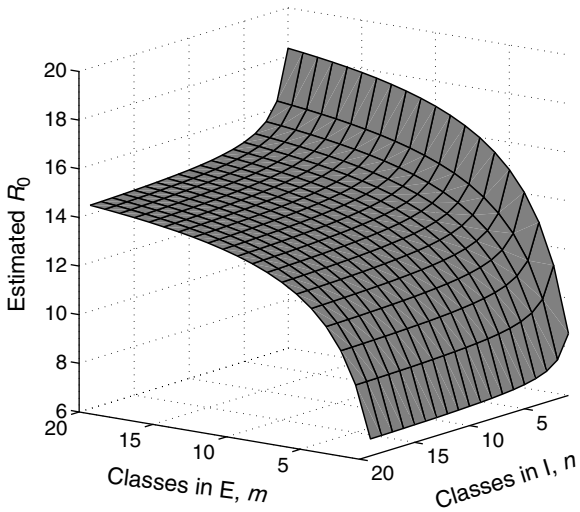


Figure 3.16. Estimates of R_0 from data on the initial growth rate, of an epidemic. The figure shows the effects of changing the distributions of the exposed and infectious periods on the value of R_0 estimated, with λ assumed to be 100 per year, $1/\sigma = 1/\gamma = 1$ week. The precise shape of each surface is independent of the exact value of λ.

During the early phase of an epidemic, the exponential growth rate, λ, satisfies the above equation.

Solution of this equation for a given set of epidemiological parameters (R_0, γ, σ, n, and m) is complex. We therefore look at the reverse problem—given we have data on an exponentially increasing epidemic, how does the subdivision of the exposed and infectious classes modify our estimation of R_0? Such an exercise reveals that for any empirically determined λ, the precise value of R_0 estimated depends on the fundamental assumptions made concerning the distributions of incubation and infectious periods. Specifically, the following equation determines the relationship between R_0 and the epidemic growth rate, λ:

$$R_0 = \frac{\lambda\left(\frac{\lambda}{\sigma m}+1\right)^m}{\gamma\left(1-\left(\frac{\lambda}{\gamma n}+1\right)^{-n}\right)}, \tag{3.14}$$

where m and n represent the number of subclasses in the exposed and infectious categories, respectively. The mean exposed and infectious periods are represented by $1/\sigma$ and $1/\gamma$, respectively, and are assumed to be known or estimated from independent data. This equation establishes how the value of R_0 estimated from data is influenced by a priori assumptions concerning the distributions of latent and infectious periods. This relationship is demonstrated in Figure 3.16 (Wearing et al. 2005). It reveals a very startling result: For a fixed epidemic growth rate, the value of R_0 estimated may vary from 6 to almost 20 as the number of subclasses changes.

In general, as the infectious period becomes more tightly distributed (increasing n), lower values of R_0 are estimated for any given growth rate, λ. On the other hand, as the variance in the exposed period is reduced (increasing m), *higher* values of R_0 are estimated. Indeed, we may use the relationship given by equation (3.14) to arrive at the following general principle: If we ignore the exposed period, then models with an exponentially distributed infectious period will *always* overestimate the basic reproductive ratio. When the latent period is included, however, this finding is *reversed* whenever the growth rate is large (Wearing et al. 2005). By closely examining equation (3.14), we note that the basic reproductive ratio estimated from a model without an exposed class ($\sigma \to \infty$) is always smaller than the estimate from the corresponding model when an exposed period is included ($1/\sigma > 0$). Therefore, when faced with a rapidly spreading infection, either entirely ignoring the exposed period or assuming exponential distributions will lead to an underestimate of R_0 and therefore will underestimate the level of global-control measures (such as mass vaccination) that will be needed to control the epidemic.

The distribution of infectious periods can have a profound impact on the epidemic behavior when the models are stochastic and individual-based (Keeling and Grenfell 1997a; Chapter 6). When $n = 1$, and the infectious period is exponentially distributed, transmission relies on the few individuals with longer than average periods—this increases the variability in transmission. In contrast, when n is large all individuals contribute equally to transmission, thus lowering the amount of variation. Such effects are most pronounced when the number of infectious individuals is low, such as at the start of an epidemic.

When births and deaths are included into this formalism, the equilibrium levels are independent of n (Hethcote and Tudor 1980):

$$S^* = \frac{1}{R_0} \qquad I^* = \frac{B}{\gamma + \mu} - \frac{\mu}{\beta},$$

and therefore the standard vaccination thresholds apply. However, the stability of this fixed point is reduced as n increases (Grossman 1980; Lloyd 2001); conceptually, this is because the fixed duration of infection adds a natural period of oscillation to the dynamics.

3.3.2. Models with Memory

The above models have achieved different infectious period distributions by modifying the number of compartmental subdivisions used to partition the infectious class; however, a similar but more flexible distribution of periods can be achieved by explicitly accounting for the time since infection. Hence, these models need to keep track of the history of infection—they are models with memory. Frequently we may wish to specify the probability distribution of both the exposed (latent) and infectious periods that have been derived from data, and are labeled $P_E(t)$ and $P_I(t)$ respectively. The probability that an individual infected at time 0 is still in the exposed class at time t is

given by:

$$\text{Prob(exposed at } t) = \mathbb{P}_E(t) = \int_t^\infty P_E(s)ds,$$

that is, the probability that the exposed period is longer than t. The probability it is infectious is calculated as:

$$\text{Prob(infectious at } t) = \mathbb{P}_I(t) = \int_0^t P_E(s) \int_{t-s}^\infty P_I(\tau)d\tau ds,$$

which is the probability that the individual is in the exposed class for less that t, but is still in the infectious class.

We now let $C(t)$ be the rate at which individuals are infected at time t, so that $\int_a^b C(t)dt$ is the fraction of new cases (relative to the population size) between times a and b. This variable C, together with the distributions P_E and P_I, allows us to calculate the number of exposed and infectious individuals at any time. Ignoring the exposed class for the moment, the basic SIR equations now become:

$$\begin{aligned} \frac{dS}{dt} &= \nu - C(t) - \mu S, \\ C(t) &= \beta S \int_0^\infty C(t-s)\mathbb{P}_I(s)ds, \end{aligned} \tag{3.15}$$

where the integral gives the proportion of individuals who are currently infectious, and so plays a similar role to I in the standard transmission term. Although one differential equation and one integral equation (3.15) may not look much more complex than the standard two differential equations needed to solve the SIR model (Chapter 2), the need to keep track of the historical values of C means that this technique is far more computationally challenging to implement.

A similar technique can be adapted for when we have data on the transmission rate as a function of the time since infection. For deterministic models, this produces a similar formulation of equations to those given above:

$$\begin{aligned} \frac{dS}{dt} &= \nu - C(t) - \mu S, \\ C(t) &= S \int_0^\infty \beta(s) \exp(-\mu s) C(t-s)ds, \end{aligned} \tag{3.16}$$

where the integral is over the time since infection, s, with $\beta(s)$ defining the transmission rate during the infectious period. For this formulation we do not need to explicitly model an exposed and infectious period; we simply allow the parameter $\beta(s)$ to capture to the known infection profile. This model also includes a term $\exp(-\mu s)$ that accounts for the death of individuals since infection.

This type of distributed infection model has been fit to several recent epidemics including SARS (Donnelly et al. 2003; Riley et al. 2003), foot-and-mouth (Ferguson et al. 2001a,b) and BSE (Ferguson et al. 1997b). These detailed analyses all confirm that, in general, the infectious period is rarely exponentially distributed, which has profound implications for the timing of control strategies. In particular, consider a control measure that isolates an individual some time after he or she has become infectious; although late control would still remove the tail of an exponential distribution, it may not do so for constant periods. Chapter 8 discusses this point in detail.

3.3.3. Application: SARS

Taken from the work of Donnelly et al. (2003) and Riley et al. (2003), we consider the dynamics and parameter estimates for the SARS epidemic in Hong Kong. The host-level lifecycle of the virus is as follows. After the initial infection, it takes about 6 days before individuals start to show clinical symptoms and become infectious. In the early part of the epidemic, in then took a further 4 days before the individual was hospitalized—although this delay dropped once more strict control measures were in place. Finally, individuals remained in hospitals for about 23 days if they recovered, or 36 days if the disease proved fatal. For modeling purposes, the population can therefore be partitioned into five distinct classes with respect to their infection or disease status: Susceptible, S; Exposed, E; Infectious, I; Hospitalized, H, and Recovered, R. The time spent in the exposed, infectious, and hospitalized classes can all be captured with a gamma distribution (Donnelly et al. 2003), and therefore are comparable with further subdividing each of these classes.

As shown above, if we wish to correctly estimate R_0, and therefore the level of control needed, it is important to correctly capture the distribution of the exposed and infectious periods. For the SARS epidemic, there was fortunately sufficiently detailed information on individual patients for the distributions to be estimated with some degree of confidence. When dealing with such public health issues as the control of SARS, it is generally important that models are as accurate as possible—reflecting the known pathogen and host behavior. Here we contrast the dynamics of standard models with distributed period models, assuming either equal R_0 or parameterized to fit to the same initial exponential growth. Although the qualitative dynamics are comparable, the quantitative differences could be very important to public health planning.

Given that the periods are all approximately gamma distributed, two equivalent formulations of the model are possible—one using memory and the other using multiple classes. Both models are given below to contrast the two mechanisms.

If $C(t)$ is the rate of infection at time t, then the proportions of the population in each of the classes are given by:

$$\begin{aligned} E(t) &= \int_0^\infty C(t-s) \int_s^\infty P_E(\tau) d\tau ds, \\ I(t) &= \int_0^\infty C(t-s) \int_0^s P_E(\tau) \int_{s-\tau}^\infty P_I(T) dT d\tau ds, \\ H(t) &= \int_0^\infty C(t-s) \int_0^s P_E(\tau) \int_0^{s-\tau} P_I(T) \int_{s-\tau-T}^\infty P_H(W) dW dT d\tau ds. \end{aligned} \tag{3.17}$$

So the proportion of the population that are hospitalized is all those who have been infected, have left the exposed class, have left the infectious class, but have not yet left hospitals. Throughout we will ignore births and deaths due to the rapid time-scale of the epidemic. The dynamics of the SARS infection are then governed by similar equations to before:

$$\begin{aligned} \frac{dS}{dt} &= -C(t), \\ C(t) &= S(\beta_I I + \beta_H H), \end{aligned} \tag{3.18}$$

where the transmission rate within hospitals (β_H) is allowed to be different from that in the community (β_I).

Alternatively, we may wish to formulate the model by subdividing the exposed, infectious, and hospitalized classes as illustrated in Section 3.1:

$$\begin{aligned}
\frac{dS}{dt} &= -S(\beta_I I + \beta_H H), \\
\frac{dE_1}{dt} &= S(\beta_I I + \beta_H H) - \sigma m E_1, \\
\frac{dE_i}{dt} &= \sigma m E_{i-1} - \sigma m E_i \qquad i = 2, \ldots, m, \\
\frac{dI_1}{dt} &= \sigma m E_m - \gamma_I n I_1, \\
\frac{dI_i}{dt} &= \gamma_I n I_{i-1} - \gamma_I n I_i \qquad i = 2, \ldots, n, \\
\frac{dH_1}{dt} &= \gamma_I n E_n - \gamma_H q H_1, \\
\frac{dH_i}{dt} &= \gamma_H q H_{i-1} - \gamma_H q H_i, \qquad i = 2, \ldots, q
\end{aligned} \tag{3.19}$$

where m, n, and q are the number of subdivisions for the exposed, infectious, and hospitalized classes, respectively.

Figure 3.17 compares the predicted dynamics of the SARS epidemic (without control) under two different modeling assumptions. The dashed lines give the results when the periods are exponentially distributed, which is equivalent to there being just one exposed, one infectious, and one hospitalized compartment. The solid lines correspond to using the distribution of periods taken from clinical observations; the periods were fit with a gamma distribution so the model is equivalent to having three exposed, three infectious, and ten hospitalized compartments. Following Riley et al. (2003), we assume that individuals in the hospitalized class transmit infection at a fifth of the rate of those in the infectious class ($\beta_H = \frac{1}{5}\beta_I$).

As seen in the generic model (Figure 3.15), the assumption of tighter distributions for the periods within the SARS model leads to a faster increase of the epidemic and a more pronounced epidemic peak for the same value of R_0. The differences between the results assuming either exponential and gamma-distributed periods are substantial, with significant public health implications associated with the discrepancies—even when both models are parameterized to match the same initial growth rate. The more accurate gamma-distributed periods consistently predict a larger peak number of hospitalized cases, which could place more strain on health services. In addition, the gamma-distributed periods are associated with a far more rapid decline in the epidemic, and a more rapid eradication of infection.

This work on SARS illustrates a crucial issue in modeling. Although the standard compartmental models, with their associated exponential distributions, are a useful tool for understanding the dynamics of infection, when very precise predictions are required

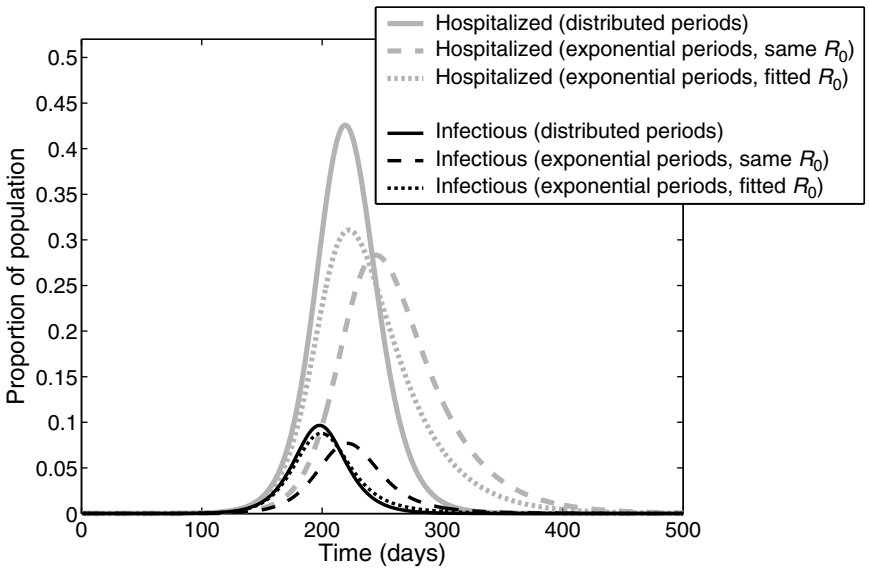


Figure 3.17. The predicted dynamics of SARS infection under two modeling assumptions: The dashed lines and dot-dashed lines correspond to a model with no memory and just one compartment for each of exposed, infectious, and hospitalized; the solid lines correspond to a model with a gamma-distributed period for the time spent in each class parameterized from the data. For the dashed line the model has equal R_0 to the distributed periods, for the dot-dashed line the model has the same initial growth rate as the distributed period model. (Exposed period distribution $1/\sigma = 6$ days, $m = 3$. Infectious period distribution $1/\gamma_I = 4$ days, $n = 3$. Hospitalized period distribution $1/\gamma_H = 30$ days, $q = 10$. Solid and dashed lines $R_0 = 2.8 \Rightarrow \beta_I = 0.28$ per day, $\beta_H = 0.056$ per day. Dot-dashed line $R_0 = 3.03$ with β_I and β_H similarly scaled.) Initialized with $S(0) = 1$, $E(0) = 10^{-6}$, $I(0) = H(0) = R(0) = 0$.

it becomes necessary to include much more of the available information; details often matter.

3.4. FUTURE DIRECTIONS

Although we have concentrated on the risk structure in sexually transmitted infections, almost all populations and associated infections have some form of risk structure. Thus even though infections such an influenza appear to spread randomly through the population, some individuals (super spreaders) have much more contact with the general public and therefore are more likely to catch and transmit infection. In the coming years, it is likely that more public-health motivated predictive modeling will incorporate this level of heterogeneity.

A second cause of risk structure may be due to genetics or the general health of the host, so that some individuals are much more susceptible to infection or more much infectious than others. This form of modeling has already been attempted, by considering a separate hospitalized, highly susceptible core group in addition to the general population (Austin and Anderson 1999b; Lloyd-Smith 2003). Such heterogeneities will be difficult to observe in practice, but again can play a determining role in the dynamics of infection and the

ability to control or eradicate the disease. Therefore, although determining this form of risk structure may be difficult, incorporating some generic form of heterogeneity may provide a better match between models and data.

An important future development may be the direct measurement of the general contact matrix ($\boldsymbol{\beta}$) for airborne infections as has been attempted for sexually transmission infections (Potterat et al. 2002). One method of ascertaining the network of potential transmission routes is through a diary-based approach (Edmunds et al. 1997), where contacts are recorded over a short period by volunteers. However, with the increased used of mobile phones and GPS technology, it will soon be possible to directly track the movement of a large proportion of the population and determine their position to within a few meters; this provides the opportunity to remotely gather vast quantities of information on the potential mixing patterns of the population. Dealing with this volume of data and using it to parameterize suitable models clearly has many potential public health applications, from estimating the likely spread of influenza to the real-time control of a deliberate or accidental release of an infectious agent.

Finally, as we have seen with the example of the SARS infection, the detailed natural history of infection at an individual level can be very important in determining the population-level dynamics. In particular, without capturing the correct distribution of exposed and infectious periods, it may be impossible to match to both the initial and long-term epidemic behavior. The inclusion of more patient-level information derived from detailed medical observations is ever more important as models are increasingly being expected to provide accurate quantitative predictions.

3.5. SUMMARY

➤ When structuring the population into different classes, the single transmission parameter in the unstructured models is replaced by a matrix of values.

➤ Because the transmission matrix generally has more terms than the structured data, simplifications are needed to overcome this deficit.

➤ The basic reproductive ratio, R_0, for structured populations is found using an eigenvalue approach. In general, this is greater than if the structure were ignored, and bounded by the basic reproductive ratios for each class.

➤ Assortative mixing is common, such that high-risk individuals mix more frequently with other high-risk individuals. Increased assortative mixing tends to increase R_0.

➤ The initial growth of a structure model depends on the initial conditions, not necessarily the basic reproductive ratio.

➤ It is no longer true that $S(\infty) = 1/R_0$; in many structured models that have high associativity, the equilibrium level of susceptibles is much higher and density-dependent saturation effects occur earlier. Hence, although the equilibrium prevalence is low, the infection may be difficult to eradicate because R_0 is still large.

➤ Targeting vaccination or other control measures is more efficient than random control. It is generally better to over-target rather than under-target.

➤ Although age is a continuous parameter, age-structured models usually group individuals into a limited number of classes—often representing school years.

➤ Age-structured models differ from the risk-structured models due to the regular progression of individuals into increasingly older age classes.

➤ At equilibrium, seroprevalence of SIR-type infections must increase with age. This allows the force of infection within an age class to be calculated.

➤ Subdividing the infected period allows some control over the distribution of this period, scaling from exponential (when there is just one class) to constant (when there are many classes).

➤ A greater subdividing of the infected population leads to more rapid growth rate and a shorter epidemic, necessitating different parameters for different models.