[Preprint of chapter 2 of *Epidemic Models: their Structure and Relation to Data*, ed. Denis Mollison, Cambridge UP 1995.]

The Structure of Epidemic Models

Denis Mollison

Summary

This paper reviews the basic components of epidemic models, and discusses some of the different ways of combining them, and relations between the resulting models. The fundamental aim is to help understanding of the relation between assumptions and the resulting dynamics: because without such understanding even a model which fits data perfectly can be of no scientific value.

Analysis of the structure of epidemic models is vital because of (1) the scarcity of good data and (2) the sensitive dependence of results on assumptions. In evaluating model dynamics, we need to look carefully at their dependence, not only on parameters, but also on the structure of the model: for instance, whether the population is treated as stochastic or deterministic, discrete or continuous, and how the timing and distribution of infectious contacts within the population is modelled. The practical target is to identify those parts of models that have most effect on dynamics: a few key parameters can drive a model (see *e.g.* Mollison 1984, 1985, Cairns, this volume).

The approach taken here is to analyse models in terms of their elements: expressing them in terms of simple key parameters that reflect individual life-histories, flows between states, and contact relationships. Basic definitions must be in terms of what one individual does to another; this implies that discrete models are basic, and that the stochastic aspect is usually important, if only in formulating and interpreting models.

Although more complex, stochastic models can have advantages in showing structure more clearly, as for instance in the technique of *coupling* which allows elegant comparisons of related models (see Ball, this volume).

Some results can be very sensitive to model assumptions, including hidden assumptions implicit in seemingly innocent parts of our model structure. Other results are so robust that they can be derived by 'pre-model' arguments, that is, by considering relations between basic components without choosing a specific model. As an illustration of this approach, the final section (Section 3) tries to express some basic results on epidemic models in their simplest and most general form, so as to analyse the range and limits of their applicability.

1 Introduction

1.1 Reasons for caring about structure

The aim of epidemic modelling is to understand and if possible control the spread of disease. To do this, it tries to relate disease dynamics at the population level to basic properties of the host and pathogen populations and of the infection process. Epidemic models thus express scientific hypotheses. Like other scientific models, if they are to be of value they need to be falsifiable; and if they are falsified, we need to know which part of the model has been disproved. There are two basic reasons why this is seldom easy.

The first is the nature of the data available for validating and testing models. The scope for experimental investigation of disease dynamics is severely limited, for both practical and ethical reasons. Data therefore are usually incomplete, and often complicated by many factors not of direct interest.

The second is that the dependence of modelling conclusions on assumptions is seldom straightforward. Some conclusions – for instance the existence of an epidemic threshold – are so robust that virtually any model will fit the data. Others may depend very sensitively on parameter values or, more insidiously, on assumptions implicit in the type of model chosen, for instance on the way in which it represents units of the population and contacts between them.

For both these reasons, it is essential to analyse the structure of epidemic models, and the relation between this structure and the resulting dynamics. To facilitate this, it is important to keep models clear and simple as far as possible. Thus we aim to find a small set of model components that determine the dynamics, and to describe these as far as possible in terms of simple parameters with clear ecological interpretations, such as the *basic reproductive ratio* (or *number*), R_0, and the *mean generation gap*, τ, of the disease. [R_0 is the mean number of infectious contacts made by an infective in a wholly susceptible population (see Dietz, this volume, Section 3); the generation gap is the time interval between an individuals's being infected and its infecting others.]

This approach should help us to see the similarities between many of the bewildering number of apparently different models in the literature; and thus allow comparison and synthesis of results on individual specific models into more general understanding. It can also clarify what data are needed to fit and to test a model.

1.2 Epidemic stages

It is generally helpful to distinguish three main epidemic stages: *Establishment, Spread* and *Persistence*. To these we might add *Arrival*, the question

of how infection reaches the population under consideration; however, for an existing disease, this can be considered, on a larger spatial scale, as part of the process of spread (see *e.g.* Cliff, this volume). Also, *Evolution* is required to explain the first arrival of any disease, and can play a key role in long term persistence, though this is an aspect in which diseases vary widely, from the evolutionary stability of smallpox or measles to the instability of myxomatosis (Fenner and Myers 1978) or influenza (Cliff *et al.* 1986); see also Hamilton and Howard (1994).

Given that an infection arrives in a population, the first question is that of *Establishment*, that is whether it has a chance to infect a sizeable proportion of the host population, rather than just a few individuals. In the establishment stage, it is common to ignore any overlap between infections by different individuals, so that growth is governed by a branching process, or by linear equations; in either case, the threshold condition for establishment to be possible is $R_0 > 1$ (see, *e.g.*, Diekmann *et al.*, Ball, Jacquez *et al.*, all in this volume; also Nåsell, this volume, regarding the definition of the threshold for stochastic models). However, where mixing is heterogeneous, and particularly in the case where individuals interact only with their spatial neighbours, the linear approximation can be poor, and the threshold value of R_0 may be appreciably greater than unity (Mollison 1991).

For cases where the infection has initial success, we then require to model its *Spread* through the population. This may be expected to depend both on heterogeneity between individuals: for instance, the spread of a sexually transmitted disease may be restricted largely within a 'core' group, at least initially; and on heterogeneity of mixing: for instance where contacts are spatially local we may expect spread in a regular wave-like manner at a steady velocity (see Metz and van den Bosch, this volume). Note that in the case of spatial waves the number of infectives grows only linearly with time (and the cumulative total quadratically), in contrast to the simpler cases where a linear or branching process model is a reasonable initial approximation, when numbers of infected accordingly grow exponentially. Intermediate rates of growth may be expected in intermediate situations, such as where the population is divided into a hierarchy of mixing groups, but this is an area where useful theoretical results have so far proved hard to develop.

Finally, the conditions for long term *Persistence* of an infection, whether at a steady level or as a sequence of outbreaks, may be expected to involve other factors. Bartlett (1957) introduced the idea of a *critical community size* for a given disease, below which an isolated population cannot sustain the disease long term. This critical community size, N_c, will depend primarily on the relation between the timescale of the infection itself and that of the regrowth of susceptible numbers (see §3.1). For a diseases such as measles with a mean generation gap of 10-14 days, N_c is around 250,000, which explains why such

diseases first became persistent in the human population in the Middle East four to five thousand years ago with the development of the earliest large cities there (McNeill 1976, Cliff *et al.* 1993). Measles persists similarly today, through reservoirs of infection in large cities, from which occasional epidemics are sparked off in rural and island communities (Cliff *et al.* 1993).

In populations of a more constant density, as in the case of many animals and plants, a disease may persist through wandering patches (see *e.g.* Mollison and Levin 1994) without any one population being continuously infected. The population size required for persistence depends on the spatial structure and connectivity of the population as well as on the parameters of the infection itself. Geographical connectivity is also important for human diseases. For instance the relatively one-dimensional connectivity of the Japanese population may at least partly explain why numerous epidemics of measles, 36 from the 11th to 19th centuries, failed to make it persist, even though Japan's population (30 million by 1868) was well over the Bartlett threshold (Cliff *et al.* 1993).

2 Building epidemic models

2.1 Components of models

Perhaps the most basic modelling components are those describing the time history of an individual infective. From the point of view of the individual, the course of the disease is best described in terms of the times at which it starts and ceases to feel ill; but from the epidemiological viewpoint, the essential element is the distribution over time and among the population of the infectious contacts made by the individual, relative to its own time and location of infection. This can be handled quite generally using a kernel describing the numbers of such contacts over time and location (see Metz and van den Bosch, this volume, Mollison 1991).

One convenient simplification is to assume constant transition rates from the incubating to the infectious and from the infectious to the removed or recovered state; the mathematical motivation for this assumption is to obtain a Markov process or differential equation model. An alternative simplification, the discrete time equivalent of this, is to assume a fixed incubation period and instantaneous infectious period, thus giving a constant generation gap.

Such simplifying assumptions will make little or no difference to some aspects of model behaviour: as we shall see below (Section 3), there are a number of basic formulae where only the mean of the generation gap or the infectious period is required. Other aspects, however, such as the stability of endemic conditions, may depend sensitively on the distribution of the generation gap (Mollison 1984, 1985).

Turning now to the contacts made by an individual, the simplest case is homogeneous mixing, where victims are selected from the whole population independently and with equal probabilities. Heterogeneous mixing can be through preference for some type of individual (*e.g.* 'high activity', or opposite sex), or from within some neighbourhood, whether defined socially (see *e.g.* Morris, Jacquez *et al.*, both in this volume) or by geography (see *e.g.* Durrett, Metz and van den Bosch, both in this volume, or Mollison and Levin 1994). The definition of geographic neighbours needs to take into account frequency of communication, not simply distance (see Cliff, this volume, Sattenspiel and Powell 1993).

In modelling, variability in the number of contacts made by an individual, and correlation between the locations of the victims, is often ignored. Where the numbers of infectives are large, this may often be justified; indeed, for linear models, and for nonlinear spatial models where the 'linear conjecture' applies (see next subsection), only the expected numbers matter, and so variability and correlation have no effect. However, they can in practice be very significant at the beginning of an outbreak. [Cliff *et al.* 1993 give many interesting examples relating to measles, for instance how it was introduced to Fiji and most effectively spread, along with the news of the islands' new colonial status, by their king in 1875.] And for stochastic models, not only do these details matter; taking them into account can actually be theoretically advantageous (see next section).

Careful consideration of the probabilities of contact with different possible victims is of particular importance where the population is divided into groups. Where there is wide variation in contact rates, perhaps both within and between groups as in the case of sexually transmitted diseases, the outcome of the epidemic may depend sensitively on the contact structure (see Morris, this volume). Where the population is divided into a large number of broadly similar groups, as for instance in the spread of airborne infections among households, it may be possible to develop hierarchical models, in which the groups are treated as individuals at the higher level of the model (Becker and Dietz 1994, Ball *et al.* 1994).

An alternative approach to modelling the infectious process is to look from the susceptible's rather than the infective's viewpoint, working in terms of the infectious pressure to which a susceptible is subject. This approach may be forced on us, for instance if the probability of infection depends only on there being some infectives and not on their number, as in the Greenwood model (see *e.g.* Bailey 1975); it also has some advantages of analytical convenience, for instance in the consideration of equilibrium conditions (see Anderson and May 1991). However, the loss of the idea of a link between infective and susceptible removes a major avenue for structural analysis of the model.

Most deterministic models use the approximation of treating populations

as continuous. This has considerable advantages of simplicity and generality, but we need to be aware of circumstances where this approximation is not good enough, notably where numbers oscillate, sometimes reaching low levels, or where mixing is heterogeneous with each individual interacting with only a small proportion of the population: an example combining both these factors is the differential equation model of Murray *et al.* (1986), which relies for its repeated waves of spread on minute fractions of an infectious individual – the 'atto-foxes' of Mollison (1991). Note that it may be the treating of the population as continuous, rather than determinism *per se*, that is the main problem here (Mollison 1991, Durrett and Levin 1994).

There has only been space here to scratch the surface of the wide subject of model choice. But I hope enough has been said to indicate the need to be aware of the process of model-building: while we must simplify, it is essential to understand the likely limits of our simplifications. The more easily we can interpret our model components – and compare them to available data – the easier it will be to understand how the structure of our model relates to reality, and its limitations. Thus, for example, it is traditional in many basic models to use a transmission rate parameter, β; but its units are 'time^{-1} population^{-1}', which make it difficult to interpret. Reinterpretation in terms of the more easily understood parameter R_0 greatly facilitates analysis of the various assumptions commonly made concerning β (see de Jong *et al.*, this volume, Mollison 1985).

2.2 Using the structure of models

The realistic detail of a stochastic model, specifying such things as the probability that one individual will infect another at a particular time and place, has long been recognised as a strength from the point of view of understanding and fitting models, but has generally been regarded as a grave handicap when it comes to analysis; even stochastic analyses have traditionally dealt whenever possible with massed variables such as the total number of infectives.

However, in recent years there has been an increasing recognition that the 'unnecessary' detail of a stochastic model framed in terms of individuals and their interactions can in many cases allow insights not possible from a 'higher level' stochastic or deterministic model.

A simple example of this is the use of the basic undirected random graph *G(n,p)* as an internal description of the Reed-Frost chain-binomial epidemic in a homogeneous population (Barbour and Mollison 1989). The 'unnecessary' detail here is that we identify each infection by one individual of another (represented by a link in the graph), rather than just dealing in the total numbers of susceptibles, infectives and removed cases. This is particularly simple

because of the independence we can allow in the random graph model between different infections made by an individual, and the symmetric relation between individuals; these allow us to represent the possibility that either a will infect b, or that b will infect a, by the same link in an undirected graph.

This idea of representing the relation between an infective and its victim by a link in a graph can be generalised to re-frame most of the common models for epidemics in a fixed population, though in general we must use a *directed* graph. We can then look at the question of who becomes infected in the course of the epidemic separately from the time structure: it only depends on the 'lists' of potential contacts of each individual, each such contact being represented by a link in the directed graph ('$a \to b$' meaning that b is on a's list of contacts). The strength of this approach is well illustrated by the beautiful theorem of Cox and Durrett (1988) on the existence of velocities for spatial epidemics with removal: although such a result clearly deals with the epidemic's development in time, much of the hardest part of the proof is accomplished through consideration of the graph structure of who may infect whom, without any explicit consideration of time.

The representation of contact structures by a random graph can be generalised to allow correlated links (see Lefèvre and Picard, this volume), and to compare two or more models (see next subsection).

Another stochastic technique, which exploits the structure of the model in a quite different but equally elegant way, is the use of martingales to estimate parameters (see Section 3 of Becker, this volume; note that his θ is our R_0).

There is typically less structure to exploit in the case of deterministic models. An interesting illustration of this is the proof of a monotonicity result by Kendall and Saunders (1983; see Ball, this volume) for the total number infected by two competing epidemics. This monotonicity seems 'intuitively obvious' for the deterministic model, but the proof requires exploitation of the structure of a corresponding stochastic model.

Nevertheless, monotonicity arguments, and similar comparisons of a model with variants, are often possible for deterministic models. An important, if only partly rigorous, example is the 'linear conjecture' for deterministic spatial epidemic models, which in turn leads to the possibility of analysing such models through a single structural element, the reproduction and dispersal kernel (see Metz and van den Bosch, this volume).

2.3 Relations between models

We here discuss briefly ways in which stochastic models can be related to each other, and to deterministic models. Though important for understanding, these could be considered rather theoretical aspects; we also discuss the crucial practical question of the relations between simple and complex models.

Stochastic models that include detailed 'internal descriptions', as described in the preceding subsection, can be used in a variety of ways to make precise comparisons between different models. The basic technique is that of *coupling* (see Ball, this volume), in which two or more different models are defined using the same probability space. Usually this representation is chosen so as to exploit the similarities between the different processes For instance, it may be possible to demonstrate a monotone relation between models where one process can be regarded as being the same as the other but for the addition of certain infections, or, more subtly, where the correlation between an individual's contacts is less in one model than in the other (see *e.g.* Kuulasmaa 1982).

Coupling can also be used to compare the outcome of the same process from different initial conditions, for instance showing that the 'contact process' (a spatial epidemic with recovery) is *additive* (see *e.g.* Mollison 1986).

Some aspects of relations between stochastic and deterministic models have already been touched on. Deterministic models are normally derived (explicitly or otherwise) by considering how the average numbers change in a stochastic model: because taking averages does not treat nonlinearities correctly, such a derivation will in general only give an exact relation for simple linear (branching process) models.

Nonlinearities are especially important where individuals interact only with a local group, whether defined socially or spatially. Thus, it is a defect of spatial continuous population models that they take little account of the spatial dimension, treating one and two dimensions very similarly (see Metz and van den Bosch, this volume), whereas in nonlinear discrete models, whether stochastic or deterministic, the very different nature of two dimensional space comes through (see, *e.g.*, Durrett, this volume, Fisch *et al.* 1991). Linear stochastic models also essentially ignore dimensionality, and it is mutual ignorance that allows them, in certain basic cases, to have an exact relation to well-known *nonlinear* differential equations (McKean 1975, Mollison and Daniels 1993).

It is possible to prove quite general results showing that 'as numbers get large' the behaviour of stochastic population processes tends to a deterministic limit, typically with diffusion process variability about that limit (Kurtz 1981). However, this could be considered to be the wrong way round, in that the use of deterministic models would be better justified if we could establish that a given stochastic process could be approximated as a limit of deterministic processes; therefore such results, though very useful, need to be treated with caution where the number with whom an individual interacts is small or where we wish to consider the process over a long time span (see §2.1 above).

Lastly, the relations between simple and complex models, though seldom mathematically elegant, are of great practical importance. From the applied

point of view it may be natural to include many parameters when setting up a model, yet its dynamics will often be almost exactly the same as that of a model with only a few basic components. Cairns (this volume) discusses the identification and estimation of such basic components, with application to modelling variable infectiousness during HIV infection.

Multi-parameter simulation models provide other examples where the complexity involved in an attempt at realism can hide crude (and unrealistic) assumptions about such basic components. For instance, the detailed spatial simulation model of Voigt *et al.* (1985) for fox rabies includes over thirty parameters; one effect of this is that important components may be handled too crudely; because of the way they discretize time, their value for the mean generation gap τ seems to be mistakenly taken as 2.5 months instead of its intended value of less than 1 month. More seriously, their conclusions as to the effect of varying population density, whether by culling or vaccination, all depend on their implicit assumption that R_0 is simply proportional to population density. This is a vital applied point: modelling should bring such crucial and debatable (see Mollison 1985, de Jong *et al.*, this volume) assumptions to ecologists' attention, not hide them.

3 Some simple general relations

In this final section, we turn to some almost 'model-free' results concerning epidemic models, relating such basic model components (see §2.1 above) as the basic reproductive ratio R_0 and mean times spent by an individual in various states.

There are a number of relations between basic population and disease parameters that can be expressed very simply. For example, we have the following three expressions for R_0:

$$R_0 = \beta N \tau_I = N/S = L/A, \tag{3.1}$$

where N is the population size (or in spatial models density), β the 'transmission parameter', τ_I the mean infectious period, S the equilibrium number of susceptibles, L the mean lifetime, and A the mean age of acquiring the disease. The first of these, $R_0 = \beta N \tau_I$, is the most general, being little more than a restatement of the definition of R_0 (see §3.1). The equivalence of the last two, $N/S = L/A$, relies on the disease being in endemic equilibrium (see §3.2); while their both being equal to R_0 relies on the assumption of homogeneous mixing (see §3.3).

Most of these simple relations can be found either as exact or approximate formulae in the literature. Dietz (1975) seems to have been the first to note that $R_0 \approx L/A$. More recently, many of these relations appear in Anderson

and May (1991)'s comprehensive survey of deterministic epidemic models. However, they are there derived for the most part as approximations, and this is one part of their otherwise impressive survey that could be improved in terms of elegance and generality.

We shall derive a number of such relations here in as general a way as possible, discussing the assumptions they rely on. For some of these results, it does not seem to have been recognised previously that they hold *exactly* in quite general circumstances. This is probably because there are alternative definitions for some of the parameters involved; these typically differ only by amounts too small to be of practical importance, but can render the simple relations unrecognisable.

Where there are such alternatives, the advantage of the simple relations, beyond their explanatory appeal, is that they will usually be of greater generality, or will at least indicate how far results can be generalised.

As well as deriving some of these simple exact results, I shall give examples to show how effectively a slight change of definition can disguise their simplicity.

3.1 Formulae for R_0

Let us first consider a simple and quite general formula, in that it does not require equilibrium conditions, concerning the *basic reproductive ratio* R_0. If we assume that infectives make contacts at a fixed rate βN, where N is the population total or density, throughout an infectious period of mean length τ_I, then their mean total number of contacts is exactly given by

$$R_0 = \beta N \tau_I. \tag{3.2}$$

This result can easily be modified to cover various different and more general assumptions. For instance, we could replace βN by a constant independent of N, so that R_0 is independent of N rather than proportional to it (Mollison 1985, de Jong *et al*, this volume); we could let β vary over time, in which case $\beta\tau_I$ should be replaced by $\int \beta(t)dt$, or over both time and space, as in Metz and van den Bosch (this volume)'s γ of their equation (2.3).

This simple formula, $R_0 = \beta N \tau_I$, is often hidden because it is common not to use the exact mean infectious period for an infective, τ_I, but instead what may seem a simpler parameter, τ_0, defined as the mean infectious period in the absence of other effects such as natural mortality.

As an example, consider the non-fatal disease model described in Anderson and May (1991, §4.4) in which individuals at birth possess immunity, which they lose at rate d, and in which infected individuals pass through successive latent and infectious stages with respective forces of removal σ and v; let us take the case of 'Type II' mortality, *i.e.* with an age-independent rate μ of

natural mortality. Then $\tau_0 = 1/v$; τ_I can be evaluated by multiplying the probability of an individual's surviving the latent period, $q_L = \sigma/(\sigma + \mu)$, by the mean time spent in the infectious state if it does so, $\tau_I' = 1/(v + \mu)$; thus $\tau_I = \sigma/[(\sigma + \mu)(v + \mu)]$. If we follow Anderson and May in excluding immune individuals from the effective population size, which is therefore $N' = Nd/(d + \mu)$, then '$R_0 = \beta N \tau_I$' gives exactly their equation (4.55):

$$R_0 = \frac{\beta N \sigma d}{(d + \mu)(\sigma + \mu)(v + \mu).} \tag{3.3}$$

The approximation, $R_0 \approx \beta N \tau_0$, can of course be deduced from this equation, but the advantage of the present approach is to clarify how the error in this approximation arises, through the component approximations: $q_L \approx 1$, $\tau_I' \approx \tau_0$, $N' \approx N$: each of which can be critically examined in a specific application. [For the case of age-dependent mortality, see §3.3 below.]

3.2 Equilibrium formulae: the microcosm principle

A number of simple equalities follow immediately from what I shall call the *microcosm principle*, which says that, for a quite general population process in equilibrium, the proportion of the population π_i in each state i is proportional to the mean time τ_i an individual spends in that state, and hence

$$\pi_i = \tau_i / L \tag{3.4}$$

[This result can be generalised to the case of a population growing at a steady rate r, essentially by including r as a discount rate – an individual's proportional contribution to the population diminishes exponentially, $\propto \exp(-rt)$. Thus in the righthand side of the equation τ_i is replaced by $\int p_i(t) e^{-rt} dt$, and L by the sum of such terms, $G = \int p(t) e^{-rt} dt$; where $p_i(t)$ denotes the probability of being in state i at age t, $p(t)$ the overall probability of being alive at that age.]

Now suppose we have a disease for which individuals are susceptible from birth, but immune once they have had the disease. Let A be the mean age of catching the disease, or of death for individuals who never get the disease. In this case the mean time spent susceptible, τ_S, is simply A, so from the microcosm principle we immediately have that

$$\pi_S \equiv S/N = A/L. \tag{3.5}$$

Note that this result makes no assumption about the epidemic process; it applies wherever individuals with a mean lifetime L start in a special state (here susceptible) and cannot return to that state once they have left it. The result is easily generalised to cases where individuals do not start susceptible.

For example, if individuals begin life with a period of mean M spent in an immune state we would have $\pi_S = (A - M)/L$. And it can again be adapted to the case of a growing population, along the lines mentioned above (see also Anderson and May 1991, §4.1 and §13.1.1).

Where an individual may visit a state i either less or more than once, the mean time spent in that state, τ_i, will not be the same as the mean time of a single sojourn, but often this can easily be allowed for. For instance, in the situation we are currently considering of the equilibrium state of a disease which can only be caught once, the mean time spent infectious will be $p_I\tau_I$, where p_I is the probability that an individual will become infectious at some time during their life. Hence

$$\pi_I = p_I\tau_I/L \tag{3.6}$$

For a typical human 'childhood disease' such as measles, L/τ_I is of the order of several thousand. This provides an elementary explanation of why the critical community size for such diseases is so large, of the order of 250,000 (Bartlett 1957). This size corresponds to an average number infected at any one time of around 100; in view of the seasonally oscillatory nature of measles it is not surprising that such a population size is necessary if the disease is not to die out through stochastic fluctuations.

As an illustration of the equation 3.6, consider a measles model of Grenfell *et al.* (this volume), which is the same as the example of §3.1 above, except that it omits immunity at birth. In our notation, their equation (3) becomes

$$\pi_I = \frac{\mu\sigma}{(\mu+\sigma)(\mu+v)} - \frac{\mu}{\beta N}. \tag{3.7}$$

To identify this with our equation 3.6, first use $R_0 = \beta N\tau_I$ and note that $\tau_I = \sigma/[(\sigma+\mu)(v+\mu)]$ as in §3.1. We can then deduce that $p_I = 1 - 1/R_0$; or this can be derived independently, by noting first that $1 - p_I =$ Prob.{Susceptible at death}, which because mortality does not depend on age is simply = Prob.{Susceptible}, $= 1/R_0$ from equation 3.9 below.

3.3 Equilibrium under homogeneous mixing

We could allow the rate of contacts to depend on the number of infectives, which would imply that an infective's mean total number of contacts also does so, in which case we have to replace R_0 by $R(I) = \beta(I)N\tau_I$; the usual definition of R_0 identifies it with $R(1)$. [To be pedantic, '$R(1)$' may not be quite right here, as we should allow for the possibility that the infectious period of the first infective may overlap with those of some of its victims, but it will do to make the point.]

If we now assume that mixing is homogeneous as regards susceptibles, so that the probability that a potentially infectious contact is with a susceptible is simply π_S, then the mean number of successful contacts per infective is $R(I)\pi_S$. But if the process is in equilibrium this number must equal 1, which immediately tells us that in that case

$$R(I_\star) = 1/\pi_S \tag{3.8}$$

where $I_\star$ is the equilibrium number of infectives.

If we also assume that the mean total number of contacts per infective is independent of the number of infectives, then $R(I) = R_0$, and so we have

$$R_0 = 1/\pi_S (= N/S) \tag{3.9}$$

Note that Anderson and May (1991, Equation 4.13) are wrong in claiming that this equation relies only on 'weak homogeneous mixing'; it is only the previous, rather less useful, equation ($R(I_\star) = 1/\pi_S$) that holds in that case. The general issue they point to, of how the number of contacts depends on the numbers of susceptibles and infectives, is nevertheless a crucial one – already raised in the contrasting Reed-Frost and Greenwood models of the 1930s. The answer is likely to depend on the mode of transmission, for instance physical contact as opposed to aerosol, and on the heterogeneous social structure of the population (see §2.1 above).

A more hair-splitting reservation concerning equation 3.9 arises if mortality is age-dependent, because that induces (in practice usually negligible) variation with age in the mean infectious period, and thus (through equation 3.2) in R_0. [In the simple homogeneous mixing case, the age distribution of infectives conditional on survival is exponential in equilibrium, whereas during initial spread it is uniform.] In specific cases, for instance where everyone lives to exactly age L (Anderson and May's 'Type I mortality'), it is possible to write down the probability that an infective will die of natural mortality,

$$p_e = \frac{\lambda(\mathrm{e}^{-\lambda L} - \mathrm{e}^{-\upsilon L})}{(\upsilon - \lambda)(1 - \mathrm{e}^{-\lambda L})}, \tag{3.10}$$

hence calculate τ_I, and thus derive exact expressions for $R(I_\star)$ and π_I (for the latter, see Anderson and May 1991, equation (4.41)). However for endemic measles p_e is of at most of order 10^{-3}, and p_I not much larger (for 'Type I mortality'), so that the errors in the approximate equations

$$R(I_\star) \approx R_0 \approx \beta N \tau_0 \qquad \text{and} \qquad \pi_I \approx \tau_0/L \tag{3.11}$$

are minute compared with the error in estimating (for instance) τ_0 or τ_I.

Another simple equilibrium formula relates the force of infection λ to parameters already introduced. Equating inward and outward flows of attempted infections gives $p_I R_0 = \lambda L$.

A rather less neat result, but of considerable interest, concerns the period of oscillations about equilibrium. For both continuous and discrete time simple endemic models, this period is approximately $2\pi\sqrt{\tau A/p_I}$; thus, as Anderson and May (1991) nicely remark, it is proportional to the geometric mean of the two basic time scales of the process: the typically short time scale of the infection, as represented by the mean generation gap τ, and the longer time scale for replenishment of susceptibles. Similar results for simple fatal disease models were found by Mollison (1985), who also showed that the stability of oscillations was sensitive to the difference between continuous and discrete time models (with less stability for the fixed delay feedback of the discrete time model, as one might expect).

3.4 Discussion

We have considered in this section a number of simple formulae that can be found over and again in models in the literature, usually disguised to a lesser or greater extent. Often the conclusions drawn from those models depend essentially on the validity of the relationships described by these simple formulae, or the way in which they are used. For instance, it is common to take β constant, which through equation 3.2 embodies the questionable assumption that $R_0 \propto N$ (see de Jong *et al.*, this volume, Mollison 1985).

The estimation of R_0 presents a key difficulty in epidemic modelling, and several of the simple formulae are relevant to this, especially $R_0 = 1/\pi_S$ or $R_0 = L/A$. If the aim is to estimate the proportion we need to vaccinate, p_V say, where homogeneous mixing theory suggests we need $p_V \geq 1 - 1/R_0$, we can in fact short-cut the argument, omitting the estimation of R_0 itself: it would seem that $p_V > 1 - \pi_S$ should suffice, not only in the homogeneous mixing case. However, this deduction relies on treating the susceptibles remaining after vaccination as being similarly distributed within the population to the susceptibles in the endemic state when there is no vaccination, and in a heterogeneous situation (whether age or space dependent) this assumption would need careful examination.

Some of the simple formulae are known to require major correction, or to be simply invalid, under certain types of heterogeneous mixing. For instance, for simple spatial endemic models for a fatal disease, Mollison and Kuulasmaa (1985) found that π_S and π_I were respectively much larger and much smaller than the values given by the homogeneous mixing model. [Though the formulae can be adjusted to explain this; for instance the increase in π_S is inversely proportional to the reduction in the frequency of {infective, susceptible} pairs relative to the homogeneous situation.]

Further, the homogeneous model's oscillations, although they carry over to differential equation models (Murray *et al.* 1986), do not occur in the stochas-

tic models, a result confirmed by Durrett and Levin (1994)'s comparison of different types of model for spatial competition. The stochastic spatial models instead have patterns of wandering patches; by a nice irony, in the case of fox rabies these have a 'turnover period' numerically similar to the homogeneous mixing model's period of oscillations (Mollison 1986). Endemic patterns in human diseases are more complex, but here too there is some evidence that the spatial structure deters chaotic and oscillatory behaviour (Ellner *et al.*, Grenfell *et al.*, both in this volume).

Although they have such limitations, the simple formulae do have the cardinal virtue of clarity. Consideration of the basic relationships that these formulae describe can clarify the assumptions inherent in a model; in contrast, complex formulae give a spurious appearance of precision that may distract our attention from structural faults in the model. It is only if any shortcomings are recognised that we can correct for them, or at least make some allowance for the error involved: an approximate answer to the right question is better than a precise answer to the wrong question.

References

Anderson, RM, and RM May (1991) *Infectious Diseases of Humans: Dynamics and Control*, OUP, Oxford.

Bailey, NTJ (1975) *The Mathematical Theory of Infectious Diseases and its Applications*, Griffin, London.

Ball, Frank (1994) 'Coupling methods in epidemic theory' (this volume)

Ball, FG, D Mollison and G Scalia-Tomba (1994) 'Epidemics in populations divided into groups or households' (in preparation)

Barbour, AD and D Mollison (1989) 'Epidemics and random graphs', in *Stochastic Processes in Epidemic Theory* (eds J P Gabriel, C Lefevre and Ph Picard), *Lec Notes in Biomathematics* **86**, pp. 86-89.

Bartlett, MS (1957) 'Measles periodicity and community size', *J Roy Statist Soc* **A 120**, 48-70.

Becker, Niels (1994) 'Statistical challenges of epidemic data' (this volume)

Becker, NG, and K Dietz (1994) 'The effect of the household distribution on transmission and control of highly infectious diseases' (in preparation)

Cairns, Andrew (1994) 'Primary components of epidemic models' (this volume)

Cliff, Andrew (1994) 'Incorporating spatial components into models of epidemic spread' (this volume)

Cliff, AD, P Haggett and JK Ord (1986) *Spatial Aspects of Influenza Epidemics*, Blackwell, Oxford.

Cliff, AD, P Haggett and M Smallman-Raynor (1993) *Measles: an Historical Geography*, Pion, London.

Cox, JT, and R Durrett (1988) 'Limit theorems for the spread of epidemics and forest fires', *Stoch Procs Applics* **30**, 171-191.

de Jong, M, O Diekmann and JAP Heesterbeek (1994) 'How does transmission of infection depend on population size?' (this volume)

Diekmann, O, JAJ Metz and JAP Heesterbeek (1994) 'The legacy of Kermack and McKendrick' (this volume)

Dietz, K (1975) 'Transmission and control of arbovirus diseases', in *Epidemiology* (eds. D Ludwig and KL Cooke), SIAM, Philadelphia, pp. 104-121.

Dietz, K (1994) 'Some problems in the theory of infectious disease transmission and control' (this volume)

Durrett, Richard (1994) 'Spatial epidemic models' (this volume)

Durrett, R, and SA Levin (1994) 'The importance of being discrete (and spatial)', *Theor Pop Biol* (to appear)

Ellner, S, R Gallant and J Theiler (1994) 'Detecting nonlinearity and chaos in epidemic data' (this volume)

Fenner, F, and K Myers (1978) 'Myxoma virus and myxomatosis in retrospect: the first quarter century of a new disease', in *Viruses and the Environment* (eds. E Kurstak and K Maramorosch), Academic Press, New York and London, pp. 539-570.

Fisch, Robert, Janko Gravner and David Griffeath (1991) 'Threshold-range scaling of excitable cellular automata', *Statistics and Computing* **1**, 23-39.

Grenfell, BT, B Bolker and A Kleczkowski (1994) 'Seasonality, demography and the dynamics of measles in developed countries' (this volume)

Hamilton, WD and JC Howard (eds) (1994) *Infection, Polymorphism and Evolution*, *Phil Trans R Soc Lond* **B** (to appear)

Jacquez, J, C Simon and J Koopman (1994) 'Core groups and R_0s for subgroups in heterogeneous SIS and SI models' (this volume)

Kendall, WS and IW Saunders (1983) 'Epidemics in competition II: the general epidemic', *JR Statist Soc* **B 45**, 238-244.

Kurtz, TG (1981) *Approximation of population processes*, SIAM, Philadelphia.

Kuulasmaa, Kari (1982) 'The spatial general epidemic and locally dependent random graphs'*J Appl Prob* **19**, 745-758.

Lefèvre, C and Ph Picard 'Collective Reed-Frost processes: a general modelling approach to the final outcome of SIR epidemics' (this volume)

Longini, I, E Halloran and M Haber (1994) 'Some current trends in estimating vaccine efficacy' (this volume)

McKean, HP (1975) 'Application of Brownian Motion to the equation of Kolmogorov-Petrovskii-Piscunov', *Comm Pure Appl Maths* **28**, 323-331.

McNeill, William H (1976) *Plagues and Peoples*, Doubleday.

Metz, JAJ, and F van den Bosch (1994) 'Velocities of epidemic spread' (this volume)

Mollison, Denis (1984) 'Simplifying simple epidemic models', *Nature* **310**, 224-225.

Mollison, Denis (1985) 'Sensitivity analysis of simple endemic models', in *Population Dynamics of Rabies in Wildlife* (ed. PH Bacon), Academic Press, London, pp. 223-234.

Mollison, Denis (1986) 'Modelling biological invasions: chance, explanation, prediction', *Phil Trans R Soc Lond* **B 314**, 675-693.

Mollison, Denis (1991) 'Dependence of epidemic and population velocities on basic parameters', *Math Biosc.* **107**, 255-287.

Mollison, D and HE Daniels (1993) 'The simple deterministic epidemic unmasked', *Math Biosciences* **117**, 147-153.

Mollison, D and K Kuulasmaa (1985) 'Spatial epidemic models: theory and simulations', in *Population Dynamics of Rabies in Wildlife* (ed. PH Bacon), Academic Press, London, pp. 291-309.

Mollison, D, and SA Levin (1994) 'Spatial dynamics of parasitism', in *Ecology of Infectious Diseases in Natural Populations* (eds. A Dobson and BT Grenfell), CUP, Cambridge, pp. ??.

Morris, Martina (1994) 'Data driven network models for the spread of disease' (this volume)

Murray, JD, EA Stanley and DL Brown (1986) 'On the spatial spread of rabies among foxes', *Proc R Soc Lond* **B 229**, 111-150.

Nåsell, Ingemar (1994) 'The threshold concept in stochastic epidemic and endemic models' (this volume)

Sattenspiel, L and C Powell (1993) 'Geographic spread of measles on the island of Dominica, West Indies', *Human Biology* **65**, 107-129.

Voigt, DR, RR Tinline and LH Broekhoven (1985) 'A spatial simulation model for rabies control', in *Population Dynamics of Rabies in Wildlife* (ed. PH Bacon), Academic Press, London, pp. 311-349.