The legacy of Kermack and McKendrick

O. Diekmann, J.A.P. Heesterbeek, J.A.J. Metz

Department of Analysis, Algebra and Geometry

Report AM-R9304 July 1993

CWI is the National Research Institute for Mathematics and Computer Science. CWI is part of the Stichting Mathematisch Centrum (SMC), the Dutch foundation for promotion of mathematics and computer science and their applications.
SMC is sponsored by the Netherlands Organization for Scientific Research (NWO). CWI is a member of ERCIM, the European Research Consortium for Informatics and Mathematics.

P.O. Box 94079, 1090 GB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

The Legacy of Kermack and McKendrick

O.Diekmann[1,2], J.A.P. Heesterbeek[1,3] , J.A.J. Metz[2,1]

1. CWI
P.O. Box 4079, 1009 AB Amsterdam
The Netherlands

2. Institute of Theoretical Biology
P.O. Box 9516, 2300 RA Leiden, The Netherlands

3. Isaac Newton Institute for Mathematical Sciences
20 Clarkson Road, Cambridge, CB3 0EH, U.K.

Abstract

Starting from the pioneering work by Kermack and McKendrick of 1927, we review some issues of recent interest in deterministic epidemic modelling such as: classification of dynamics, the influence of heterogeneity and submodels for the contact process.

1991 Mathematics Subject Classification: 92D30.
Keywords & Phrases: Epidemic models, endemicity, basic reproduction ratio, contact process, heterogeneity.
Note: This paper will appear in: *Epidemic Models, their Structure and Relation to Data* D. Mollison (ed.), Cambridge University Press, 1993.

Report AM-R9304
ISSN 0924-2953
CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

1. Introduction

In 1927, Kermack and McKendrick published a contribution to the mathematical theory of epidemics in which they considered the following situation:

1. a single infection triggers an autonomous process within the host (i.e., they look at 'microparasites' and not at 'macroparasites');
2. the disease results in either complete immunity or death;
3. contacts are according to the law of mass-action;
4. all individuals are equally susceptible;
5. the population is *closed*, i.e. at the time-scale of disease transmission the inflow of new susceptibles into the population is negligible;
6. the population size is large enough to warrant a deterministic description.

The aim of this paper is to first very briefly review the results of Kermack and McKendrick and then to give an overview of some new issues which have received a lot of attention more recently, to wit

- more complicated dynamics (relaxing 5)
- heterogeneity (relaxing 4)
- submodels, in particular for the contact process (relaxing 3)

while staying within the deterministic setting (i.e. retaining 6, although we will make one or two remarks on the difficulty of defining the 'border' between those situations where a deterministic model makes sense and those where it does not). We will not consider partial and/or temporary immunity (Anderson and May (1991)), nor say anything about models for macroparasites (see Kretzschmar and Adler (1993) for a comparison of different approaches and Roberts, Smith and Grenfell (1993) for a recent review; see also Dobson, this volume). For a general introduction to modelling infectious diseases, especially those incorporating vertical transmission, consult Busenberg and Cooke (1993).

Admittedly, the overview will be strongly biassed by personal interest and knowledge. At the positive side of this we want to acknowledge the pleasant cooperation over several years with M.C.M. de Jong and M. Kretzschmar which has contributed a lot to any insight we might have.

2. Invasion, time course and final size according to Kermack and McKendrick

The assumptions listed in section 1 directly lead to the integral equation

$$\dot{S}(t) = S(t) \int_0^\infty \bar{A}(\tau)\dot{S}(t-\tau)d\tau, \tag{2.1}$$

where $S(t)$ denotes the (spatial) density of susceptibles (i.e., number of individuals per unit area) at time t and where, by definition,

$$\bar{A}(\tau) = \text{expected infectivity of an individual that became infected } \tau \text{ units of time ago.} \tag{2.2}$$

In order to understand equation (2.1) one has just to realise that, by the closedness of the population, $-\dot{S}(t)$ is precisely the incidence at time t, so $-\dot{S}(t-\tau)$ is the number of individuals arising per unit of time who at time t have been infected for τ time-units.

At this point we make a short digression. Even though it has been emphasised repeatedly that Kermack and McKendrick deal with the case of a general time-kernel $\bar{A}(\tau)$ for infectivity, most people keep referring to the system of ordinary differential equations

$$\begin{aligned} \dot{S}(t) &= -\beta S(t)I(t) \\ \dot{I}(t) &= \beta S(t)I(t) - \gamma I(t) \end{aligned} \tag{2.3}$$

as *the* Kermack and McKendrick model. This should stop!* Note that (2.3) is derived from the special case

$$\bar{A}(\tau) = \beta e^{-\gamma\tau} \tag{2.4}$$

by defining $I(t) := -\frac{1}{\beta}\int_0^\infty \bar{A}(\tau)\dot{S}(t-\tau)d\tau = -\frac{1}{\beta}\int_{-\infty}^t \bar{A}(t-\tau)\dot{S}(\tau)d\tau$ and differentiating. In conclusion of our digression, we pass on the observation of Klaus Dietz that already in 1917, in a neglected paper, Ross and Hudson (1917) discussed a model where infectivity can be a general function of the time elapsed since infection took place (and, incidentally, also discuss (2.3) with differential mortality). Although Kermack & McKendrick were the first to give a detailed analysis of such a model, they were clearly influenced by Ross and Hudson (see Aitchison & Watson (1988)). The spirit of the Kermack and McKendrick and Ross-Hudson papers is very much one of generality. Their aim is to analyse large classes of models in one go. In recent years, a trend in the opposite direction can be discerned, leading to page after page on the analysis of the umpteenth variant of model $4A\beta*$, without any more extensive insights or methods coming to the fore (see Hethcote (1993) for a critique of this phenomenon).

The definition of $\bar{A}(\tau)$ as an *expected* infectivity emphasises that some heterogeneity, viz. variability in infectivity, is already incorporated. Indeed, one should always realise that even deterministic models are built from stochastic considerations at the individual level. Sometimes it is clear how to take averages and sometimes, as we will see below, it is less clear or even not clear at all.

To illustrate this remark we recall the usual interpretation of (2.3): individuals are infected for an exponentially distributed period of time (with exponent γ) and have a constant infectivity β. Hence

$$\bar{A}(\tau) = \int_0^\infty A(\tau,\xi)f(\xi)d\xi, \tag{2.5}$$

where $f(\xi) = \gamma e^{-\gamma\xi}$ and $A(\tau,\xi) = \beta$ when $\tau \leq \xi$ and $A(\tau,\xi) = 0$ when $\tau > \xi$. So here the 'type' ξ of individuals refers to the length of their period of infectivity. In this manner any compartmental model (indicated as some finite sequence of characters from $\{S,E,I,R\}$) for a closed population may be reduced to (2.1) with an appropriate kernel $\bar{A}$, (see Metz (1978)).

Kermack and McKendrick derive an invasion criterion and the equation for the final size of the epidemic in the general setting of (2.1) and they obtain more detailed information about the time course of the epidemic for the special case described by (2.3).

The *invasion* criterion is based on the linearisation in which at the right hand side of (2.1), $S(t)$ is replaced by S_0, the density of the population at the start of the epidemic with everyone susceptible. The linearised equation has a solution $\dot{S}(t) = ce^{rt}$ with $r > 0$ if and only if $R_0 > 1$ where, by definition,

$$R_0 = S_0\int_0^\infty \bar{A}(\tau)d\tau. \tag{2.6}$$

Hence, R_0 can be interpreted as the expected number of secondary cases produced by one typical primary case and it describes the growth of the epidemic in the initial phase on a generation basis. In the same vein r, the real root of the characteristic equation

$$1 = S_0\int_0^\infty \bar{A}(\tau)e^{-r\tau}d\tau \tag{2.7}$$

describes the 'real-time' growth in the initial phase. The positivity of $\bar{A}$ guarantees the equivalence

$$R_0 > 1 \Leftrightarrow r > 0 \tag{2.8}$$

* The often heard 'excuse' for only citing but not reading the Kermack and McKendrick papers, i.e. that it is difficult to obtain them, can no longer be upheld due to the recent reprinting in ***Bulletin Math. Biol.*** **53**: 33-55, 57- 87 and 89-118, (1991).

but one should note that, if one compares different kernels, the ordering of the R_0-values does not necessarily correspond to the ordering of the r-values (early or late 'reproduction' does not matter for R_0 but it does matter for r). Anyhow, the invasion criterion clearly is $R_0 > 1$.

The invasion criterion is actually very often used negatively, viz. as an eradication/elimination criterion*: whether or not a certain control measure (e.g. a vaccination programme) is strong enough to eradicate/eliminate the disease is determined by whether or not it is capable of reducing the value of the net reproduction ratio R to below 1. Incidentally, we remark that explicit expressions for R may be helpful in suggesting which component(s) of the transmission cycle are most sensitive to control measures. See Dietz (1993) for a more elaborate discussion.

Dividing (2.1) by $S(t)$ and integrating we obtain

$$\ln \frac{S(t)}{S_0} = \int_0^\infty \bar{A}(\tau)\{S(t-\tau) - S_0\}d\tau \tag{2.9}$$

and subsequently a limit argument yields the *final size* equation

$$\ln \frac{S(\infty)}{S_0} = R_0\left(\frac{S(\infty)}{S_0} - 1\right) \tag{2.10}$$

which can easily be analysed graphically (here we are concentrating on a negligibly small inoculum; a more elaborate presentation of the arguments involved can be found in Metz and Diekmann (1986), section IV.4.1). The outcome is most conveniently presented pictorially as in figure 1.

Concerning the *time course* of the epidemic when (2.3) is used, the density of infecteds reaches its peak value when $S = \frac{\gamma}{\beta}$. This can be seen directly from the second equation of (2.3).

3. The problem of endemicity

In 1932 and 1933, Kermack and McKendrick (1932,1933) addressed the problems that arise when one relaxes assumption 5 that the population is closed, and assumption 2 of permanent immunity. Here, possibly by the nature of the problem and by the lack of computers, they did not arive at such clear conclusions.

What type of model is appropriate depends on the *time-scale*

1. of disease transmission;
2. of population turnover (demography);
3. which interests us.

First of all one has to decide whether one takes the rate at which newborns are added to the population

a. constant

or

b. (linearly) related to population size.

In case *a*, demography influences the disease dynamics but not really vice versa, whereas in case *b* one can study the regulation problem: can an infectious agent control the size of its host population? Anderson and May called attention to this regulation problem in influential papers in 1978 and 1979.

Before briefly listing the typical results for cases *a* and *b*, we want to emphasise a problem which is, in our opinion, not understood at all: when can we expect *repeated outbreaks* (i.e. epidemics separated by

---

* One can argue whether or not a new symbol should be introduced in the case of a population that is not wholly susceptible because some control measures have been applied. One could denote the corresponding net reproduction ratio by R. However, from a mathematical point of view, R_0 and R are reproduction ratios calculated in precisely the same manner, the only difference being that the virgin and the controlled populations differ in density of susceptibles at the moment of invasion. Under the assumptions in this section $R = \frac{S}{S_0}R_0$, where S/S_0 is the fraction of the population that is actually susceptible.

disease free periods, e.g. measles in Iceland) and when an endemic situation? Here we touch upon the difficulty that deterministic dynamics may lead to a situation where the deterministic approximation ceases to be meaningful (the 'atto-fox' of D. Mollison (1991)). If we draw the somewhat symbolic picture in figure 2, we can distinguish three types of stochastic aspects:

1. At the introduction of the infectious agent the possibility of a minor outbreak exists, but *if* a major outbreak starts off then the deterministic description applies.
2. If population turnover is slow relative to disease transmission, we reach almost the final size situation of the closed epidemic before the gradual inflow of new susceptibles has any effect. In this situation, there are very few infecteds and the density of susceptibles is of the order of $S_0 e^{-R_0}$, which is far below threshold. It will take a long time before the density is above threshold again, and during this period demographic stochasticity may easily lead to the extinction of the infectious agent. But when and how to switch to a stochastic model?
3. Even if the infective agent escapes extinction after the first outbreak and becomes established, there is still the possibility of extinction as a result of chance fluctuations. Relevant questions are: what is the expected time until extinction if we are currently in the 'endemic' regime? (See Nåsell, this volume); what is the expected total number of cases during that time?

A general feeling is that the possibility of extinction due to demographic stochasticity is strongly enhanced if the deterministic dynamics is characterised by oscillations, rather than by a stable endemic steady state.

Bartlett (1960), Dietz & Schenzle (1985) and others mainly concentrate on 3. The distinction between 2 and 3 is also mentioned in passing as epidemic fade-out versus endemic fade-out in the book by Anderson and May (1991, p. 20). The notion of a *critical community* size (Bartlett (1960), Dietz (1982), Schenzle and Dietz (1987)) seems to come in for two reasons. The first is that Ne^{-R_0} may still be reasonably large if the total population size N is large (a low density over a large domain may yield an appreciable number). The second is that the geographical distribution by itself may necessitate a reconsideration of the process of disease transmission. If local epidemics are out of phase, then the proneness to extinction may be much smaller (cf. metapopulation models in ecology, Gilpin and Hanski (1991)). See Grenfell (this volume) for this aspect and others, related to age structure, of the fade-out phenomenon in the context of measles. Concerning 2, see Rand and Wilson (1991) for other aspects.

Let us now consider case *a* where the birthrate is assumed to be a constant. Basic issues are the existence, uniqueness, representation and stability of endemic *steady states*. The representation is important when one wants to use estimates of endemic levels for parameter estimation (cf. Hasibeder, this volume). Stability has both a local aspect (where do the roots of the characteristic equation lie?) and a global aspect (e.g. can we find a Lyapunov function?). See e.g. Hethcote (1976), Bailey (1975). The generic picture that emerges is that there exists a (unique) steady state if and only if $R_0 > 1$ (but see section 5 below). Recent issues are: if the steady state is unstable, what sort of dynamics can we expect? Periodic oscillations or chaos?; if contact rates are periodic (seasons, school system) then how are the periods of the solutions related to the driving period? For results in this line we refer to Smith (1983), Aron and Schwarz (1984), Dietz (1976), Kuznetsov and Piccardi (1992), Schenzle (1984), Schwarz (1985, 1988), Schaffer (1985), Hethcote and Levin (1989) and Grenfell, this volume.

The regulation case *b* is characterised by the existence of multiple thresholds if there is differential mortality. Typically we find four regimes for a contact rate parameter according to the following types of dynamic behaviour:

- the infectious agent dies out;
- the infectious agent grows exponentially, but at a slower rate than the host population (the dilution effect is that the *proportion* of infecteds in the population goes to zero);
- host and infectious agent grow at the same rate, which is reduced relative to the host growth rate in the disease free situation;
- either the common growth rate is negative or a steady state obtains, depending on how exactly one models the contact rate (strictly homogeneous or only asymptotically homogeneous).

For this type of result see Anderson (1979), Diekmann and Kretzschmar (1991), Busenberg and Van den Driessche (1990), Busenberg and Hadeler (1990), Anderson and May (1991), Thieme (1992) and Mena-Lorca

and Hethcote (1992).

Results become more subtle when one allows for the interaction of disease effects with other density dependent effects. This area currently receives much attention, see, e.g., Brauer (1990), Pugliese (1990), Greenhalgh (1992a,b), Greenhalgh and Das (1992) and Gao and Hethcote (1992), but we feel that time-scale aspects deserve more prominence (see Andreasen (1989, 1992a,b)). For a recent general approach to SIS density-dependent models, see Zhou and Hethcote (1993).

The following table summarises our classification of aspects of infectious-disease dynamics.

- closed population
  - invasion/elimination
  - time course
  - final size
- repeated outbreaks
- endemic situation
  - steady state
  - periodic oscillations
  - chaos
- feedback to demography
  - independent growth
  - partial regulation: reduced growth of host
  - regulation.

4. Heterogeneity

Suppose that not all individuals are equally susceptible, but that certain traits (e.g. age, gender or whether or not one suffers from a sexually transmitted disease causing ulcers, when we consider HIV transmission) have a marked influence. Of course one then has to specify these traits, their dynamics and their frequency in the susceptible population. Having done that, one question is: can we average and if so, how to do it?

In order to have a common formulation for both static and dynamic traits it is most convenient to parametrise by the trait an individual has at the moment it becomes infected (we will also write 'at birth'). Let now A be a function of three variables defined by

$$A(\tau,\xi,\eta) = \begin{array}{l}\text{the expected infectivity of an individual that was infected } \tau \text{ units of time ago} \\ \text{while having trait value } \eta \text{ towards a susceptible with trait value } \xi\end{array} \tag{4.1}$$

then exactly the same reasoning which led to (2.1) yields, in the case of a closed population,

$$\frac{\partial S}{\partial t}(t,\xi) = S(t,\xi)\int_\Omega\int_0^\infty A(\tau,\xi,\eta)\frac{\partial S}{\partial t}(t-\tau,\eta)d\tau d\eta \tag{4.2}$$

where Ω denotes the set of trait values. So the structure remains essentially the same as in (2.1), but the way to proceed is slightly more involved. We have to deal with distributed quantities and replace straightforward multiplication by an operator mapping a function onto a new function. The linearised version of (4.2) has a solution of the form $\frac{\partial S}{\partial t}(t,\xi) = \Psi(\xi)e^{\lambda t}$ if and only if $\Psi(\cdot)$ is an eigenvector of the operator K_λ defined by

$$(K_\lambda\phi)(\xi) = S_0(\xi)\int_\Omega\int_0^\infty A(\tau,\xi,\eta)e^{-\lambda\tau}d\tau\phi(\eta)d\eta \tag{4.3}$$

(here $S_0(\cdot)$ is the demographic steady state at the start of the epidemic). Note that K_0 is the *next-generation operator* corresponding to the linearisation. This means that, given a generation of infecteds distributed as ϕ with respect to trait value at birth and of size $\int_\Omega\phi(\eta)d\eta$, the function $K_0\phi$ describes both the size and the distribution of the next generation. The positivity of A guarantees that K_0 is a positive operator on $L_1(\Omega)$, the space of integrable functions, and under appropriate minor extra conditions (minor in the sense

that they will generally be fulfilled in practical situations) one can conclude that K_0 has a ***strictly*** dominant eigenvalue R_0. We can rightfully identify this eigenvalue with R_0 since, under those minor extra conditions, we have the asymptotic relation

$$K_0^n \phi \sim R_0^n c(\phi)\phi_d, \qquad n \to \infty, \tag{4.4}$$

where $c(\phi)$ is a scalar depending on the initial generation, and ϕ_d the eigenvector corresponding to R_0. So, if we iterate the next-generation operator, the distribution of infected individuals over all trait values stabilises to the form described by ϕ_d, while numbers are multiplied by R_0 from generation to generation. In other words, ϕ_d describes the distribution of the 'typical' infected individual and R_0 is the number of secondary cases. With the normalisation $\int_\Omega \phi_d(\eta)d\eta = 1$, the eigenvector has the interpretation of a probability distribution for the trait value at the moment of infection.

Remark: In the host-vector case, the population can be divided into two subpopulations which do not communicate internally and hence transmission has a well-defined cycle (of length two). The next-generation matrix then has the anti-diagonal structure

$$K_0 = \begin{pmatrix} 0 & K_{hv} \\ K_{vh} & 0 \end{pmatrix}$$

and $R_0 = \sqrt{\lambda_d(K_{hv}K_{vh})} = \sqrt{\lambda_d(K_{vh}K_{hv})}$, where $\lambda_d(M)$ denotes the dominant eigenvalue of the (positive) matrix M. In this case it is, in the biological literature, more usual to choose R_0^2 for the quantity one calls R_0, in particular since R_0^2 admits a more direct interpretation, as the host-to-host (and vector-to-vector) multiplication factor. Useful as this may seem, one should realise that such a choice can lead to a proliferation of R_0's (think for example of two loosely coupled groups, where one can define R_0^{rig} ('return-in-group') as the number of first offspring in ones own group, i.e. the sum of the cases in ones own group produced either directly or indirectly via an arbitrarily long transmission chain in the other group). We therefore sympathise with Hasibeder's suggestion (Hasibeder, this volume) to use the notation R_0^c for the dominant eigenvalue of the iterated next-generation matrix in the case of a well-defined transmission cycle.

In the heterogeneous case we have, as before, the equivalence

$$R_0 > 1 \Leftrightarrow r > 0 \tag{4.5}$$

where now r is defined as the (real) value of λ for which K_λ has dominant eigenvalue one, i.e. r is the 'real-time' growth rate (in contrast to the homogeneous case, the proof of the equivalence requires some work, see Heesterbeek (1992) for one possible proof).

Here a cautionary remark is in order. We have to worry about how irreducible the kernel A and how dominant r really is before we can rightfully conclude that the epidemic grows as e^{rt} in real time. For instance, when we consider a very large spatial domain, then the speed of the epidemic is not described by r but rather by the asymptotic speed of propagation c_0 (see Metz and van den Bosch, this volume). In this connection we also mention that the Perron root of K_0 (which is in a certain way a measure of *local* changes; see Jagers (1992), Shurenkov (1992)) and the spectral radius R_0 (which is more concerned with total growth) may differ when the epidemic drifts off towards infinity while growing (e.g. think of a focus of a fungal plant disease in a field of wheat where the spread of the epidemic can be heavily influenced by a strong prevailing wind direction). In the theory of branching processes r, there usually referred to as the Malthusian parameter, recently has become associated with the Perron root rather than the spectral radius, see Jagers (1992), Shurenkov (1992), Taib (1992). However, the two notions can only lead to different results for non-compact trait spaces (Shurenkov (1992)). As another example, consider two very loosely coupled subpopulations, one small but highly active and one big with low activity. Then it may very well be that the nonlinearity comes into play in the small group before enough time has elapsed for the stable invasion distribution to be attained.

Anyhow, concerning the invasion/elimination criterion we can conclude that there is a systematic way of performing the right averaging: compute the dominant eigenvalue R_0 (R) of the next-generation operator and compare it with 1. This still leaves us with important basic problems:

- how to express the kernel A in terms of ingredients of submodels for the contact- and transmission process?
- how to actually compute R_0 (R) for a given kernel A?

We refer to Heesterbeek (1992), Diekmann, Heesterbeek and Metz (1990) and Diekmann (1991a,b) and the references therein for various results in this direction. In De Jong, Diekmann and Heesterbeek (1993a) an algorithm is given to compute the elements of the next-generation matrix for discrete-time multigroup models where the individuals are allowed to change their 'type'. Dietz (1993) gives a survey of various methods to estimate reproduction ratios.

The final size equation for the closed population takes the form

$$\ln \frac{S(\infty,\xi)}{S_0(\xi)} = \int_\Omega \int_0^\infty A(\tau,\xi,\eta) d\tau (S(\infty,\eta) - S_0(\eta)) d\eta \tag{4.6}$$

and has, in particular when ξ is a discrete variable, been studied by Radcliffe and Rass (1984) (see also Diekmann (1978), Thieme (1977a)). One can easily show that a nontrivial positive solution exists if and only if $R_0 > 1$, but apart from that very little can be said in any generality. The nonlinearity is an obstruction for a further simplification of (4.6), the problem remains higher-, usually even infinite-, dimensional and cannot be summarised in terms of one or two numbers as in the case of the invasion problem.

In heterogeneous populations, the invasion/elimination problem is the only part of the classification table in section 3 that has been addressed in some generality. As far as the other aspects listed in that table are concerned, results are only available for special kinds of structure. There are many results on age-structured models, on spatial structure and on models incorporating a finite number of groups. While not at all claiming to list all important contributions, we mention a few papers that tackle aspects of table 3 in heterogeneous populations. For a recent review of models for periodic outbreaks see Hethcote and Levin (1989) and Liu (1992). For the endemic equilibrium see for example Hethcote and Yorke (1984), Lajmanovich and Yorke (1976), Lin and So (1990), Busenberg and Van den Driessche (1992), Hethcote and Thieme (1985) and Beretta and Capasso (1986) in the case of a finite number of individual types, and Busenberg, Iannelli and Thieme (1991), Greenhalgh (1988) and Inaba (1990) in the case of age structure. For the interaction with demography see for example Busenberg, Cooke and Thieme (1991) for multigroup models, and May, Anderson and McLean (1988a,b), Tuljapurkar and John (1991) for age structured models.

5. Submodels for the contact process

What exactly constitutes a 'contact' depends on the disease being modelled. For example, for sexually transmitted diseases contacts take place at two levels (partners and sexual contacts within partnerships), and for some host-vector diseases a different type of contact is involved for the two transmission steps. Even if contacts are symmetric, the transmission probability, given contact, need not be so.

Disregarding heterogeneity for a moment, a first question is how many contacts an individual makes as a function of population size. (Here we have to carefully distinguish between the cases where our variables describe numbers and where they describe (spatial) densities (see De Jong, Diekmann and Heesterbeek (1993b) for a discussion of both theoretical arguments and experimental results concerning this point); in this paper we consistently work with densities.) This is the functional response question from predator-prey ecology, but now in an epidemic context. The following approaches have been taken:

a.) mass-action: the per capita number of contacts per unit of time is a linear function of density;
b.) saturation: the linear function is replaced by $\frac{aN}{1+bN}$ as a convenient phenomenological description without mechanistic underpinning (see Dietz (1982));
c.) extreme saturation: the linear function is replaced by a constant. Even though this does not make sense at extremely low densities, it seems a reasonable assumption for, e.g., sexual contacts or blood meals taken by mosquitoes;

d.) Holling squared: In Heesterbeek and Metz (1993) a submodel for the contact process is considered for pair formation at a short time-scale, together with a quasi-steady-state assumption to derive the functional response in much the same way as one derives the famous Holling disc expression of b) above from a submodel of prey search and handling by a predator. In the simplest situation the argument works as follows. Let x denote the (local spatial) density of 'free' individuals and let c denote the density of pairs of individuals involved in a contact. One assumes that there are constants ρ and σ, respectively the contact rate parameter and the inverse of the average duration of contacts, such that

$$\begin{aligned} \frac{dx}{dt} &= -\rho x^2 + 2\sigma c \\ \frac{dc}{dt} &= \frac{1}{2}\rho x^2 - \sigma c. \end{aligned} \tag{5.1}$$

Solving the steady state equations of (5.1) under the constraint that $x + 2c = n$, one obtains for the number of contacts per individual per unit of time $\frac{\rho x^2}{n}$, the expression

$$\frac{2\rho n}{1 + 2\rho n/\sigma + \sqrt{1 + 4\rho n/\sigma}} \tag{5.2}$$

which behaves like $2\rho n$ for small densities n, while approaching the limit σ for $n \to \infty$.

If we add heterogeneity again, matters quickly get complicated. The approach d) still works but usually no longer leads to explicit expressions (for several possible 'next steps', e.g. numerical integration, this need not trouble us too much).

When one has a constant functional response, as in c), strange things may happen since the reduction of the size of some groups may *increase* the infective 'pressure' on an other group. Indeed, Huang, Cooke, Castillo-Chavez (1992) showed by way of example that one may have a bifurcation diagram as sketched in figure 3, where there is bistability for $R_0 < 1$ (see Jacquez, Simon and Koopman, this volume, for a related phenomenon: R_0 of the whole population may increase when some subpopulation reduces its contact intensity with another group, if one works with a constant functional response).

In addition, we have to face the consistency problem: the total number of contacts (per unit of time) of all individuals of type ξ with all individuals of type η has to be equal to the same quantity with ξ and η interchanged. This is automatically satisfied for the extended version of (5.1) and for the mass-action type selective mixing of Morris, this volume, but has to be achieved by a suitable choice of mixing pattern in the case c) of a constant functional response. In Busenberg and Castillo-Chavez (1991) all possibilities have been classified. In this purely descriptive approach, the free parameters have no clear interpretation. Other authors like Sattenspiel (1987), Sattenspiel and Castillo-Chavez (1990), and Jacquez et al. (1988) have introduced submodels which allow somewhat more of a mechanistic interpretation. However, the matching of supply and demand by a market mechanism incorporating preferences remains a difficult problem in combinatorial sociology. Of course, the acquisition and analysis of actual data adds another difficult component to this problem (see Morris, this volume).

Another special case is when an individual is either single or paired to one other individual, where pairs remain together for a substantial period of time (as opposed to the short time-scale pairing of d) above) which however may not cover the entire period of infectiousness (as opposed to the long time-scale bonds described in the previous paragraph), and where partnerships are formed at random. One can again derive a next-generation operator which has a dominant eigenvalue deserving the name R_0. This is explained in detail in Diekmann, Dietz and Heesterbeek (1991) and used in Dietz, Heesterbeek and Tudor (1993) to analyse, among other things, the effect that 'wasted' contacts (i.e. contacts between two infected partners) have on the value of R_0. An annoying feature of the comparison of various members of a family of models with each other, that arises here (and elsewhere), is that it is not directly obvious how to gauge the models in order to make them comparable in the first place.

A final question concerning the contact process is the following: is the deterministic limit meaningful? Clearly when individuals interact only with their nearest neighbours on a spatial lattice, the answer is no and

one has to turn to cellular automata (or interacting particle systems, in another jargon; see Durret and Levin (1992) for a nice overview; also see Mollison, this volume, and Durret, this volume). If the neighbourhood structure has no regularity, but still individuals interact only with a fixed group of others, at least for some period of time, one is considering an epidemic on a random graph, see Blanchard, Bolz and Krüger (1990). However, very little headway with such models has been made on a general level.

6. Final remarks

Differences in behaviour necessitate subtle ways of averaging. In the linearisation appropriate for the initial phase of an epidemic one can do this in quite some generality and arrive at one or two numbers (R_0 and r or c_0 (the asymptotic speed of spatial propagation, see Metz and van den Bosch, this volume)) to describe the dynamics. Things are far less clear for other aspects of the dynamics, such as the final size of an epidemic, the 'size' and stability of an endemic steady state, the growth rate reduction in the 'regulation' setting. Here we need clever case studies, where 'clever' means that one has to consider simplifications which are not too unrealistic while making the problem tractable.

In addition there is a need for quasi-mechanistic submodels especially for the contact process, dealing with such issues as handling time, satiation, virus transport by aerosols, modes of spore dispersal, lining up of seals on a sandbank, etc. (The last example illustrates a further difficulty in the modelling of contact rates: when numbers go down, the *effective* density may stay constant since the nearest neighbour distances remain roughly the same.)

We have to contemplate what is the time-scale of the various processes that we are combining into one model, in order to decide what should be considered as constant and what as variable on the time-scale that we focus on (see Andreasen (1992a,b) for a nice example of the power that the exploitation of differences in time-scale may have).

We reiterate that the border between stochastic and deterministic phases of the dynamics of disease is hard to define or to determine, yet is very relevant to our overall understanding.

In modelling one can either use a top-down approach, where one starts to build a general abstract framework and then gradually gives a more concrete specification of various ingredients, or a bottom-up approach, where one first concentrates on the data concerning a specific disease in a given population and then gradually tries to describe and analyse the essential mechanisms and phenomena. R_0 and c_0 are among the success-stories of the top-down approach, the current understanding of measles dynamics illustrates the potential power of the bottom-up approach. Although we classify ourselves as 'top-down' people, and have written this paper in a top-down spirit, we think that both approaches should be followed. A chain of people with partially overlapping interests and knowledge should bridge the gap between the two starting points by striving for unity in the formulation of mathematical models and the type of data that are collected. The Cambridge workshop was an excellent catalyst for the formation of such chains.

After these somewhat pedantic remarks we like to close with paraphrasing Simon Levin's observation during the conference dinner:

"to are or to R_0, that was the question."

Acknowledgement: This paper was written during a stay at the Isaac Newton Institute for Mathematical Sciences; we thank the Newton Institute for its hospitality. The second author would like to thank the Netherlands Organisation for Scientific Research NWO and the Royal Society for financial support during the writing of this paper.

References

Adler F.R. & M. Kretzschmar (1992): Aggregation and stability in parasite-host models. *Parasitology* **104**: 199-205.

Aitchison, J. & G.S. Watson (1988): A not-so-plain tale from the Raj. In: *The Influence of Scottish Medicine*, D. Dows (ed.), Parthenon, Park Ridge, New Jersey, pp. 113-128.

Anderson, R.M. (1979): The persistence of direct life cycle infectious diseases within populations of hosts. *Lect. on Math. in Life Sciences* **12**: 1-67.

Anderson, R.M. & R.M. May (1978): Regulation and stability of host-parasite population interactions. I. Regulatory processes. *J. Anim. Ecology* **47**: 219-247.

Anderson, R.M. & R.M. May (1979): Population biology of infectious diseases, part 1. *Nature* **280**: 361-367.

Anderson, R.M., & R.M. May (1991): *Infectious Diseases of Humans, Dynamics and Control.* Oxford University Press, Oxford.

Andreasen, V. (1989): Disease regulation of age-structured host populations. *Theor. Pop. Biol.* **36**: 214-239.

Andreasen, V. (1992a): The effect of age-dependent host mortality on the dynamics of an endemic disease, preprint.

Andreasen, V. (1992b): Instability in an SIR-model with age-dependent susceptibility, preprint.

Aron, J.L. & I.B. Schwarz (1984): Seasonality and period-doubling bifurcations in an epidemic model. *J. Theor. Biol.* **110**: 665-679.

Bailey, N.J.T. (1975): *The Mathematical Theory of Infectious Diseases and Its Application.* Griffin, London.

Bartlett, M.S. (1960): *Stochastic Population Models in Ecology and Epidemiology.* Methuen, London.

Beretta, E. & V. Capasso (1986): On the general structure of epidemic systems. Global asymptotic stability. *Comp. & Maths. with Appls.* **12A**: 677-694.

Blanchard Ph., Bolz, G.F. and T. Krüger (1990): Modelling AIDS-Epidemics or any venereal disease on random graphs. In: *Stochastic Processes in Epidemic Theory*, J.-P. Gabriel, C. Lefèvre & P. Picard (eds.), Lecture Notes in Biomathematics No. 86, pp. 104-117.

Brauer, F. (1990): Models for the spread of universally fatal diseases. *J. Math. Biol.* **28**: 451-462.

Busenberg, S.N. & C. Castillo-Chavez (1991): A general solution of the problem of mixing of subpopulations and its application to risk- and age-structured epidemic models for the spread of AIDS. *IMA J. Math. Appl. Med. Biol.* **8**: 1-29.

Busenberg, S.N. & K. Cooke (1993): *Vertically Transmitted Diseases, Models and Dynamics.* Biomathematics vol. 23, Springer-Verlag, Berlin.

Busenberg, S.N., Cooke, K. & H.R. Thieme (1991): Demographic change and persistence of HIV/AIDS in a heterogeneous population. *SIAM J. Appl. Math.* **51**: 1030-1052.

Busenberg, S.N. & K.P. Hadeler (1990): Demography and epidemics. *Math. Biosc.* **101**: 63-74.

Busenberg, S.N., Iannelli, M. & H.R. Thieme (1991): Global behavior of an age-structured SIS epidemic model. *SIAM J. Math. Anal.* **22**: 1065-1080.

Busenberg, S.N. & P. van den Driessche (1990): Analysis of a disease transmission model in a population with varying size. *J. Math. Biol.* **28**: 257-270.

Busenberg, S.N. & P. van den Driessche (1992): Disease transmission in multigroup populations of variable size. To appear in: *Proceedings 3rd International Conference on Mathematical Population Dynamics*, Pau, 1992.

De Jong, M.C.M., Diekmann, O. & J.A.P. Heesterbeek (1993a): The computation of R_0 for discrete-time epidemic models with dynamic heterogeneity. *Math. Biosc.*, in press.

De Jong, M.C.M., Diekmann, O. & J.A.P. Heesterbeek (1993b): How does infection-transmission depend on population size? This volume.

Diekmann, O. (1978): Thresholds and travelling waves for the geographical spread of infection. *J. Math. Biol.* **6**: 109-130.

Diekmann, O. (1991a): Modelling infectious diseases in structured populations. In: B.D. Sleeman & R.J. Jarvis (eds) *Ordinary and Partial Differential Equations, vol. III.* Pitman RNiMS 254: 67-79. Longman, Harlow.

Diekmann, O. (1991b): An invitation to structured (meta)population models. *CWI-report* AM-R9113, to appear in: *Patch Dynamics.*

Diekmann, O., Dietz, K. & J.A.P. Heesterbeek (1991): The basic reproduction ratio for sexually transmitted diseases, part 1: theoretical considerations. *Math. Biosc.* **107**: 325-339.

Diekmann, O., Heesterbeek, J.A.P. & J.A.J. Metz (1990): On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**: 365-382.

Diekmann, O. & M. Kretzschmar (1991): Patterns in the effects of infectious diseases on population growth. *J. Math. Biol.* **29**: 539-570.

Dietz, K. (1976): The incidence of infectious diseases under the influence of seasonal fluctuations. Lect. Notes Biom. vol. 11, pp. 1-15.

Dietz, K. (1982): Overall population patterns in the transmission cycle of infectious disease agents. In: *Population Biology of Infectious Diseases*, R.M. Anderson, R.M. May (eds), pp. 87-102, Springer-Verlag, Berlin.

Dietz, K. (1993): The estimation of the basic reproduction number for infectious diseases. *Stat. Meth. Med. Research* **2**: 23-41.

Dietz, K., Heesterbeek, J.A.P. & D.W. Tudor (1993): The basic reproduction ratio for sexually transmitted diseases, part 2: The influence of variable HIV-infectivity. *Math. Biosc.*, in press.

Dietz, K. & D. Schenzle (1985): Mathematical models for infectious disease statistics. In: *A Celebration of Statistics* A.C. Atkinson & S.E. Feinberg (eds.), Springer, Berlin.

Dobson, A. (1993): Population dynamics of a parasitic nematode and its host. This volume.

Durret, R. (1993): Spatial epidemic models. This volume.

Durret, R. & S.A. Levin (1992): Stochastic spatial models: a user's guide to ecological applications, preprint.

Gao, L.Q. & H.W. Hethcote (1992): Disease transmission models with density-dependent demographics. *J. Math. Biol.* **30**: 717-731.

Gilpin, M.E. & I. Hanski (1991): *Metapopulation Dynamics: Empirical and Theoretical Investigations.* Academic Press.

Greenhalgh, D. (1988): Threshold and stability results for an epidemic model with an age-structured meeting rate. *IMA J. Math. Appl. Med. Biol.* **5**: 81-100.

Greenhalgh, D. (1992a): Vaccination in density-dependent epidemic models. *Bull. Math. Biol.* **54**: 733-758.

Greenhalgh, D. (1992b): Some threshold and stability results for epidemic models with a density dependent death rate. *Theor. Pop. Biol.* **42**: 130-151.

Greenhalgh, D. & R. Das (1992): An SIR epidemic model with a contact rate depending on population density, preprint.

Grenfell, B. (1993): Chaos and biological realism in measles dynamics. This volume.

Hasibeder, G. (1993): Heterogeneous disease transmission: estimating the basic reproduction number from prevalences. This volume.

Heesterbeek, J.A.P. (1992): R_0, PhD-Thesis, University of Leiden.

Heesterbeek, J.A.P. & J.A.J. Metz (1993): The saturating contact rate in marriage- and epidemic models. *J. Math. Biol.* **31**: 529-539.

Hethcote, H.W. (1976): Qualitative analysis for communicable disease models. *Math. Biosc.* **28**: 335-356.

Hethcote, H.W. (1993): A thousand and one epidemic models. To appear in: *Frontiers in Theoretical Biology*, S.A. Levin (ed.), Lect. Notes in Biomathematics, vol. 100, Springer, Berlin.

Hethcote, H.W. & S.A. Levin (1989): Periodicity in epidemic models. In: *Applied Mathematical Ecology*, S.A. Levin, T.G. Hallam, L.J. Gross (eds.), pp. 193-211, Springer-Verlag, Berlin.

Hethcote, H.W. & H.R. Thieme (1985): Stability of the endemic equilibrium in epidemic models with subpopulations. *Math. Biosc.* **75**: 205-227.

Hethcote, H.W. & J.A. Yorke (1984): *Gonorrhea Transmission Dynamics and Control.* Lect. Notes Biomath. Vol. 56, Springer-Verlag, Berlin.

Huang, W., Cooke, K.L. & C. Castillo-Chavez (1992): Stability and bifurcation for a multiple group model for the dynamics of HIV/AIDS transmission. *SIAM J. Appl. Math.* **52**: 835-854.

Inaba, H. (1990): Threshold and stability results for an age-structured epidemic model. *J. Math. Biol.* **28**: 411-434.

Jacquez, J.A., Simon, C.P., Koopman, J., Sattenspiel, L. & T. Perry (1988): Modeling and analyzing HIV transmission: the effect of contact patterns. *Math. Biosc.* **92**: 119-199.

Jacquez, J.A., Simon, C.P. & J.S. Koopman (1993): Core groups and the R_0's for subgroups in heterogeneous SIS and SI models. This volume.

Jagers, P. (1992): The deterministic evolution of general branching populations. Preprint 1992-25, Department of Mathematics, Chalmers University Göteborg.

Kermack, W.O. & A.G. McKendrick (1927): Contributions to the mathematical theory of epidemics, part I. *Proc. Roy. Soc. Edinb. A* **115**: 700-721.

Kermack, W.O. & A.G. McKendrick (1932): Contributions to the mathematical theory of epidemics, part II (the problem of endemicity). *Proc. Roy. Soc. Edinb. A* **138**: 55-83.

Kermack, W.O. & A.G. McKendrick (1933): Contributions to the mathematical theory of epidemics, part III (further studies of the problem of endemicity). *Proc. Roy. Soc. Edinb. A* **141**: 94-122.

Kretzschmar, M. & F.R. Adler (1993): Aggregated distributions in models for patchy populations. *Theor. Pop. Biol.* **43**: 1-30.

Kuznetsov, Yu.A. & C. Piccardi (1992): Bifurcation analysis of periodic SEIR and SIR epidemic models. *J. Math. Biol.*, to appear.

Lajmanovich, A. & J.A. Yorke (1976): A deterministic model for Gonorrhea in a nonhomogeneous population. *Math. Biosc.* **28**: 221-236.

Lin, X. & J.W.-H. So (1990): Global stability of the endemic equilibrium in epidemic models with subpopulations, preprint.

Liu, W.M. (1992): Models of recurrent outbreaks of infectious diseases. To appear in: *Proceedings 3rd International Conference on Mathematical Population Dynamics*, Pau, 1992.

May, R.M., Anderson, R.M. & A.R. McLean (1988a): Possible demographic consequences of HIV/AIDS epidemics: I. Assuming HIV infection always leads to AIDS. *Math. Biosc.* **90**: 475-505.

May, R.M., Anderson, R.M. & A.R. McLean (1988b): Possible demographic consequences of HIV/AIDS epidemics: II. Assuming HIV infection does not necessarily lead to AIDS. In: C. Castillo-Chavez, S.A. Levin, C.A. Shoemaker (eds.) *Mathematical Approaches to Problems in Resource Management and Epidemiology*, Lect. Notes Biomath. vol. 81: 220-248.

Mena-Lorca, J. & H.W. Hethcote (1992): Dynamic models of infectious diseases as regulators of population sizes. *J. Math. Biol.* **30**: 693-716.

Metz, J.A.J. (1978): The epidemic in a closed population with all susceptibles equally vulnerable; some results for large susceptible populations and small initial infections. *Acta Biotheoretica* **27**: 75-123.

Metz, J.A.J. & F. van den Bosch (1993): Velocities of epidemic spread. This volume.

Metz, J.A.J. & O. Diekmann (eds) (1986): *The Dynamics of Physiologically Structured Populations.* Lecture Notes in Biomath. vol. 68, Springer, Heidelberg.

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

Mollison, D. (1993): The structure of epidemic models. This volume.

Morris, M. (1993): Data driven network models for the spread of disease. This volume.

Nåsell, I. (1993): The threshold concept in stochastic epidemic and endemic models. This volume.

Pugliese, A. (1990): Population models for diseases with no recovery. *J. Math. Biol.* **28**: 65-82.

Radcliffe, J. & L. Rass (1984): The spatial spread and final size of the deterministic non-reducible n-type epidemic. *J. Math. Biol.* **19**: 309-327.

Rand, D.A. & H. Wilson (1991): Chaotic stochasticity: a ubiquitous source of unpredictability in epidemics. *Proc. R. Soc. Lond. B* **246**: 179-184.

Roberts, M.G., Smith, G. & B. Grenfell (1993): Mathematical models for macroparasites of wildlife, to appear in: *Ecology of Infectious Diseases in Natural Populations*, B.T. Grenfell & A.P. Dobson (eds.), Cambridge University Press, Cambridge.

Ross, R. & H.P. Hudson (1917): An application of the theory of probabilities to the study of a priori pathometry, part III. *Proc. Roy. Soc. Lond. A* **43**: 225-240.

Sattenspiel, L. (1987): Population structure and the spread of disease. *Hum. Biol.* **59**: 411-438.
Sattenspiel, L. & C. Castillo-Chavez (1990): Environmental context, social interactions and the spread of HIV. *Am. J. Hum. Biol.* **2**: 397-417.
Schaffer, W.M. (1985): Can nonlinear dynamics elucidate mechanisms in ecology and epidemiology? *IMA J. Math. Appl. Biol. Med.* **2**: 221-252.
Schenzle, D. (1984): An age-structured model of pre- and post-vaccination measles transmission. *IMA J. Math. Applied in Med. Biol.* **1**: 169-191.
Schenzle, D. & K. Dietz (1987): Critical population sizes for endemic virus transmission. In: *Räumliche Persistenz und Diffusion von Krankheiten*, W. Fricke & E. Hinz (eds.), Heidelberger Geographische Arbeiten, vol. 83, Heidelberg, pp. 31-42.
Schwarz, I.B. (1985): Multiple stable recurrent outbreaks and predictability in seasonally forced nonlinear epidemic models. *J. Math. Biol.* **21**: 347-361.
Schwarz, I.B. (1988): Nonlinear dynamics of seasonally driven epidemic models. In: *Proc. 12th IMACS World Congress* R. Vichnevetsky, P. Borne and J. Vignes (eds.), vol. IV. pp. 166-169.
Shurenkov, V.M. (1992): On the relationship between spectral radii and Perron roots, preprint 1992-17, Chalmers Univ. of Technology, Dept. of Math.
Smith, H.L. (1983): Multiple stable subharmonics for a periodic epidemic model. *J. Math. Biol.* **17**: 179-190.
Taib, Z. (1992): *Branching Processes and Neutral Evolution*, Lect. Notes Math. Biol. vol. 93, Springer-Verlag, Berlin.
Thieme, H.R. (1977a): A model for the spread of an epidemic. *J. Math. Biol.* **4**: 337-351.
Thieme, H.R. (1977b): The asymptotic behaviour of solutions of nonlinear integral equations. *Math. Z.* **157**: 141-154.
Thieme, H.R. (1980): On the boundedness and the asymptotic behaviour of the non-negative solutions of Volterra-Hammerstein integral equations. *Manuscr. Math.* **31**: 379-412.
Thieme, H.R. (1992): Epidemic and demographic interaction in the spread of potentially fatal diseases in growing populations. *Math. Biosc.* **111**: 99-130.
Tuljapurkar, S.D. & A.M. John (1990): Disease in changing populations: growth and disequilibrium. *Theor. Pop. Biol.* **40**: 322-353.
Zhou, J. & H.W. Hethcote (1993): Population size dependent incidence in models for diseases without immunity, preprint.

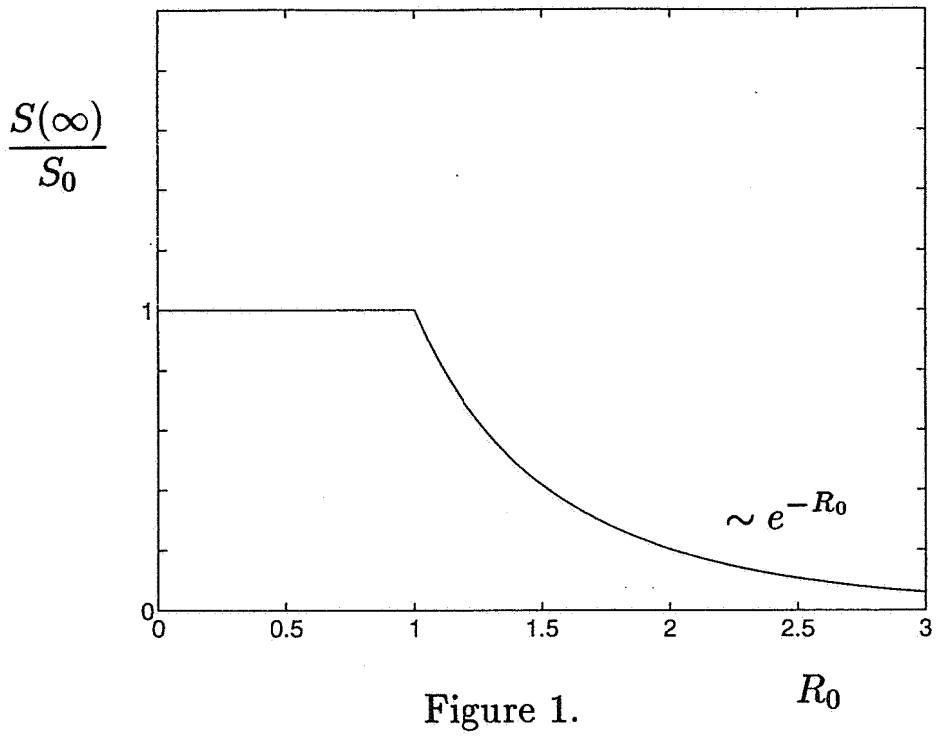

Figure 1.

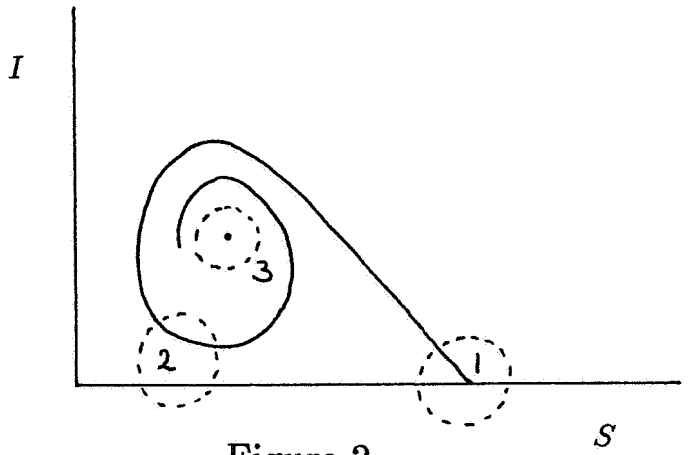

Figure 2.

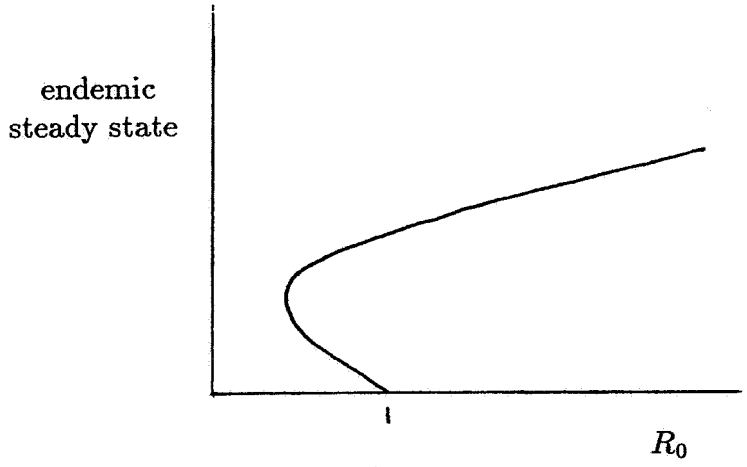

Figure 3.