INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

UNITED NATIONS INDUSTRIAL DEVELOPMENT ORGANIZATION

INTERNATIONAL CENTRE FOR SCIENCE AND HIGH TECHNOLOGY
c/o INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS 34100 TRIESTE (ITALY) VIA GRIGNANO, 9 (ADRIATICO PALACE) P.O. BOX 586 TELEPHONE 040-224572 TELEFAX 040-224575 TELEX 460449 APH I

**SMR.648 - 23**

# SECOND AUTUMN WORKSHOP ON MATHEMATICAL ECOLOGY

**(2 - 20 November 1992)**

---

# "Modelling Infectious Diseases in Structured Populations"

**O. Diekmann**
**Centrum voor Wiskunde en Informatica**
**Kruislaan 413**
**1098 SJ Amsterdam**
**The Netherlands**

---

**These are preliminary lecture notes, intended only for distribution to participants.**

# Modelling Infectious Diseases in Structured Populations

Odo Diekmann
*Centre for Mathematics and Computer Science*
*Kruislaan 413, 1098 SJ Amsterdam, the Netherlands*
*&*
*Institute of Theoretical Biology, University of Leiden*
*Kaiserstraat 63, 2311 GP Leiden, the Netherlands*

After a brief survey of epidemic modelling we concentrate on the invasion problem. The basic reproduction ratio $R_0$ is defined as the dominant eigenvalue of the next-generation operator. Several examples involving various structuring variables are then presented.

*1980 Mathematics Subject Classification:* 92A15
*Keywords & phrases:* Epidemic models, basic reproduction ratio, basic reproductive rate.
*Note:* Paper presented at the 11th Dundee Conference on the Theory of Ordinary and Partial Differential Equations, 3-6 July 1990.

## 1. Introduction

Epidemiology is the study of the spread, in space and time, of diseases, with the objective to trace factors that are responsible for, or contribute to, the occurrence of a disease. Epidemiology is concerned with human, animal as well as plant diseases, but antropocentrism manifests itself in the predominant attention for human diseases and diseases affecting economically important livestocks and crops.

For a long time infectious diseases dominated the field (note, however, that for most diseases it was not always clear that they were caused by an infective agent). Gradually, and notably in the industrialized world, degenerative diseases, like cancer and cardio-vascular dysfunctions, became prominent. Recent events, in particular the advance of the human immunodeficiency virus and the resurgence of malaria, have put infectious diseases back to the forefront.

The work of an epidemiologist comprises

- description: laying out an atlas displaying the incidence (the number of new cases per unit of time) or the prevalence (the total number of cases) as a function of geographical position and time (see Cliff & Haggett (1988) for a master piece)
- data analysis: trying to find causative effects, such as contaminated drinking water in the case of cholera (see Cliff & Haggett (1988), Ch. I) or estimating relevant parameters (see N.G. Becker (1989)) by a careful (statistical) examination of the available data
- explanation: developing theories based on mechanistic considerations at the individual level
- prevention: designing control strategies, for instance screening or vaccination programs
- prediction: forecasting future developments in incidence and prevalence

The present paper is concerned with mathematical models for the spread of infectious diseases. The understandable expectation that such models are primarily useful as a tool for prediction is actually (and unfortunately) not warranted, since, as a rule, our knowledge of key parameters and our understanding of the details of the transmission mechanism is so poor, that accurate quantitative results are out of reach. However, as I hope to demonstrate, models do enhance our understanding of the causal relationship between phenomena at the population level and mechanisms which act in and between individuals. Moreover, they are helpful in evaluating the possible effect of control measures.

Report AM-R9017
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

In: B.D. Sleeman & R.J. Jarvis (eds) Ordinary and Partial Differential Equations, vol. III, Pitman RNiMS 254 (1991) 67-79 (Longman, Harlow)

Let me briefly list some of the justifications of mathematical modelling exercises in epidemiology (see Hethcote (to appear) for a more extensive list).

A mathematical model requires a precise formulation of hypotheses concerning the physiological and behavioural mechanisms involved. Thus we become aware of *working hypotheses* which may otherwise go unnoticed and which, actually, may deserve scrutiny.

Mathematical modelling may catalyse the introduction of key *concepts*. For instance, the basic reproduction ratio $R_0$ (commonly but incorrectly called 'basic reproductive rate') was introduced by Sir Ronald Ross in 1909 when be analysed a mathematical model for malaria transmission. $R_0$ is (roughly speaking: a precise mathematical definition is given below) the number of secondary cases produced by a typical infected individual. By definition then the prevalence increases when $R_0>1$ and decreases when $R_0<1$. Ross showed that for malaria $R_0$ is proportional to the quotient of mosquito density and human population density and that, consequently, the disease can be eliminated in a certain area by sufficiently reducing the mosquito density.

Mathematical models allow us to do *thought experiments* where actual experiments are impossible, immoral or would demand too much time and/or money. An example is provided by the comparative study of various vaccination schedules aiming at the elimination or eradication[1] of (childhood) diseases (see Knox (1980), Anderson & May (1985.I) and van Druten e.a. (1986))

Mathematical models allow us to identify 'mechanistic' links between various kinds of data. Not only can they be used for *parameter estimation*, but they may also make us aware of our *ignorance* concerning numerical values of key parameters (e.g. the distribution of the number of different sexual partners per year) and tell us, via a *sensitivity analysis*, which parameters deserve to be called key parameters.

To put the work described below into perspective, I next want to present two ways of classifying epidemic models followed by a list of phenomena that can occur on various time scales.

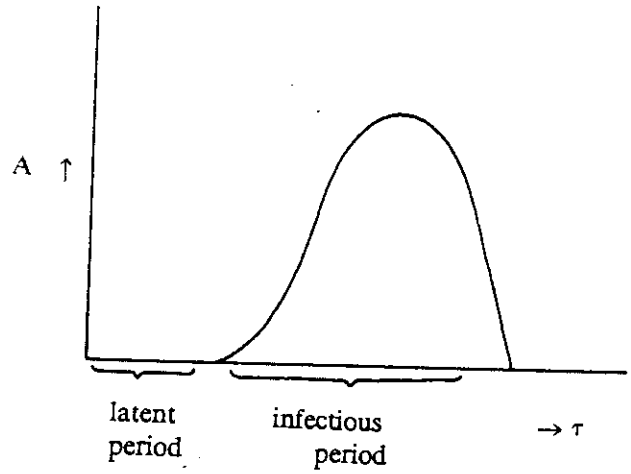


FIGURE 1

Suppose that after infection the disease develops as an autonomous process within the infected individual or, more precisely, that the invading organism reproduces within the host at such a rate that further infections are irrelevant (examples include measles, influenza and rabies and fall under the general heading of micro-parasitic diseases). Irrespective of the precise nature of the physiological variables required to describe the 'state' of an individual accurately, we can then work with a '$d$-age' representation, where $d$-age refers to the time $\tau$ elapsed since infection. In particular we can use a

1. Elimination requires that the disease-free state is made stable: when the disease causing organism has disappeared worldwide, but re-introduction would cause an epidemic, we speak of eradication.

function $A$, describing the average *infectivity* as a function of $\tau$, as our main modelling ingredient. Typically the graph of $A$ is as depicted in Figure 1. One can interpret $A$ either as a deterministic property or as an expectation, where we imagine very many 'stochastic' individuals (distributed in a manner which does not change under the influence of the progressing epidemic). In this way the familiar 'compartmental' models of the S-E-I-R (Susceptible-Exposed-Infectious-Removed) type are included (here residence time in each compartment is exponentially distributed while the infectivity is a fixed constant during the stay in the I-compartment; see Section 4).

For other diseases infection is not a unique event, but rather a repeated process. Meanwhile reproduction within the host can or cannot take place. The number of parasites a host harbours can be taken as a rough description of the 'state' of an individual (with respect to the disease). Schistosomiasis and other worm diseases are the basic examples of such *macro-parasitic* diseases. In this paper we shall not be concerned with the corresponding class of models (see Anderson & May (1985.I), Hadeler & Dietz (1984) and Kretzschmar (1989)).

The essential mechanistic difference between both kind of models derives from the environmental impact after the first infection. The much discussed but little understood phenomenon of 'superinfection' in malaria (see Bailey (1982) and the references given there) indicates that faithful models of malaria may require a submodel for the interaction of the immune system and successive broods of parasites. So, in other words, the two types of models reflect current practice rather than a dichotomy of the natural world.

As already indicated above, we may, starting from stochastic models at the individual level, arrive at *deterministic* models at the population level by applying the law of large numbers. Truly *stochastic* models are needed to describe what happens in smaller groups, like households, schools or farms. In this paper we restrict ourselves to deterministic models. See Bartlett (1960), Bailey (1975), Becker (1989), Frauenthal (1980) and Ludwig (1974) for stochastic models.

What are relevant questions that mathematical models should help to answer? I shall briefly review:

(i) invasion
(ii) the time course of an epidemic
(iii) recurrent outbreaks
(iv) endemic steady states and oscillations
(v) regulation

(i) Given a population in a demographic steady (at least at the time scale of disease transmission) state, does the disease prevalence increase when initially very low? In other words: what is $R_0$ and is it bigger that one?

The invasion problem is actually relevant for the eradication/elimination problem as well. As a rule nonlinearities in epidemic models exhibit a certain monotonicity which makes the disease free state *globally* stable when $R_0<1$. So if a disease has already established itself in a population and we wonder whether certain control measures would be strong enough to get rid of it, we can equivalently ask whether or not the control measures would bring $R_0$ below one.

(ii) When $R_0>1$ the introduction of the disease agent will trigger an epidemic. Suppose we may consider the population as demographically closed at the time scale of disease transmission, then we may ask such questions as: when do prevalence and incidence reach their maximal values and how big will these values be? What fraction of the initialy present susceptible population escapes from ever getting the disease? As early as 1927 W.O. Kermack and A.G. McKendrick analysed such questions in the context of a rather general model (see Kendall (1956) for a detailed analysis of a special case; how much Kermack and McKendrick were ahead of their time cannot be judged from later references, which usually deal with the special case only, but only from reading the original paper!).

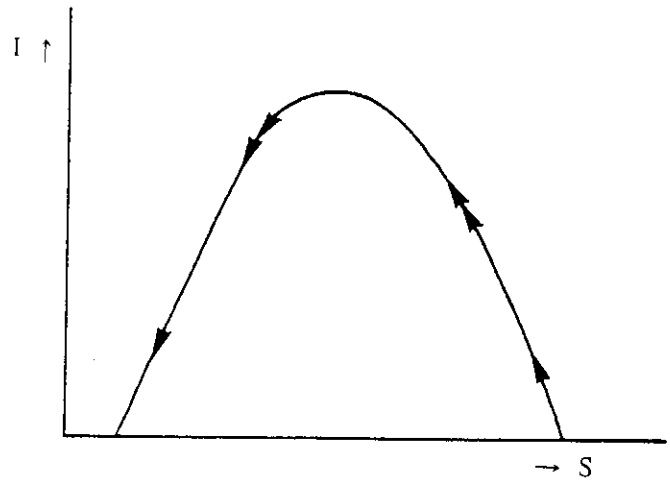


FIGURE 2

(iii) At the end of the rapid epidemic outbreak the disease agent may go extinct for lack of sufficient susceptibles (of course in the context of a model we need to take stochasticity into account to describe this phenomenon). Subsequently a new susceptible population is built up at the demographic time scale. When the susceptible population is well above threshold again, re-introduction of the disease agent (another chance event) triggers another epidemic. Etc.. Measles on Iceland are the standard example of such recurrent epidemic outbreaks at irregular intervals (see Cliff & Haggett (1988) and Bartlett (1960)).

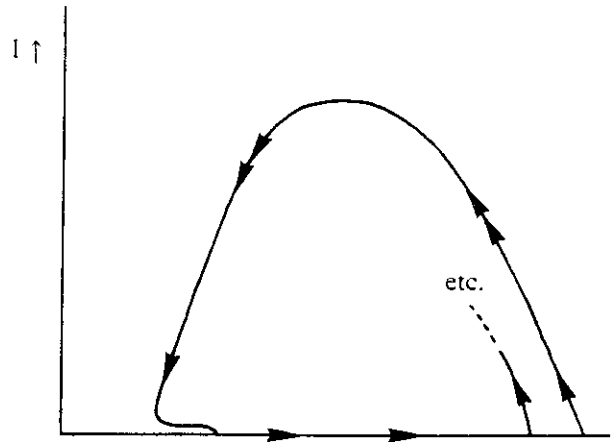


FIGURE 3

(iv) When a disease is continuously being transmitted at the demographic time scale it is called endemic. It is still an open problem to characterize the 'boundary' between the cases (iii) and (iv), for instance in terms of a critical community size or a sufficiently fast demographic turnover (see Bartlett (1960) and Schenzle & Dietz (1987)). A key question pertaining to the endemic situation is: should we expect a stable steady state or oscillations? How do the average prevalence and the period of the oscillations (if any) depend on relevant parameters? (see papers in Part III of Levin, Hallam & Gross (1989) and the references given there.)

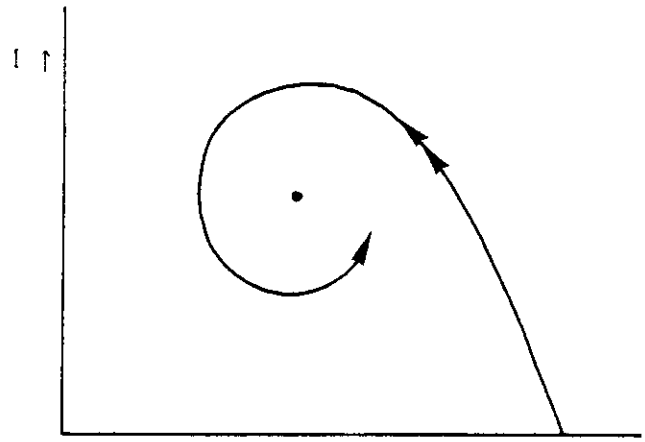


FIGURE 4

(v) How does a disease affect the growth rate of its host population? How do demographic processes and disease transmission interact?

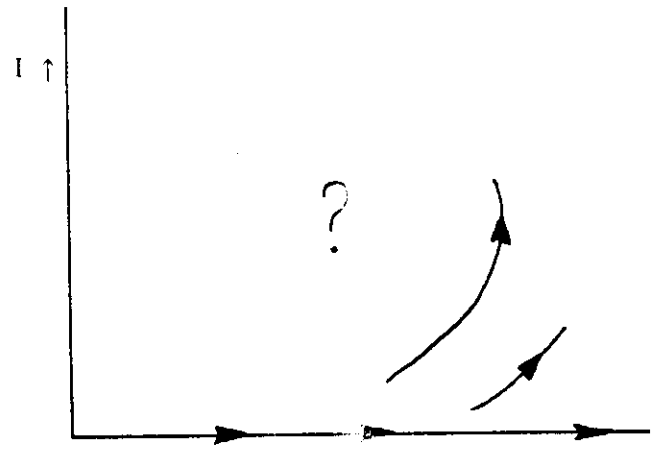


FIGURE 5

Typically one finds multiple threshold values for contact parameters. Bistability and large periodic outbreaks may occur. (see May, Anderson & McLean (1988), Diekmann & Kretzschmar (submitted) and references given there).

Note that in (i) and (ii) one ignores the demography, in (iii) demography and disease transmission are decoupled by a time scale argument, while in (iv) and (v) it is exactly the interaction of these which is the object of study (In (iv) the influence goes one way only, from demography to disease transmission, while in (v) it goes both ways.). What the appropriate question is depends partly on the relative magnitude of the various time scales and partly on our own interest: which time scale do *we* want to consider? (See the very readable introductory article 'Ecology in theory and application' by S.A. Levin in Levin, Hallam & Gross (1989).)

We refer to Diekmann, Heesterbeek, Kretzschmar & Metz (preprint 1989; to be elaborated into a book) for a more extensive survey of epidemic modelling in the style of this paper.

## 2. The invasion problem

In this and the following sections I report on results obtained in joint work with J.A.P. Heesterbeek and J.A.J. Metz.

We have introduced a function $A$ to describe the expected infectivity as a function of $d$-age $\tau$. But 'infectivity' is an ill-defined concept and a further specification may depend on the particular characteristics of the transmission mechanism. As a rule, however, the function $A$ combines information on the probability (per unit of time) that contacts take place and the probability that, given a contact, the disease agent is actually transmitted (in a sufficient quantity).

The probability that contacts take place may (or may not) depend on population density. It is common habit to write any such dependence explicitly as a factor multiplying $A$, and to incorporate in $A$ itself only information which does not depend on population density.

In 1927 W.O. Kermack and A.G. McKendrick proved that the introduction of some infectives into a (demographically closed) susceptible population of density $S$ triggers an epidemic when $R_0 > 1$, where, by definition

$$R_0 := S \int_0^\infty A(\tau) d\tau \tag{2.1}$$

In contrast, no epidemic occurs when $R_0 < 1$.

Two assumptions underly their result:

1) contacts between individuals occur according to the law of mass action, and
2) all individuals are equally *susceptible*

(recall that possible differences with respect to *infectivity* are incorporated in the definition of $A$ as an *expected* infectivity). Given these assumptions $R_0$ is the expected number of secondary cases per case and the result corresponds exactly to what common biological sense would tell.

Whenever 'satiation' reduces the tendency to make contacts, the law of mass action is not appropriate. In fact, for sexually transmitted diseases (STD's) and for diseases transmitted by mosquito's while taking blood meals, a more appropriate assumption is that the number of contacts an individual makes per unit of time is independent of population density (obviously such an assumption is violated at very low densities). In general one may assume a saturating functional response

$$\Lambda(S)$$

to describe phenomenologically the dependence of infectivity on population density. For instance, Dietz (1982) suggests taking the Holling function

$$\Lambda(S) = \frac{S}{1+cS}.$$

(Side-remark: In ecology the Holling function describes how the number of prey caught per predator per unit of time depends on prey density. The expression was derived by Holling from a mechanistic submodel for the predation process, assuming that the prey density is constant on the time scale of the predation process (or, in other words, that there are far more prey than predators). In the present context we need to model contacts *within* a population and the Holling argument breaks down. So the expression is purely phenomenological. In work in progress J.A.P. Heesterbeek and J.A.J. Metz seek to replace the Holling time scale argument by one appropriate for the epidemic situation in order to derive an expression for $\Lambda$ on the basis of mechanistic arguments.)

We conclude that relaxing the first assumption underlying the Kermack-McKendrick threshold theorem is a straightforward matter: one simply defines

$$R_0 = \Lambda(S) \int_0^\infty A(\tau) d\tau. \tag{2.2}$$

(Side-remark: When we want to take into account that 'contacts' may last a non-negligible period of time, matters become more complicated. For STD's such pair-formation and dissociation seems an

essential ingredient of the transmission mechanism. In joint work with K. Dietz and J.A.P. Heesterbeek we are developing methods to calculate $R_0$ for epidemic models that incorporate pair-formation.)

We now concentrate on relaxing the second assumption. When there are differences in both susceptibility and infectivity, it is no longer immediately clear how to take expectations, since the transmission process involves the product of susceptibility and infectivity. In order to proceed we introduce a variable $\xi$, which we call the *h-state* ($h$ for heterogeneity), and which takes values in a set $\Omega$, the $h$-state space. The $h$-state can be static or dynamic, discrete or continuous. Examples include age, sex, suffering from another STD, average sexual activity, spatial position. Next we introduce a function

$$A(\tau,\xi,\eta)$$

to describe the expected infectivity of an individual which was infected $\tau$ units of time ago while having $h$-state $\eta$, towards a susceptible of $h$-state $\xi$. Note that, by parametrizing with $h$-state at infection we avoid, for the time being, having to specify the dynamics of $h$-state. Thus we retain generality without too much notational burden. At a later stage we have to make submodels in order to specify $A$ on the basis of biological and or sociological assumptions concerning the disease and the contact process.

Apart from the function $A(\tau,\xi,\eta)$, which is the basic modelling ingredient, we need to specify

$$S(\xi)$$

the distribution of the population density $S$ with respect to $h$-state in the demographic steady state and, if mass action is not appropriate, the functional response map $\Lambda$. The questions I want to address are: given $A, S$ and $\Lambda$, how do we define $R_0$ and how do we compute $R_0$?

Fist of all we *linearize*. Since we did not yet write any equations the reader may wonder what we have in mind. In the present context linearization simply means that we ignore the fact that the density of susceptibles decreases due to the infection process. In our calculations we keep the density of susceptibles at the constant level $S$, rather than taking its time change, as the disease makes victims, into account.

Next, we ignore the real time development and concentrate, instead, on the associated *generation* process which evolves in discrete steps. More precisely we introduce the *next-generation operator* $K(S)$ defined by

$$(K(S)\phi)(\xi) = \Lambda(S)(\xi)\int_{\Omega}\int_{0}^{\infty} A(\tau,\xi,\eta)d\tau\,\phi(\eta)d\eta. \tag{2.3}$$

Here $\phi$ describes, for a certain generation, the distribution with respect to $h$-state at the moment infection occurred (so $\int_\Omega \phi(\eta)d\eta$ is the total number of cases in that generation) and $K(S)\phi$ does exactly the same for the next generation. Technically we consider $K(S)$ as a linear operator mapping $L_1(\Omega)$ into itself. The question is whether or not we can meaningfully summarize the properties of $K(S)$ in one number and if so, what that number is.

The operator norm of $K(S)$, i.e.

$$\|K(S)\| := \sup_{\|\phi\|=1} \|K(S)\phi\|$$

gives a worst case estimate: it yields the highest multiplication factor that can occur when going from one generation to the next. Since the distribution in the second generation will be different, the multiplication factor in going from the second to the third generation will, in general, be less. This point is nicely demonstrated by the matrix

$$\begin{pmatrix} 0 & a_{12} \\ a_{21} & 0 \end{pmatrix}$$

which describes host-vector transmission or male-female sexual transmission. The norm equals the

maximum of $a_{12}$ and $a_{21}$. But if we start, for instance, with the vector $\binom{1}{0}$ the multiplication factors will be, alternately, $a_{21}$ and $a_{12}$. So, on average over the generations, the multiplication factor equals $\sqrt{a_{21}a_{12}}$.

In general the average multiplication factor over many generations is

$$\rho_d := \lim_{m\to\infty} \|K(S)^m\|^{1/m}$$

the so-called spectral radius of the operator $K(S)$. Since $K(S)$ is a positive operator (note that the interpretation implies that $S \geqslant 0$ and $A \geqslant 0$ and that, consequently, $K(S)$ maps the cone of non-negative functions into itself) one has, as a rule, that

(i) $\rho_d$ is an eigenvalue and the corresponding eigenvector $\phi_d$ is non-negative,
(ii) $\rho_d$ is strictly dominant, i.e. all other elements of the spectrum of $K(S)$ are strictly less in absolute value,
(iii) for all positive $\phi$ one has that

$$K(S)^m\phi \sim c(\phi)\rho_d^m\phi_d \quad \text{for} \quad m\to\infty$$

where all dependence on $\phi$ is through the positive number $c(\phi)$.

In other words, transients may depend on the precise manner in which the disease entered into the population (as described by $\phi$), but after a while the distribution with respect to $h$-state of subsequent generations becomes fixed and the numbers increase each generation with a factor $\rho_d$. Therefore we now take as

DEFINITION. $R_0 = \rho_d$. In words: $R_0$ is the spectral radius/dominant eigenvalue of the next generation operator $K(S)$.

Note that here we run with the hare and hunt with the hounds: we linearize because only the initial phase has our interest and then we iterate very often since we don't want to be bothered by the precise details of how the infection was introduced. If it takes very many iterations indeed before one observes the stable distribution $\phi_d$ and the multiplication factor $R_0 = \rho_d$ for the linearized problem, it may be that the nonlinearity is already important. This certainly happens for large spatial domains (where travelling waves describe the spread of an epidemic; see van den Bosch, Metz & Diekmann (to appear) and the references given there) and probably for the spread of HIV in a population which is structured according to sexual activity (see Los Alamos Science *18* (1989)). In general it is determined by two factors:

i) the difference, in absolute value, between the dominant eigenvalue and the rest of the spectrum, and
ii) quantitative aspects of (ir)reducibility (in a manner of speaking: the 'distance' from reducibility).

In any case $R_0 < 1$ guarantees that no epidemic will occur upon introduction of the disease.

In order to be sure that *any* introduction leads to an epidemic when $R_0 > 1$, we of course need to assume that $K(S)$ is *irreducible*.

An endemic steady state $\hat{S}$ is (partially) characterized by the requirement that the corresponding $K(\hat{S})$ operator has dominant eigenvalue one.

When the spectral radius of $K(S)$ depends monotonically on $S$ the condition $R_0 < 1$ guarantees *global* stability of the disease free state (the presence of the disease will lower the number of susceptibles available). But for STD's one may consider

$$\Lambda(S)(\xi) = \frac{S(\xi)}{\int_\Omega c(\eta)S(\eta)d\eta}$$

and then one does not have monotonicity. In such cases subcritical bifurcation of an endemic steady state at the critical value $R_0=1$ is possible (see for an example Huang, Cooke & Castillo-Chavez (preprint) or the paper by Castillo-Chavez, Cooke, Huang & Levin in Castillo-Chavez (ed.) (1989)).

The linearized real time course is described by the integral equation

$$i(t,\xi) = \Lambda(S)(\xi)\int_{\Omega}\int_{0}^{\infty} A(\tau,\xi,\eta)i(t-\tau,\eta)d\tau\,d\eta \tag{2.4}$$

where $i(t,\xi)$ is the incidence rate (the rate at which new cases occur). To verify that (2.4) should hold one just has to 'read' it in terms of the biological interpretation of $i$, $\Lambda(S)$ and $A$. The 'Ansatz' $i(t,\xi)=e^{\lambda t}\psi(\xi)$ leads to the requirement that $\psi$ is an eigenvector corresponding to the eigenvalue one for the operator

$$(K_\lambda\phi)(\xi) = \Lambda(S)(\xi)\int_{\Omega}\int_{0}^{\infty} A(\tau,\xi,\eta)e^{-\lambda\tau}d\tau\phi(\eta)d\eta. \tag{2.5}$$

This then defines the *rate* $\lambda_d$ at which the incidence will grow during the initial phase of an epidemic. Monotonicity arguments yield the equivalence

$$R_0>1 \iff \lambda_d>0\,.$$

The nonlinear 'real time' equation is obtained by letting $S$ in (2.4) depend on time and performing the appropriate bookkeeping for $S(t)$. Note that it does not make sense to consider a nonlinear generation process, since infectives are confronted with the prevailing 'environment' (i.e. the available susceptibles) in real time, and not on a generation basis. Only when the environment is constant, as it is in the linear approximation, can we consider the associated generation process by itself.

This ends our motivated definition of $R_0$. The claim is now, that one can use the definition of $R_0$ as a tool to derive epidemiologically interesting conclusions. In order to substantiate this claim we have to deal with two aspects:

i) the specification of $A$ and $\Lambda$ for various structuring variables on the basis of submodels

ii) the actual computation of $R_0$.

In the next section we list some special assumptions on $A$ which allow us to reduce the last problem to a very simple one. In section 4 we indicate, by presenting several examples, how $A$ can be composed from more basic modelling ingredients. All of this, however, will remain quite abstract from the point of view of someone interested in, say, public health policy. And indeed, much further work is required before one can come up with the kind of graphs (or more sophisticated pictures) that can possibly catch the attention of public health authorities. The results presented below provide a basis for subsequent work with a more quantitative signature. The thesis of van den Bosch (1990) shows at the same time that it is not always trivial to apply applied mathematics and that a successful biological operationalization of abstract mathematical results is sometimes possible (the subject is plant pathology and he shows how results on the asymptotic speed of propagation of solutions of certain integral equations yield biological insight concerning the spatial spread of plant diseases; also see Diekmann (1986) and van den Bosch, Metz & Diekmann (to appear)). It is conceivable that the work described in this paper will have a similar follow-up.

## 3. Reduction

In special cases $R_0$ allows a characterization that is amenable to calculations. We start with a case where one can even find an explicit expression for $R_0$.

Suppose that both the infectivity of an infective and the susceptibility of a susceptible are possibly influenced by their $h$-state, but in an independent manner. More precisely, we shall assume that

$$\int_{0}^{\infty} A(\tau,\xi,\eta)d\tau = a(\xi)b(\eta) \tag{3.1}$$

and speak about *separable* infectivity and susceptibility. (When $b$ is a multiple of $a$ one can also speak of 'weighted random mixing' or 'proportionate mixing'). Biologically this corresponds to the situation in which the distribution (over the $h$-state space $\Omega$) of the 'offspring' (i.e. the ones who become infected) is *independent* of the state of the 'parent' (i.e. the one who transmits the infection). Mathematically it guarantees that the operator $K(S)$ has one-dimensional range. Hence there is only one candidate for an eigenvector and we find

$$R_0 = \int_\Omega b(\eta)\Lambda(S)(\eta)a(\eta)d\eta. \tag{3.2}$$

Next, suppose that individuals make extra contacts within their own group. We take

$$\int_0^\infty A(\tau,\xi,\eta)d\tau = a(\xi)b(\eta)+c(\xi)\delta(\xi-\eta) \tag{3.3}$$

where $\delta$ is Dirac's 'function'. Some straightforward manipulations lead from the eigenvalue problem $K(S)\phi=\rho\phi$ to the *characteristic equation*

$$\int_\Omega \frac{b(\eta)\Lambda(S)(\eta)a(\eta)}{\rho-c(\eta)\Lambda(S)(\eta)}\,d\eta = 1. \tag{3.4}$$

The left hand side defines a decreasing function of $\rho$ which tends to zero for $\rho\to\infty$. The largest real root $R_0$ is larger than one if and only if either

(i) $\quad c(\xi)\Lambda(S)(\xi)>1 \quad$ for some $\xi\in\Omega$

or, otherwise,

(ii) $\quad \int_\Omega \frac{b(\eta)\Lambda(S)(\eta)a(\eta)}{1-c(\eta)\Lambda(S)(\eta)}\,d\eta>1.$

We refer to Diekmann, Heesterbeek & Metz (1990) for a further interpretation of these conditions and references to the relevant literature.

We conclude this section with a multigroup version of separable infectivity and susceptibility. Let $\xi$ be of the from $(i,\xi_i)$, where $i$ can take the values $1,2,\ldots,n$ and $\xi_i$ takes values in $\Omega_i$ (so $\Omega=\bigcup_{i=1}^n \{i\}\times\Omega_i$).

Assume that

$$\int_0^\infty A(\tau,(i,\xi_i),(j,\xi_j))d\tau = a_i(\xi_i)b_{ij}(\xi_j). \tag{3.5}$$

This is a *conditional* independence assumption. Indeed, when we normalize $a_i$ such that

$$\int_{\Omega_i} a_i(\xi_i)\Lambda(S)(i,\xi_i)d\xi_i=1$$

the conditional (on the first component being $i$) probability density function for $h$-state at infection is independent of the $h$-state of the one who infects and given by $a_i(\cdot)\Lambda(S)(i,\cdot)$. Then

$$(K(S)\phi)(i,\xi_i) = \Lambda(S)(i,\xi_i)a_i(\xi_i)\sum_j \int_{\Omega_j} b_{ij}(\xi_j)\phi(j,\xi_j)d\xi_j$$

and we conclude that, in order to be an eigenvector, necessarily

$$\phi(i,\xi_i) = \sigma_i\Lambda(S)(i,\xi_i)a_i(\xi_i).$$

Substitution then yields that the vector $\sigma$ should be an eigenvector of the matrix $M$ with entries

$$m_{ij} = \int_{\Omega_j} b_{ij}(\xi_j)\Lambda(S)(j,\xi_j)a_j(\xi_j)d\xi_j. \tag{3.6}$$

In particular, $R_0$ is the dominant eigenvalue of the matrix $M$. Thus we have reduced the dimension of the eigenvalue problem that characterizes $R_0$ from $\infty$ to $n$.

The essential point is that the independence assumption (3.5) allows us to average out one component of the $h$-state. A similar procedure works when the first component can take infinitely many values, but then, of course, the reduced problem will still be infinite dimensional.

## 4. Structure in infectivity and susceptibility

To warm up we neglect heterogeneity and show how to compute $A(\tau)$ from an underlying stochastic model at the individual level. Consider a discrete finite $d$-state space. We label the $d$-state by $i=1,2,\ldots,n$. Let $b$ with $\sum_{i=1}^{n} b_i = 1$ denote the probability distribution for $d$-state at the moment immediately following infection. Let $\Sigma$ denote the $n \times n$-matrix of transition rates. Then $\Sigma$ has non-negative off-diagonal elements and $\Sigma$ is defective (i.e., its dominant eigenvalue is less than 0) due to death or removal (i.e. the definite loss of infectivity). Let $h$ denote the vector of infectivities associated with the various $d$-states. Then

$$A(\tau) = \langle h, e^{\Sigma\tau} b \rangle \tag{4.1}$$

and

$$\int_0^\infty A(\tau)d\tau = -\langle h, \Sigma^{-1} b \rangle. \tag{4.2}$$

As an example consider a herpes infection which, after a first infective period, may become latent, resurge with a certain probability per unit of time etc. (see for instance Tudor (1990)). Let 1 denote the infective state and 2 the latent state. Let $\sigma$ denote the rate of going from 1 to 2 and $\theta$ the rate of going from 2 to 1, while $\mu$ denotes the death rate. Then

$$\Sigma = \begin{bmatrix} -\sigma-\mu & \theta \\ \sigma & \theta-\mu \end{bmatrix}$$

$b = \binom{1}{0}$ and $h_2 = 0$. So in this case

$$\int_0^\infty A(\tau)d\tau = h_1 \frac{\theta+\mu}{\mu(\sigma+\theta+\mu)}, \tag{4.3}$$

which shows how $R_0$ depends on the recrudescence parameter $\theta$.

We now switch to heterogeneity. Let $\xi$ be a measure for sexual activity. We consider $\xi$ as static and adopt the proportionate mixing assumption with $a(\xi) \sim \xi$. To express that sexual desires can become satisfied we take

$$\Lambda(S)(\xi) = \frac{S(\xi)}{\int_\Omega \eta S(\eta)d\eta}. \tag{4.4}$$

Together these assumptions lead to (Dietz, 1980)

$$R_0 \sim \frac{\int_\Omega \eta^2 S(\eta)d\eta}{\int_\Omega \eta S(\eta)d\eta} = \text{mean} + \frac{\text{variance}}{\text{mean}} \tag{4.5}$$

which shows that the variance in sexual activity can contribute substantially to the value of $R_0$! As presented here the result applies to homosexual contacts but, using the conditional separability of section 3, it is easy to extend it to heterosexual contacts (see Diekmann, Heesterbeek & Metz (1990) for more details).

As a general 'decomposition' we can write

$$A(\tau,\xi,\eta) = h(\tau)\int_{\Omega} a(\xi,\theta)P(\tau,\theta,\eta)d\theta \tag{4.6}$$

where $h$ denotes the 'intrinsic' infectivity (say, the amount of virus shed) as a function of $d$-age $\tau$, and $P(\tau,\theta,\eta)$ denotes the probability that the $h$-state is $\theta$, given that the $h$-state was $\eta$ $\tau$ units of time ago, while $a(\xi,\theta)$ is a measure for the intensity of contacts between individuals of $h$-states $\xi$ and $\theta$. In the *deterministic* case $P$ is a $\delta$ 'function'. For instance, when $h$-state is static we have $P(\tau,\theta,\eta)=\delta(\theta-\eta)$ and when we use 'age' as $h$-state variable we have $P(\tau,\theta,\eta)=\delta(\theta-\eta-\tau)$. In the *stochastic* case we have to provide a submodel for the movement through $\Omega$. In Diekmann, Heesterbeek & Metz (1990) this is worked out in detail for an STD for which both infectivity and susceptibility are enhanced by some other STD (causing ulcers and the like that enlarge the probability that transmission takes place upon contact) that is in an endemic state. The state space $\Omega$ consists of two points (suffering from the other STD or not) and individuals jump back and forth between these points depending on the force of infection (for the other STD) and the curing rate. The paper also deals with the case where sexual activity is another component of the $h$-state. Exploiting multigroup separability an explicit expression for $R_0$ is derived.

## References


Los Alamos Science **18** (1989) Aids Research.

R.M. Anderson & R.M. May (1985,I). Vaccination and herd immunity to infectious diseases. Nature **318**: 323-329.

R.M. Anderson & R.M. May (1985,II). Helminth infections of humans: mathematical models, population dynamics and control. Adv. Parasitology **24**: 1-101.

N.T.J. Bailey (1975). The Mathematical Theory of Infectious Diseases and its Applications. Griffin, London.

N.T.J. Bailey (1982). The Biomathematics of Malaria. Griffin, London.

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

N.G. Becker (1989). Analysis of Infectious Disease Data. Chapman and Hall, London, New York.

C. Castillo-Chavez (Ed.) (1989). Mathematical and Statistical Approaches to AIDS Epidemiology. Springer LNiBM **83**, Berlin etc.

A.D. Cliff & P. Haggett (1988). Atlas of Disease Distributions. Blackwell, Oxford, New York.

O. Diekmann (1986). Dynamics in bio-mathematical perspective. In: M. Hazewinkel, J.K. Lenstra & L.G.L.T. Meertens (eds.). Mathematics and Computer Science II, CWI Monographs 4, North-Holland, Amsterdam etc..

O. Diekmann, J.A.P. Heesterbeek, M. Kretzschmar & J.A.J. Metz (preprint 1989). Building Blocks and Prototypes for Epidemic Models.

O. Diekmann, J.A.P. heesterbeek & 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.

O. Diekmann & M. Kretzschmar (submitted). Patterns in the effects of infectious diseases on population growth.

K. Dietz (1980). Models for vector-borne parasitic diseases. In C. Barigozzi (ed.). Vito Volterra Symposium on Mathematical Models in Biology. Springer LNiBM **39**: 264-277, Berlin etc.

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

J.A.M. van Druten, Th. de Boo & A.D. Plantiga (1986). Measles, mumps and rubella: control by vaccination. Develop. Biol. Standard. **65**: 53-63 (S. Karger, Basel).

J.C. Frauenthal (1980). Mathematical Modeling in Epidemiology. Springer, Berlin etc.

K.P. HADELER & K. DIETZ (1984). Population dynamics of killing parasites which reproduce in the host. J. Math. Biol. **21**: 45-65.

H.W. HETHCOTE (to appear). Current issues in epidemiological modeling.

W. HUANG, K.L. CAMOKE,C. CASTILLO-CHAVEZ (preprint). Stability and bifurcation for a multiple group model for the dynamics of HIV/AIDS transmission.

D.G. KENDALL (1956). Deterministic and stochastic epidemics in closed populations. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability (J. Neyman, ed.) IV: 149-165.

W.O. KERMACK & A.G. MCKENDRICK (1927, 1932, 1933). Contributions to the mathematical theory of epidemics. Proc. Roy. Soc. A **115**: 700-721, **138**: 55-83, **141**: 94-122.

E.G. KNOX (1980). Strategy for rubella vaccination. Int. J. Epid. **9**: 13-23.

M. KRETZSCHMAR (1989). Persistent solutions in a model for parasitic diseases. J. Math. Biol. **27**: 549-573.

S.A. LEVIN, T.G. HALLAM & L.J. GROSS (eds.) (1989). Applied Mathematical Ecology, Springer, Berlin etc..

D. LUDWIG (1974). Stochastic Population Theories. Springer LNiBM **3**, Berlin etc.

G.MACDONALD (1957). The Epidemiology and Control of Malaria, Oxford Univ. Press, London.

R.M. MAY, R.M. ANDERSON & A.R. MCLEAN (1988). Possible demographic consequences of HIV/AIDS Epidemics. I. Assuming HIV infection always leads to AIDS. Math. Biosc. **90**: 475-505.

R. ROSS (1909). The Prevention of Malaria, Murray, London.

D. SCHENZLE & K. DIETZ (1987). Critical population sizes for endemic virus transmission. In: Räumliche Persistenz und Diffusion von Krankheiten, Heidelberger Geographischen Arbeiten, Heft 83: 31-41.

D. TUDOR (1990). A deterministic model for herpes infections in human and animal populations. SIAM Review **32**: 136-139.

F. VAN DEN BOSCH (1990). The velocity of spatial population expansion. Ph.D. Thesis, Leiden.

F. VAN DEN BOSCH, J.A.J. METZ & O. DIEKMANN (to appear). The velocity of spatial population expansion. J. Math. Biol.