Statistica Neerlandica (1996) Vol. 50, nr. 1, pp. 89–110

The concept of R_0 in epidemic theory

J. A. P. Heesterbeek*

Agricultural Mathematics Group (GLW-DLO), P.O. Box 100, 6700 AC Wageningen, The Netherlands

K. Dietz

Institut für Medizinische Biometrie, Universität Tübingen, Westbahnhofstraße 55, D-72070 Tübingen, Germany

In epidemiology R_0 denotes the average number of secondary cases of an infectious disease that one case would generate in a completely susceptible population. This concept is among the foremost and most valuable ideas that mathematical thinking has brought to epidemic theory. In this contribution, we first review the historical development of R_0, from demography to epidemiology, proceed to give an exposition of the recently formalised theory to define and calculate R_0 for structured populations, return to the interaction of demography and epidemiology for an example of the use of the concept to study vaccination campaigns and finally we deal with statistical aspects of estimating R_0. In the appendix we discuss some issues of current attention.

Key Words & Phrases: epidemic theory, threshold behaviour, demography, structured populations.

1 Introduction

The birth of statistics and epidemic theory and the initial development of these fields, are closely linked (for overviews see e.g. GREENWOOD, 1948, IRWIN, 1963). Arthur Ransome wrote in 1868: "There is probably no more legitimate use of the instrument of statistics than its application to the study of epidemic diseases".

The basic reproduction ratio (or number) R_0 is one of the most important concepts in epidemic theory:

> *R_0 is the expected number of secondary cases produced by a typical infected individual during its entire infectious period, in a population consisting of susceptibles only.*

It is used to study a number of different problems in infectious disease epidemiology. First, the value of R_0 characterises the ability of an infectious organism to invade into a virgin population of susceptible individuals. The quantity shows threshold behaviour. More precisely, if $R_0 < 1$ the infection cannot establish itself, and the outbreak of disease will only involve a very small number of

* j.a.p.heesterbeek@sc.agro.nl

individuals. If $R_0 > 1$ the infection can, at least initially, invade the population and cause an epidemic in the sense that there is a positive probability of a large outbreak. Second, and more importantly, R_0 is a measure to gauge the amount of uniform effort needed for a given control measure to eradicate an infection that has already become established in the population (i.e., that is endemic), and it measures the amount of effort needed to prevent an epidemic.

In this paper, we initially trace the historic development of R_0 from demography to epidemiology and indicate the parallels in the theoretical frameworks to characterise it. We proceed to review the recent advances in the theory of defining, calculating and estimating R_0 for infectious diseases spreading in (mainly large) structured populations. In the appendix we give a flavour—admittedly biased by personal interests—of current research concerning the threshold concept in epidemiology.

2 Origins in demographic theory

The threshold quantity R_0 does not originate in epidemiology. The concept is first mentioned by Richard Böckh, the Director of the Statistical Office of Berlin in his statistical yearbook for 1884, published in 1886. He refers to 'totale Fortpflanzung' (total reproduction), by which he means the number of females born from one female during her entire reproductive life. Based on a life-table for females for the year 1879, he calculated that on average, 2.172 girls were born to one female in her reproductive period ranging from 14 to 53 years of age.

It was Dublin and Lotka (1925) and Kuczynski (1928) who, also in demographic context, formalised the calculation and introduced the notation R_0. They used the term 'net reproduction rate per generation'. In Lotka (1925), the term 'net fertility' is still used to denote R_0 and he correctly talks of it as a 'ratio'. See Samuelsen (1976) for a detailed account of early confusion concerning the term (rate versus ratio among other things).

Consider a large population. Let $\mathcal{F}_d(a)$ be the survival function, i.e. the probability for a new-born individual to survive at least to age a, and let $b(a)$ denote the average number of offspring that an individual will produce per unit of time at age a. The survival function is related to the age-specific mortality $\mu(a)$ via

$$\mathcal{F}_d(a) = \mathrm{e}^{-\int_0^a \mu(\alpha)\,\mathrm{d}\alpha}$$

The function $b(\cdot)\mathcal{F}_d(\cdot)$ is called the reproduction function. The expected future offspring of a new-born individual, R_0, is given by the zeroth moment (hence the index '0') of the reproduction function:

$$R_0 = \int_0^\infty b(a)\mathcal{F}_d(a)\,\mathrm{d}a \tag{1}$$

If $R_0 > 1$, so if on average each female contributes more than one female to the next

generation, the population will grow. This characterisation of population growth is based on the growth of subsequent generations of individuals.

One can also describe changes to population size in real-time. For this, let $n(t, a)$ be the age-distribution of the population at time t, i.e. $\int_{a_1}^{a_2} n(t, a)\,da$ is the number of individuals with ages between a_1 and a_2. Consistency requires that

$$n(t, 0) = \int_0^\infty b(a) n(t, a)\,da \tag{2}$$

McKendrick (1926) first showed that the age-distribution satisfies

$$\frac{\partial n}{\partial t}(t, a) + \frac{\partial n}{\partial a}(t, a) = -\mu(a) n(t, a) \tag{3}$$

with boundary condition (2). The equation was later rediscovered by von Foerster (1959). In a more general setting than we will consider here, one can interpret (3) as an abstract Cauchy problem on some appropriate function space, and use the theory of semigroups of operators on that space to characterise solutions (see Webb, 1985, Metz and Diekmann, 1986).

An alternative due to Lotka and Feller to the formulation (2), (3) is to describe the same situation by an integral equation for the total birth-rate $B(t)$, being the total number of newborns produced per unit of time evaluated at time t. Note that $B(t) = n(t, 0)$. We obtain the so-called renewal equation

$$B(t) = \int_0^\infty b(a) \mathscr{F}_d(a) B(t - a)\,da \tag{4}$$

where we have neglected the effects of a founder population. Feller (1941) showed that, under suitable weak conditions on $b(a)$ and $\mu(a)$, the solutions of (2), (3) converge to a stable distribution of individuals over age (normalised with $\bar{n}(0) = 1$)

$$\bar{n}(a) = \bar{n}(0) \mathscr{F}_d(a)\, e^{-ra} \tag{5a}$$

in the sense that asymptotically for $t \to \infty$

$$n(t, a) \sim const \cdot \bar{n}(a)\, e^{rt} \tag{5b}$$

where r is the "intrinsic (or natural) rate of increase" of the population, which is the unique real number satisfying the characteristic equation

$$1 = \int_0^\infty b(a) \mathscr{F}_d(a)\, e^{-ra}\,da \tag{6}$$

Condition (6) is easily obtained from (4) by substituting the Ansatz $B(t) = c\, e^{rt}$. One

sees that the following relation holds between 'generation' and 'real-time' growth

$$R_0 > 1 \Leftrightarrow r > 0 \tag{7}$$

We described two ways of characterising growth and decline of populations, the first based on discrete generations, calculating the next generation (of females) from the present one, and the second based on real-time population growth. We will see that in epidemic theory, both approaches occur when characterising growth and decline of the (sub)population infected with a given infectious disease. A disadvantage of the second is that it as a rule leads to *implicit* characterisations of the growth rate, and calculation of r therefore has to be done numerically. The two methods are equivalent in the sense that $R_0 > 1$ iff $r > 0$.

The approach discussed above is deterministic. More precisely, it is stochastic as far as processes are concerned that operate on the level of individuals, but we use a law-of-large numbers argument to describe the processes that result on the population level by deterministic equations. Of the stochastic approaches, we only mention the approximations based on the theory of branching processes (for example, JAGERS, 1975) because of the close connections to the theory of characterising R_0 for structured populations (section 4). In the theory of branching processes, the growth rate r is usually referred to as the *Malthusian parameter*.

3 Development of the concept in epidemic theory

In epidemiology the concept of R_0 was first used in a rudimentary version by Ronald Ross in 1911 in connection to his work on malaria. Ross received the Nobel prize for medicine in 1902 for his discovery that the *Plasmodium* species that causes (bird) malaria is transmitted from individual to individual by mosquitoes. Ross (1911) argued that local eradication of malaria was possible by decreasing the density of mosquitoes in the area. Prior to that it was generally believed that the malaria-parasite would always survive as long as some mosquitoes were still present, and that, since total eradication of mosquitoes was impossible, the disease could not be eradicated by mosquito control. Ross showed, using a simple model, that a critical mosquito density exists, below which malaria transmission cannot be maintained in the population. Empirical corroboration was later obtained in India with the discovery of neighbouring areas with and without malaria, and mosquito densities respectively above and below the critical level. No clear statement can be found in Ross (1911) that he interpreted this observation as the number of secondary cases being greater than or less than one. In other words, Ross did not connect his idea to a critical level of unity.

The next, and most fundamental, step in the development of R_0 in epidemiology, was taken in the seminal paper by Kermack and McKendrick in 1927. We will go into details of their definition of R_0 below. Neither Ross nor Kermack and McKendrick attached a name or a symbol to their threshold concept. This was first done by Macdonald, again in connection to malaria, who called it 'basic reproduction rate'

and denoted it by z_0 (MACDONALD, 1957). Dietz in 1975 and Anderson and May in the influential Dahlem Workshop Proceedings from 1982, popularised the name 'basic reproductive rate' and the symbol R_0 in epidemiology. Wide appreciation of the importance of R_0 was certainly increased by a paper of Anderson and May in *Nature* in 1979.

Up to the present day, only the symbol has been standardised in epidemic theory, but R_0 is still being called by many names. The most natural name would be 'basic reproduction ratio' or 'basic reproduction number' as has been pointed out many times before, since R_0 is a dimensionless quantity and certainly does not deserve the affix 'rate' which suggests a dimension 'time^{-1}'. Probably the cause of it all lies in Macdonald's naming of the quantity. As noted in section 2, the same confusion arose in the original demographic context. There, at least initially, Lotka showed understanding of the intricacy by referring to 'ratio' and 'rate per generation', but unfortunately the addition 'per generation' was quickly dropped for brevity.

McKendrick extended his work on age-structured populations to the description of the spread of infectious diseases when the infectiousness of an infected individual is not necessarily constant in time. A decade earlier, models like this had been studied by ROSS and HUDSON (1917), but not in this generality and without characterising threshold behaviour. Kermack and McKendrick make the following assumptions:

1. a single infection triggers an autonomous process within the host (i.e., they have viral and bacterial diseases in mind, where the agent generally multiplies so fast within the host that additional doses acquired later have little influence; see appendix A1 for remarks on diseases caused by parasitic worms);
2. the disease results in either complete immunity or death;
3. contacts are according to the law of mass-action (i.e., the idea borrowed from chemical reaction kinetics that says that the number of contacts between susceptible and infective individuals per unit time per unit area, is proportional to the product of the respective (spatial) densities); see appendices A4 and A5 for other options;
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.

Let $S(t)$ be the density of susceptibles in the population at time t (i.e., number of susceptibles per unit area, *not* probability density). The assumption 1 allows an age-representation for the state of the infection (infectivity) of an infected individual. The time elapsed since infection is called the infection-age. The above assumptions lead to the following integral equation

$$\dot{S}(t) = S(t) \int_0^\infty A(\tau)\dot{S}(t-\tau)\,\mathrm{d}\tau \tag{8}$$

where, by definition,

$$A(\tau) = \text{expected infectivity of an individual with infection-age } \tau$$

(cf. the renewal equation (4)). In analogy to the demographic context, we can interpret A as the reproduction function, where offspring produced has to be interpreted as new infections caused. In order to understand equation (8) one just has to realise that, by the closedness of the population, $-\dot{S}(t)$ is precisely the incidence $i(t,0)$, i.e. the density of new infecteds arising per unit of time, evaluated at time t (and so, $-\dot{S}(t-\tau) = i(t-\tau,0) = i(t,\tau)$ gives the incidence of infecteds that at time t have been infected for a time τ). We can reformulate (8) in terms of the incidence $i(t,\tau)$ of individuals that became infected at time $t-\tau$ by writing

$$\frac{\partial i}{\partial t} + \frac{\partial i}{\partial \tau} = 0 \tag{9}$$

$$i(t,0) = S(t) \int_0^\infty A(\tau) i(t-\tau,0)\, \mathrm{d}\tau \tag{10}$$

cf. system (2–3). The integral equation formulation is more convenient for our purpose. The expression $\Lambda(t) := \int_0^\infty A(\tau) i(t,\tau)\, \mathrm{d}\tau$ is called the *force of infection*. It is the per capita probability per unit of time to become infected.

Example: The function A is typically zero during the latency period, then rises to a maximum and declines to reach zero upon recovery. Various other shapes (e.g., a function which rises again from low values after the first maximum in the case of HIV) are encountered. Typically, the graph of A does *not* look like (a combination of) decreasing exponentials, but this is nevertheless what underlies ordinary differential equation models. If we choose

$$A(\tau) = \beta \mathrm{e}^{-\gamma\tau} \tag{11}$$

(8) reduces to an ordinary differential equation by calculating the total density of infectives present at time t, $I(t) := \int_{-\infty}^{t} \mathrm{e}^{-\gamma(t-\tau)} \dot{S}(\tau)\, \mathrm{d}\tau$ and differentiating. We obtain

$$\begin{aligned} \dot{S}(t) &= -\beta S(t) I(t) \\ \dot{I}(t) &= \beta S(t) I(t) - \gamma I(t) \end{aligned} \tag{12}$$

which is often called the Kermack–McKendrick system. This name is inaccurate for two reasons: first, it is only a very special case of their actual model (9), (10), and secondly, the system (12) can already be found in Ross and Hudson (1917) where differential mortality due to the infection is also taken into account; so, if anything, (12) should be called the Ross–Hudson system.

To characterise growth or decline of the infective population we have the same two routes as in the demographic theory. Since we wish to characterise R_0, we can

start by replacing $S(t)$ in (8) by the constant S_0, the density of the population at the start of the epidemic when every individual is susceptible. We take S_0 to be the demographic steady state population density in the absence of the infection. Our equations are now linear and like in the demographic theory, one can show that the linearised (8) has solutions of the form $-\dot{S}(t) = i(t,0) = c\,e^{\lambda t}$ with $\lambda > 0$ if and only if $R_0 > 1$ where

$$R_0 = S_0 \int_0^\infty A(\tau)\,d\tau \tag{13}$$

(cf. equation (!)). The real-time growth rate λ in the exponential phase of the epidemic is found as the unique real root of the characteristic equation obtained by substituting the 'Ansatz' into (8)

$$1 = S_0 \int_0^\infty A(\tau)\,e^{-\lambda\tau}\,d\tau \tag{14}$$

cf. equation (6). Since this generally is an implicit relation for λ we see the advantage of the generation approach—where we do get an explicit expression—over the real-time approach. This is even more so in the case of structured populations. Of course, one has to show that growth in generation-time corresponds to growth in real-time. In this case one has $R_0 > 1 \Leftrightarrow \lambda > 0$ by the positivity of A.

In the special case (11) we find from (13), $R_0 = \beta S_0/\gamma$, which we could have written down right away, since each infective makes on average βS_0 successful transmissions per unit of time, and does this for on average $1/\gamma$ time-units. In the same case we find from (12) that $\lambda = \beta S_0 - \gamma$. This is the only case where a direct relation between R_0 and the real-time growth rate λ exists, other than through the relation $R_0 - 1 > 0 \Leftrightarrow \lambda > 0$. In general, the ordering of R_0-values need not correspond to the ordering of λ-values (only in real-time does it matter whether one 'reproduces' early or late).

For recent reviews of theoretical developments see the papers in MOLLISON (1995).

Remark 1 *Stochastic basis of A.* One can either view A as a deterministic function describing infectivity, or as a true expectation. In METZ (1978), the stochastic basis of A is clarified, to highlight the fact that at the individual level there is stochasticity in the progression of the disease. Let the 'type' of an individual with respect to disease progression be x, taking values in some space X. One could let x be the total length of the infectious period. Let $a(\tau, x)$ be the infectivity of an individual with infection-age τ and type x, and let $m(\cdot)$ be some probability measure for the different possible progressions of the disease within hosts. Then

$$A(\tau) = \int_X a(\tau, x) m(\{dx\})$$

For example, if x denotes length of infectious period, (11) is obtained in the special case that $m(\{\mathrm{d}x\}) = f(x)\,\mathrm{d}x$ with $f(x) = \gamma \mathrm{e}^{-\gamma x}$ and $a(\tau, x) = \beta$ when $\tau \leq x$ and $a(\tau, x) = 0$ when $\tau > x$, and $X = \mathbb{R}^+$. This reflects the traditional approach that individuals are infective for an exponentially distributed period of time (with parameter γ) and have a constant infectivity β during that period.

Remark 2 *Stochastic models.* The application of branching processes to epidemic theory was stressed in NEYMAN and SCOTT (1964); see appendix A2. The characterisation of invasion thresholds in a stochastic setting was first studied by WHITTLE (1955); see also BALL (1995), BARBOUR (1994), and the contributions by F. G. Ball and W. S. Kendall to the discussion of MOLLISON, ISHAM and GRENFELL (1994). Under the assumptions listed above (apart from 6), let $s(t)$ and $i(t)$ be the fractions of the population susceptible and infective, respectively, and let N be the total population size. Whittle argues that the initial stages of the invasion, i.e. the process describing $Ni(t)$, can be approximated by a linear birth–death-process with birth rate $\beta s(0)$ and death rate γ. The following threshold result holds: if $\beta s(0) < \gamma$, the outbreak of infection will involve only a very small number of individuals; if $\beta s(0) > \gamma$, a large outbreak—$O(N)$ individuals—will occur with probability approximately $1 - (\beta s(0)/\gamma)^{-i(0)}$, and a small outbreak—$O(1)$ individuals—with probability approximately $(\beta s(0)/\gamma)^{-i(0)}$. Thus, even for $R_0 > 1$ the result can still be only a minor epidemic with positive probability. See BALL (1983) and METZ (1978) for details. NÅSELL (1995) also discusses threshold results for stochastic epidemic models.

4 R_0 for structured populations

Suppose that not all individuals are equally susceptible, but that certain traits (e.g., age, gender, genetic composition, whether or not one suffers from another disease) have a marked influence. Of course one then has to specify these traits, their dynamics and their frequency in the susceptible population.

Let the individual's trait be characterised by a variable ξ, taking values in some state space $\Omega \subset \mathbb{R}^m$. Let $S = S(\xi)$ denote the density function of susceptibles, describing the steady demographic state in the absence of the disease (we emphasise that $S(\cdot)$ is not a probability density function, its integral equals the total population size in the steady demographic state). Note that one now immediately has a valid question concerning the definition of R_0. It is possible to characterise the expected number of new cases of type $\xi \in \Omega$ caused by infectives of type $\eta \in \Omega$, along the lines of section 3, for any combination of traits (ξ, η). However, it is not immediately clear how the resulting numbers should be averaged. The aim is for the average to (i) have the original biological interpretation of R_0, and (ii) show the same threshold behaviour. The abstract methodology to solve this averaging problem for deterministic models is described in DIEKMANN, HEESTERBEEK and METZ (1990).

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

$$A(\tau,\xi,\eta) = \text{the expected infectivity of an individual that was infected } \tau \text{ units of time ago while having trait value } \eta \text{ towards a susceptible with trait value } \xi \tag{15}$$

then exactly the same reasoning which led to (8) yields, in the case of a closed population,

$$\frac{\partial S}{\partial t}(t,\xi) = S(t,\xi)\int_{\Omega}\int_{0}^{\infty} A(\tau,\xi,\eta)\frac{\partial S}{\partial t}(t-\tau,\eta)\,\mathrm{d}\tau\,\mathrm{d}\eta = -S(t,\xi)\Lambda(t,\xi) \tag{16}$$

where Λ is again the force of infection. Note that in the case of a single trait, taking just one value (16) collapses to (8). So the structure remains essentially the same as in (8), 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.

Instead of the real-time process (16), we consider the associated generation process as in sections 2 and 3. We regard generations of *infected* individuals. The first generation consists of the initial infected individuals in the population; the second generation consists of all infections caused by members of this first generation irrespective of their timing, etcetera. Let ϕ denote the distribution over the trait space Ω of the present generation. The next generation of infecteds with type ξ is then given by

$$\big(K(S)\phi\big)(\xi) = S(\xi)\int_{\Omega}\int_{0}^{\infty} A(\tau,\xi,\eta)\,\mathrm{d}\tau\,\phi(\eta)\,\mathrm{d}\eta \tag{17}$$

which tells us both how many secondary cases arise from the generation ϕ and how these new cases are distributed over Ω. We will call $K(S)$ the *next-generation operator*. In the special case $\Omega = \{1, \ldots, m\}$, $K(S)$ is an $m \times m$-matrix; in the case of a single trait taking only one value, (17) collapses to the Kermack–McKendrick value for R_0, equation (13).

Since S and A are non-negative, we usually interpret K as a positive operator on the Banach space $L_1(\Omega)$ of integrable functions. This has pleasant consequences. Let the spectral radius of K be denoted by $r(K)$. By the positivity of K, $r(K)$ is an eigenvalue of K (the dominant eigenvalue), and as a rule—under certain irreducibility conditions—one has for an initial generation ϕ, convergence to a steady distribution

$$K(S)^n\phi \sim c(\phi)r(K)^n\psi_d \quad \text{for} \quad n \to \infty \tag{18}$$

where ψ_d is the positive eigenfunction corresponding to $r(K)$, and $c(\phi)$ is a scalar which is positive whenever ϕ is non-negative and not identically zero. So, if we iterate the next-generation operator, the distribution of infected individuals over all trait

values stabilises to the form described by ψ_d, while numbers are multiplied by $r(K)$ from generation to generation. In other words, ψ_d describes the distribution of the 'typical' infected individual and $r(K)$ is the number of secondary cases. If we normalise $\|\psi_d\| = 1$, ψ_d is the probability distribution for the trait value at the moment of infection. Thus we are led to define

$$R_0 = r(K) = \text{the spectral radius/dominant eigenvalue of } K(S)$$

In general, the spectral radius of an infinite dimensional operator is difficult to compute explicitly. In one case, however, it is a triviality: when the operator has one-dimensional range. Biologically this corresponds to the situation where the distribution (over Ω) of the 'offspring' (i.e., the ones who become infected) is *independent* of the particular trait value of the 'parent' (i.e., the one who transmits the infection). Assume that

$$\int_0^\infty A(\tau, \xi, \eta)\, d\tau = a(\xi)b(\eta) \tag{19}$$

then one easily finds $\psi_d = Sa$ and

$$R_0 = r(K) = \int_\Omega b(\eta)S(\eta)a(\eta)\, d\eta \tag{20}$$

An obvious mathematical generalisation of this so-called separable mixing is to assume that $K(S)$ has a finite dimensional range. For this see DIEKMANN, HEESTERBEEK and METZ (1990).

To prove a threshold theorem that relates this R_0 defined on a generation basis to the development of the epidemic in real-time (both in the linearised version) is now no longer a triviality as in the homogeneous case. The basic idea is to show that the linearised real-time equation (16) has an exponentially growing solution with a positive growth-rate if and only if $R_0 > 1$. The linearised version of (16) 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 eigenfunction of the operator K_λ defined by

$$(K_\lambda \phi)(\xi) = S_0(\xi) \int_\Omega \int_0^\infty A(\tau, \xi, \eta)\, e^{-\lambda\tau}\, d\tau\, \phi(\eta)\, d\eta \tag{21}$$

Note that K_0 is the next-generation operator corresponding to the linearisation. One uses various positivity arguments to show from the family of operators defined by (21) that there exists a unique real $\lambda > 0$ characterising real-time growth if and only if $r(K_0) = R_0 > 1$. See HEESTERBEEK (1992) for details of the proof (where we note that the additional conditions given there for part of the proof are not sufficient, as was pointed out by Grimmer and Desch (pers. communication), but that this shortcoming is reparable).

5 Modelling A

The major modelling effort involved for a question pertaining to a specific host/infection system is to make precise how the infectivity kernel A depends on individual types, on infection-age, and to make precise how A changes as the type and infection-age of the infected individual change over time. The dependence of A of a given infected individual on the types of susceptibles is purely through the frequency of contacts (for example, if the type denotes gender, we could specify other contact rates for homosexual and heterosexual contacts). What is needed is to express the ideas/wishes about a given infection and the way it progresses in the host into a sub-model to try and derive a suitable infectivity kernel A. As is typical for structured population models, the process of obtaining A involves detailed stochastic modelling of events at the individual level, because on this level one often has possibilities to experimentally measure or estimate parameters. A law-of-large-numbers argument brings us back to the deterministic setting at the population level. Building these submodels is in general a difficult task. The expected infectivity is not only concerned with how infectious an infected individual with given type and infection-age is as a function of infection-age, the outcome in a way of the internal struggles with the immune system, but it is also important to which type of susceptibles he is infectious in what degree and how many contacts to these susceptibles occur per unit time.

A fruitful approach to determine A is to model the underlying changes in an infected individual's type by a Markov process on Ω. In $A(\tau, \xi, \eta)$, η refers to the type of an infected individual, say individual 'x', at the moment it contracted its own infection. Let $b(\xi, \theta)$ denote the infectivity and let $p(\tau, \theta, \eta)$ be the conditional probability that x is still alive at time τ and that the type of x is θ at that time, given that it was η at time 0. Then the expected infectivity towards ξs of x at time τ since infection is

$$A(\tau, \xi, \eta) = \int_{\Omega} b(\xi, \theta) p(\tau, \theta, \eta)\, \mathrm{d}\theta \tag{22}$$

and for $K(S)$ we find

$$(K(S)\phi)(\xi) = S(\xi) \int_{\Omega} \int_{0}^{\infty} \left(\int_{\Omega} b(\xi, \theta) p(\tau, \theta, \eta)\, \mathrm{d}\theta \right) \phi(\eta)\, \mathrm{d}\tau\, \mathrm{d}\eta \tag{23}$$

If type-value is static, $p(\tau, \theta, \eta) = \delta(\theta - \eta)$ (where δ denotes the Dirac-delta 'function') and we see that 'b' is equal to 'A'. One then sees the convenience of having parametrised A with the state *at birth* of the infecting individual, since the development of the theory of section 4 works for either case with the above interpretation. An algorithm for calculating R_0 for discrete-time epidemic models with $\Omega = \{1, \ldots, m\}$ and Markovian dynamics in type-change can be found in De Jong, Diekmann and Heesterbeek (1994).

For the remainder of this section we concentrate on age as trait. There are many reasons to incorporate age-structure in infectious disease models. For example, patterns of social behaviour and sexual activity correlate with age. The seriousness of many infectious diseases correlates with the infected's age and susceptibility and infectivity are often age-related. Data on the distribution of the random variable 'age at infection' contain information about the prevailing force of infection in an endemic situation.

We follow tradition and denote the types in this case by letters a and α. The stable age-distribution of susceptibles in the absence of infection is given by equation (5a) and now reads $S(a) = S(0)\,\mathrm{e}^{-ra}\mathscr{F}_d(a)$. Let $b(\tau, a, \alpha)$ be the average infectivity of an individual of age α and infection-age τ towards an individual of age a. Then

$$A(\tau, a, \alpha) = b(\tau, a, \alpha + \tau)\frac{\mathscr{F}_d(\alpha + \tau)}{\mathscr{F}_d(\alpha)}$$

where the quotient simply describes the probability that the individual has not died since becoming infected (we neglect death from disease-related causes). It makes sense to decompose b as

$$b(\tau, a, \alpha + \tau) = \pi(\tau, \alpha)c(a, \alpha + \tau)$$

Here c denotes the contact coefficient, i.e. for a population with age-density $n(a)$, the expected number of contacts per unit of time between individuals of ages a and $\hat{a}$ equals $c(a, \hat{a})n(a)n(\hat{a})$. The factor $\pi(\tau, \alpha)$ denotes the probability of transmission, given a contact between a susceptible and an infective, and given that the infected individual was itself infected τ time-units ago while having age α.

R_0 is the dominant eigenvalue of the next-generation operator

$$\big(K(S)\phi\big)(a) = S(0)\,\mathrm{e}^{-ra}\mathscr{F}_d(a)\int_0^\infty\int_0^\infty b(\tau, a, \alpha + \tau)\frac{\mathscr{F}_d(\alpha + \tau)}{\mathscr{F}_d(\alpha)}\,\phi(\alpha)\,\mathrm{d}\alpha\,\mathrm{d}\tau \tag{24}$$

If we separate the influence of the attributes of the infective and the type of the susceptible involved in the contact by assuming

$$b(\tau, a, \alpha) = f(a)g(\tau, \alpha)$$

then this leads to a one-dimensional range for K and we find

$$R_0 = S(0)\int_0^\infty\int_0^\infty g(\tau, \alpha + \tau)\mathscr{F}_d(\alpha + \tau)\,\mathrm{e}^{-r\alpha}f(\alpha)\,\mathrm{d}\alpha\,\mathrm{d}\tau \tag{25}$$

More generally, one can consider

$$b(\tau, a, \alpha) = \sum_{k=1}^{n} f_k(a)g_k(\tau, \alpha) \tag{26}$$

and find R_0 as the dominant eigenvalue of the $n \times n$-matrix M with entries

$$m_{ij} = S(0) \int_0^\infty \int_0^\infty g_i(\tau, \alpha + \tau) \mathcal{F}_d(\alpha + \tau) \, \mathrm{e}^{-r\alpha} f_j(\alpha) \, \mathrm{d}\alpha \, \mathrm{d}\tau$$

As a concrete example of such a "finite dimensional range" situation, decompose the age axis into n non-overlapping intervals $I_1, \ldots, I_n$, and assume that $c(a, \hat{a}) = c_{ij}$ for $a \in I_i$, $\hat{a} \in I_j$ (see DIEKMANN, 1995). Now let χ_I denote the characteristic function of a set I, and take $f_k = \chi_{I_k}$ and

$$g_k(\tau, \alpha) = \pi(\tau, \alpha) \sum_{j=1}^{n} c_{kj} \chi_{I_j}(\alpha + \tau)$$

For an application of reproduction ratios to the study of control methods, consider an infection that is endemic in a population with an age-distribution that is stable relative to the time-scale of disease transmission. In that case, the force of infection is constant over time, $\Lambda(t, \cdot) = \Lambda(\cdot)$ and

$$S(a) = S(0) \, \mathrm{e}^{-ra} \mathcal{F}_d(a) \mathcal{F}_i(a)$$

where $\mathcal{F}_i$ with

$$\mathcal{F}_i(a) = \mathrm{e}^{-\int_0^a \Lambda(\alpha) \, \mathrm{d}\alpha}$$

describes the probability that an individual is not infected before reaching age a, given that the individual will not die before reaching age a (i stands for infection).

The incidence of new cases is described by $-\frac{\partial S}{\partial t}$ (from equation (16)), which can be rewritten here as the consistency condition

$$\Lambda(a) = S(0) \int_0^\infty \int_0^\infty b(\tau, a, \alpha + \tau) \mathcal{F}_d(\alpha + \tau) \, \mathrm{e}^{-r\alpha} \mathcal{F}_i(\alpha) \Lambda(\alpha) \, \mathrm{d}\alpha \, \mathrm{d}\tau$$

This is a nonlinear integral equation for the unknown Λ. (Note that linearisation at $\Lambda \equiv 0$ and the transformation $\phi(a) \sim \mathrm{e}^{-ra} \mathcal{F}_d(a) \Lambda(a)$ lead us back to the eigenvalue problem for $K(S)$.)

If b satisfies relation (26), then necessarily $\Lambda(a) = \sum_k \psi_k f_k(a)$ with ψ an eigenvector corresponding to eigenvalue one of the matrix $M(\psi)$ with entries

$$m_{ij} = S(0) \int_0^\infty \int_0^\infty g_i(\tau, \alpha + \tau) \mathcal{F}_d(\alpha + \tau) \, \mathrm{e}^{-r\alpha} \, \mathrm{e}^{-\sum_k \psi_k \int_0^a f_k(\sigma) \, \mathrm{d}\sigma} f_j(\alpha) \, \mathrm{d}\alpha \, \mathrm{d}\tau$$

For the special case of n non-overlapping age-intervals, with $f_k = \chi_{I_k}$, one has that ψ_k is the force of infection for individuals in interval I_k and $M(\psi)$ can be interpreted as the so-called WAIFW-matrix ("Who Acquires Infection From Whom") introduced by Anderson and May as an approach to linking this theory to population data

(see ANDERSON and MAY, 1991, and GREENHALGH and DIETZ, 1994, for detailed treatment). One should realise that data (from e.g. serological studies or age-stratified case records) will at most allow the estimation of the force of infection vector ψ. These n values found from data do not allow one to reconstruct the n^2 entries of matrix $M(\psi)$. Additional assumptions on the structure of mixing between age-classes are needed. See ANDERSON and MAY (1991, chapter 9) and GREENHALGH and DIETZ (1994).

We finally look at the use of reproduction ratios in the evaluation of vaccination strategies. The idea is to calculate a reproduction ratio R_v for invasion into a population of susceptibles in a demographic steady state where a vaccination schedule v is in operation. Schedule v can lead to eradication of the infection if $R_v < 1$ can be accomplished. Let

$$\mathcal{F}_v(a)$$

denote a vaccination 'survival function', i.e. the conditional probability that an individual of age a that is alive and not yet infected, is susceptible (and not made immune by vaccination). Then

$$R_v = S(0) \int_0^\infty \int_0^\infty g(\tau, \alpha + \tau) \mathcal{F}_d(\alpha + \tau) \, \mathrm{e}^{-r\alpha} \mathcal{F}_v(\alpha) f(\alpha) \, \mathrm{d}\alpha \, \mathrm{d}\tau$$

and one can calculate whether or not a particular vaccination schedule, as described by $\mathcal{F}_v(a)$, leads to $R_v < 1$. Assuming among other things short disease duration one can simplify this formula to, DIETZ and SCHENZLE (1985),

$$R_v = \frac{\int_0^\infty f^2(\alpha) \mathcal{F}_d(\alpha) \, \mathrm{e}^{-r\alpha} \mathcal{F}_v(\alpha) \, \mathrm{d}\alpha}{\int_0^\infty f^2(\alpha) \mathcal{F}_d(\alpha) \, \mathrm{e}^{-r\alpha} \mathcal{F}_i(\alpha) \, \mathrm{d}\alpha}$$

As an example of the use of such a formula, suppose we vaccinate at age a_v. What is the minimum proportion q of the population to be covered to obtain $R_v \leq 1$? We put $R_v = 1$ and solve for the proportion q to find

$$q = \frac{\int_0^\infty f^2(\alpha) \mathcal{F}_d(\alpha) \, \mathrm{e}^{-r\alpha} \, \mathrm{d}\alpha - \int_0^\infty f^2(\alpha) \mathcal{F}_d(\alpha) \, \mathrm{e}^{-r\alpha} \mathcal{F}_i(\alpha) \, \mathrm{d}\alpha}{\int_{a_v}^\infty f^2(\alpha) \mathcal{F}_d(\alpha) \, \mathrm{e}^{-r\alpha} \, \mathrm{d}\alpha}$$

This expression can be evaluated by estimating the various ingredients from population data.

6 Statistical aspects of estimating R_0

A survey on problems of estimating R_0 is given by DIETZ (1993). Four approaches are discussed for estimating R_0:

—using data about individual parameters
—from the initial growth rate of the epidemic

—based on the final size of an epidemic
—using data from equilibrium situations

An attempt to estimate R_0 on the basis of the individual parameters is only meaningful for diseases where contacts are clearly defined such that they can be counted. This requirement excludes infectious diseases like measles or cholera where the spread is either airborne or due to contamination of food and/or water. Therefore, attempts to determine R_0 from individual parameters have been restricted to vectorborn infections based on estimates of the number of human blood meals which one vector takes per unit of time and the number of vectors contacting one human host per unit of time and to sexually transmitted diseases based on estimates of the number of new partners per person per unit of time and the number of contacts per partner. The most important diseases for which such attempts have been carried out are malaria as a representative of vectorborn infections and AIDS as an example for sexually transmitted diseases.

In rare situations where a new disease is introduced into a susceptible population one can observe the initial rate of growth of the epidemic and derive an estimate of R_0 from the rate of initial growth provided one has further information about the distribution of the infectious period. For a detailed application of this approach to AIDS see CAIRNS (1995).

Often it is not possible to observe the time dependent course of an epidemic. One is restricted to count the number of individuals infected during the epidemic and to estimate the final proportion of individuals still susceptible in the population. From this information one can calculate the initial proportion of susceptibles. The following formula provides an estimate of R_0 for a homogeneously mixing population:

$$R_0 = \frac{\ln u_0 - \ln u_\infty}{u_0 - u_\infty}$$

where u_0 denotes the initial and u_∞ the final proportion of susceptibles. It is important to note that one must not neglect that initially the population usually is not totally susceptible because this would lead to a severe underestimation of R_0.

In practice one is faced with the problem of estimating R_0 in equilibrium situations where the prevalence of the infection stays more or less constant over time. If the population is homogeneously mixing, the force of infection is identical for all ages and is proportional to $R_0 - 1$. This expression can then be used to estimate R_0 if one knows the force of infection and the constant of proportionality which equals the overall death-rate. If the contact matrix for different ages can be factored like in Equation (19) above then the estimation of R_0 only involves the estimation of the age-dependent force of infection on the basis of data that are all censored either from the left or from the right: For each individual it is only known whether for a given age the infection has not yet taken place (right-censored) or that the infection has taken place at some unknown time in the past (left censored). In a discussion paper to the Royal

Statistical Society, KEIDING (1991) presented a nonparametric method to estimate the age-specific prevalence and the corresponding smoothed force of infection. He applies this method to a data set for hepatitis A in Bulgaria and obtains $R_0 = 3.8$. The nonparametric method makes use of the left-continuous derivative of the convex minorant of a suitably defined empirical distribution function (GROENEBOOM and WELLNER, 1992). In general, however, as has been pointed out in the previous section, the contact rates do depend both on the age of the infective and on the age of the susceptible individual so that the information about the age specific force of infection is not sufficient to identify the contact matrix. This basic problem leads to intrinsic uncertainties about the magnitude of R_0 because different assumptions about the mixing matrix are compatible with the observed force of infection (GREENHALGH and DIETZ, 1994). Lower and upper bounds can be determined which may be sufficient for practical applications.

References

ADLER, F. R. and M. KRETZSCHMAR (1992), Aggregation and stability in parasite-host models, *Parasitology* **104**, 199–205.

ALTMANN, M. (1995), Susceptible-infected-removed epidemic models with dynamic partnerships, *Journal of Mathematical Biology* **33**, 661–675.

ANDERSON, R. M. and R. M. MAY (1979), Population biology of infectious diseases, part 1, *Nature* **280**, 361–367.

ANDERSON, R. M. and R. M. MAY (eds.) (1982), *Population biology of infectious diseases*, Springer-Verlag, Berlin.

ANDERSON, R. M. and R. M. MAY (1991), *Infectious diseases of humans, dynamics and control*, Oxford University Press.

BAILEY, N. J. T. (1975), *The mathematical theory of infectious diseases and its application*, Griffin, London.

BALL, F. G. (1983), The threshold behaviour of epidemic models, *Journal of Applied Probability* **20**, 227–241.

BALL, F. G. (1995), Coupling methods in epidemic theory, in: D. Mollison (ed.), *Epidemic models: their structure and relation to data*, Cambridge University Press, 34–52.

BALL, F. G. and P. J. DONNELLY (1995), Strong approximations for epidemic models, *Stochastic Processes and Applications* **55**, 1–21.

BARBOUR, A. D. (1994), Threshold phenomena in epidemic theory, in: F. P. Kelly (ed.), *Probability, statistics and optimization*, Wiley, Chichester.

BARTLETT, M. S. (1955), *An introduction to stochastic processes*, Cambridge University Press, London.

BECKER, N. G. and K. DIETZ (1995), The effect of the household distribution on transmission and control of highly infectious diseases, *Mathematical Biosciences* **127**, 207–219.

BÖCKH, R. (1886), *Statistisches Jahrbuch der Stadt Berlin*, Volume 12, Statistik des Jahres 1884, Berlin.

CAIRNS, A. (1995), Primary components for epidemic models, in: D. Mollison (ed.), *Epidemic models: their structure and relation to data*, Cambridge University Press, 350–371.

DE JONG, M. C. M., O. DIEKMANN and J. A. P. HEESTERBEEK (1994), The computation of R_0 for discrete-time epidemic models with dynamic heterogeneity, *Mathematical Biosciences* **119**, 97–114.

DIEKMANN, O. (1995), Age dependence, Unpublished lecture notes, CWI, Amsterdam.

DIEKMANN, O., M. C. M. DE JONG and J. A. J. METZ (1995), A deterministic epidemic model taking account of repeated contacts between the same individuals, CWI-report AM-R95.

DIEKMANN, O., K. DIETZ and J. A. P. HEESTERBEEK (1991), The basic reproduction ratio for sexually transmitted diseases, part 1: theoretical considerations, *Mathematical Biosciences* **107**, 325–339.

DIEKMANN, O., J. A. P. HEESTERBEEK and 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, *Journal of Mathematical Biology* **28**, 365–382.

DIETZ, K. (1975), Transmission and control of arbovirus diseases, in: D. Ludwig and K. L. Cooke (eds.), *Epidemiology*, SIAM, Philadelphia, 104–121.

DIETZ, K. (1989), The role of pair formation in the transmission dynamics of HIV, in: M. H. Gail and N. L. Johnson (eds.), *Sesquicentennial invited paper sessions*, Am. Statistical Association, Alexandria, VA, 609–621.

DIETZ, K. (1993), The estimation of the basic reproduction number for infectious diseases, *Statistical Methods in Medical Research* **2**, 23–41.

DIETZ, K. and K. P. HADELER (1988), Epidemiological models for sexually transmitted diseases, *Journal of Mathematical Biology* **26**, 1–25.

DIETZ, K., J. A. P. HEESTERBEEK and D. W. TUDOR (1993), The basic reproduction ratio for sexually transmitted diseases, part 2: The influence of variable HIV-infectivity, *Mathematical Biosciences* **117**, 35–47.

DIETZ, K. and D. SCHENZLE (1985), Proportionate mixing models for age-dependent infection transmission, *Journal of Mathematical Biology* **22**, 117–120.

DUBLIN, L. I. and A. J. LOTKA (1925), On the true rate of natural increase of a population, *Journal of the American Statistical Association* **20**, 305–339.

FELLER, W. (1941), On the integral equation of renewal theory, *The Annals of Mathematical Statistics* **12**, 243–267.

GREENHALGH, D. and K. DIETZ (1994), Some bounds on estimates for reproductive ratios derived from the age-specific force of infection, *Mathematical Biosciences* **124**, 9–57.

GREENWOOD, M. (1948), *Medical statistics from Graunt to Farr*, Cambridge University Press.

GROENEBOOM, P. and J. A. WELLNER (1992), *Information bounds and nonparametric likelihood estimation*, Birkhäuser, Basel.

HALLAM, T. G. and S. A. LEVIN (eds.) (1986), *Mathematical ecology, an introduction*, Biomathematics, vol. 17, Springer-Verlag, Berlin.

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

HEESTERBEEK, J. A. P. and M. G. ROBERTS (1995), Threshold quantities for helminth infections, *Journal of Mathematical Biology* **33**, 415–434.

IRWIN, J. O. (1963), The place of mathematics in medical and biological statistics, *Journal of the Royal Statistical Society A* **126**, 1–41.

JACQUEZ, J. A., C. P. SIMON and J. S. KOOPMAN (1995), Core groups and the R_0's for sub-groups in heterogeneous SIS and SI models, in: D. Mollison (ed.), *Epidemic models: their structure and relation to data*, Cambridge University Press, 279–301.

JAGERS, P. (1975), *Branching processes with biological applications*, Wiley, New York.

JAGERS, P. (1992), The deterministic evolution of general branching populations, Preprint 1992-25, Department of Mathematics, Chalmers University Göteborg, to appear in *Proceedings 3rd International Conference on Population Dynamics*, Pau, 1992, Wuerz Publ., Winnipeg.

JAGERS, P. and O. NERMAN (1984), The growth and composition of branching populations, *Advances in Applied Probability* **16**, 221–259.

JAGERS, P. and O. NERMAN (1985), Branching processes in periodically varying environment, *Annals of Probability* **13**, 254–268.

KEIDING, N. (1991), Age-specific incidence and prevalence: a statistical perspective, *Journal of the Royal Statistical Society A* **54**, 371–412.

KENDALL, D. G. (1956), Deterministic and stochastic epidemics in closed populations, in: *Proceedings 3rd Berekeley Symp. Math. Statist. Prob.*, Volume 4, 149–165.

KERMACK, W. O. and A. G. MCKENDRICK (1927), Contributions to the mathematical theory of epidemics, part I. *Proceedings of the Royal Society A* **115**, 700–721, Reprinted: *Bulletin of Mathematical Biology* (1991), **53**, 33–55.

KRETZSCHMAR, M., J. C. JAGER, D. P. REINKING, G. VAN ZESSEN and H. BROUWER (1994), The basic reproduction ratio R_0 for a sexually transmitted disease in a pair formation model with two types of pairs, *Mathematical Biosciences* **124**, 181–205.

KUCZYNSKI, R. R. (1928), *The balance of births and deaths*, Vol. 1, Macmillan Co., New York.

LOTKA, A. J. (1925), The measure of net fertility, *Journal of the Washington Academy of Sciences* **15**, 469–472.

MACDONALD, G. (1957), *The epidemiology and control of malaria*, Oxford University Press.

MACDONALD, G. (1965), The dynamics of helminth infections, with special reference of schistosomes, *Transactions of the Royal Society for Tropical Medicine* **59**, 489–506.

MCKENDRICK, A. G. (1926), Applications of mathematics to medical problems, *Proceedings of the Edinburgh Mathematical Society* **44**, 98–130.

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. and O. DIEKMANN (eds.) (1986), *The dynamics of physiologically structured populations*, Lecture Notes in Biomathematics, Vol. 68, Springer, Heidelberg.

MOLLISON, D. (ed.) (1995), *Epidemic models: their structure and relation to data*, Cambridge University Press.

MOLLISON, D., V. ISHAM and B. T. GRENFELL (1994), Epidemics: models and data, *Journal of the Royal Statistical Society A* **157**, 115–149.

NÅSELL, I. (1985), *Hybrid models of tropical infections*, Lecture Notes in Biomathematics, Vol. 59, Springer-Verlag, Berlin.

NÅSELL, I. (1995), The threshold concept in stochastic epidemic and endemic models, in: D. Mollison (ed.), *Epidemic models: their structure and relation to data*, Cambridge University Press, 71–83.

NEYMAN, J. and E. L. SCOTT (1964), A stochastic model of epidemics, in: J. Gurland (ed.), *Stochastic models in medicine and biology*, The University of Wisconsin Press, Madison, 45–83.

RANSOME, A. (1868), On epidemics, studied by means of statistics of disease, *British Medical Journal* **10**, 386–388.

ROBERTS, M. G. and J. A. P. HEESTERBEEK (1995), The dynamics of nematode infections of farmed ruminants, *Parasitology* **110**, 493–502.

ROBERTS, M. G., G. SMITH and B. GRENFELL (1995), Mathematical models for macroparasites of wildlife, in: B. T. Grenfell and A. P. Dobson (eds.), *Ecology of infectious diseases in natural populations*, Cambridge University Press, 177–208.

ROSS, R. (1911), *The prevention of malaria*, 2nd edition, J. Murray, London.

ROSS, R. and H. P. HUDSON (1917), An application of the theory of probabilities to the study of a priori pathometry, part III, *Proceedings of the Royal Society London A* **43**, 225–240.

SAMUELSON, P. A. (1976), Resolving a historical confusion in population analysis, in: D. Smith and N. Keyfitz (eds.), *Mathematical Demography*, Biomathematics, Vol. 6, Springer Verlag, Berlin, pp. 109–130.

SHURENKOV, V. M. (1992), On the relationship between spectral radii and Perron roots, preprint 1992-17, Chalmers University of Technology, Department of Mathematics (not submitted for publication).

VON FOERSTER, H. (1959), Some remarks on changing populations, in: F. Stohlman (ed.), *The kinetics of cellular proliferation*, Grune and Stratton, New York, 382–407.

WEBB, G. F. (1985), *Theory of nonlinear age-dependent population dynamics*, Marcel Dekker, New York.

WHITTLE, P. (1955), The outcome of a stochastic epidemic—a note on Bailey's paper, *Biometrica* **42**, 116–122.

YORKE, J. A., H. W. HETHCOTE and A. NOLD (1978), Dynamics and control of transmission of gonorrhea, *Sexually Transmitted Diseases* **5**, 51–56.

Appendices

A.1. *Parasitic worms.* If one regards infections caused by parasitic worms (helminths), assumption 1 (section 3) of Kermack and McKendrick no longer applies since as a rule, these disease agents do not multiply within the principal host. Disease is caused by a rise in parasite levels within the host and this rise must come from frequent re-infection. For helminth infections, 'infection-age' is not a relevant measure to gauge disease progress. Infectivity in the principal host is therefore not described by a function of infection-age, but by a function of the number of adult parasites that the host harbours. See ANDERSON and MAY (1991), and ROBERTS, SMITH and GRENFELL (1995) for recent reviews, and ADLER and KRETZSCHMAR (1992) for a comparison of different modelling approaches. The threshold concept is the expected number of (female) adult parasites produced per (female) adult parasite under optimal conditions (in the absence of density dependent constraints). In the simplest case, models are formulated in terms of mean parasite burdens—neglecting individual variation in parasite burden—and the threshold quantity for invasion can be easily derived, see ANDERSON and MAY (1991) (and the references given there). See HEESTERBEEK and ROBERTS (1995) for a more formal approach and ROBERTS and HEESTERBEEK (1995) for applications of this. For models that take individual variation in parasite burden into account, no general theory to characterise invasion yet exists, see BARBOUR (1994) for some interesting problems that can arise.

One has to be careful in characterising invasion thresholds for macroparasites where sexual reproduction of a female and a male adult is required, e.g. in schistosomes, before eggs can be produced to further the infection. Since invasion thresholds are characterised in a low (parasite) density limit, not every female parasite will be able to 'find' a mate (compare the Allee-effect in ecology, see e.g., HALLAM and LEVIN, 1986). Therefore, in the invasion-limit, the expected number of (female) adult parasites produced per (female) adult parasite is zero. This has been pointed out by NÅSELL (1985), and was first described by MACDONALD (1965) who used the effect in estimating the amount of control effort needed. If sexually reproducing parasites are picked up in clumps by the principal host from the environment (e.g., if the host actively acquires parasites by ingestion) the 'Macdonald-effect' may be avoided. However, in some cases parasites are acquired passively by the host (e.g., by skin penetration during watercontact for schistosomes) and typically not in clumps. An appropriate generalisation of the R_0-concept is not available for this type of infection.

A.2. *Branching processes.* Parameters similar to R_0 determine the asymptotic behaviour in branching processes with general state space. See JAGERS and NERMAN (1984). The approximation of epidemic spread by a branching process goes back to BARTLETT (1955) and to KENDALL (1956), who basically treats the Ross–Hudson model (12) with constant infectivity. See also NEYMAN and SCOTT (1964). The work of Kendall was generalised in METZ (1978) to the Kermack–McKendrick model with

variable infectivity; see also BALL (1983). In BALL and DONNELLY (1995), branching process approximations to stochastic epidemic models are treated extensively.

It has been pointed out by JAGERS (1992) and notably SHURENKOV (1992) that in Markov renewal theory, a more natural measure for generation growth, to relate to the real-time growth described by the Malthusian parameter, is the so-called Perron root. The Perron root $\rho(K)$ of our next generation operator is defined as

$$\frac{1}{\rho(K)} := \sup\left\{ t \in \mathbb{R}_+ : \sum_{n=0}^{\infty} t^n \tilde{A}^{(n)}(\xi, \cdot) \quad \text{is } \sigma\text{-finite for each } \xi \in \Omega \right\}$$

where $\tilde{A}(\xi, \eta) := \int_0^\infty A(\tau, \xi, \eta)\, d\tau$ where kernel-multiplication is defined in the usual way. The following relation holds: $\rho(K) \leq r(K)$, with equality if—among other things—Ω is compact (SHURENKOV, 1992). One could regard $R_0 = r(K)$ as a measure of global change on Ω, and $\rho(K)$ as a measure of more local change. In case of a non-compact trait space, the Perron root could lead to a 'better' description of epidemic spread than the spectral radius (which overestimates generation growth in that case), but the latter is often 'easier' to compute. One example where the Perron root could prove relevant is in certain models of helminth infections.

A.3. *Periodic environment.* Recently there is much interest in characterising invasion in a periodic environment. Consider the case $\Omega = \{1, \ldots, n\}$. The stability of the infection-free steady state of the real-time linearised system describing the infection dynamics, can be determined from the sign of the dominant Floquet multiplyer of the system. In HEESTERBEEK and ROBERTS (1995), this threshold quantity is described together with two others defined in terms of matrices related to the next-generation matrix K. Branching processes and the Malthusian parameter in periodically varying environments have been studied by JAGERS and NERMAN (1985).

A.4. *Sexually transmitted diseases.* For sexually transmitted infections, certainly in human populations, the practice that individuals form longer lasting partnerships can have a large influence on the spread of the infection (DIETZ and HADELER, 1988, DIETZ, 1989). If the formation and separation of partnerships is taken into account in a model for the spread of an infection, one can no longer meaningfully define the next-generation operator as above in terms of an infectivity function A that only depends on the infection-age. However, one can characterise and calculate R_0 for pair formation models of structured populations, by describing changes in infection-state and partnership state by a Markov process on a given state space. See DIETZ (1989) for the basic idea, DIEKMANN, DIETZ and HEESTERBEEK (1991) for the more general theory, and DIETZ, HEESTERBEEK and TUDOR (1993) for an application to the spread of HIV. In a model with different types of pairs (rather than different types of individuals)—e.g. long lived and short lived pairs—it can be shown that one can get different results when interpreting R_0 as referring to individuals or to pairs (see KRETZSCHMAR et al., 1994).

Recently there is much interest in the effects of concurrent partnerships. These effects are investigated with stochastic network models, see ALTMANN (1995) for details and references and a characterisation of R_0.

For applications to sexually transmitted diseases the idea of a small *core-group* of highly active individuals that can drive the epidemic in the other larger but less active groups is important (see YORKE, HETHCOTE and NOLDE, 1978, for initial development and JACQUEZ, SIMON and KOOPMAN, 1995, for a review and discussion of the relation to epidemic thresholds).

A.5. *Small contact structures.* BECKER and DIETZ (1995) derive explicit formulas for reproduction numbers of households and of individuals for arbitrary vaccination strategies in arbitrarily structured communities of households. The derivation of these formulas is based on two key ideas: Firstly, individuals and households are classified according to the number of susceptibles available at the time of infection either of the individual or of the household. Because of the usually small sizes of households, one cannot ignore the depletion of the number of susceptibles as one usually does in deriving formulas for R_0 in large populations. Secondly, it is assumed that the infection is highly infectious within a household. This assumption considerably simplifies the formulas because the final size distribution of the number of secondary cases within one household is simply given by the number of susceptibles in the household. For many diseases this is a realistic approximation. The threshold parameters are used to derive the levels of immunity required for the prevention of major epidemics in the community. For a community of households of equal size it is found that random vaccination of individuals is better than immunising all members of a corresponding fraction of households. When households have varying sizes, immunising all members of large households can be better than a corresponding vaccination coverage of randomly selected individuals.

A main point about the existence of small structures in a community like pairings, households or circles of acquaintances is that within these smaller structures repeated contacts of the same individuals cannot be neglected and the effect of these contacts, that infectious material is 'wasted' on individuals one has already infected before, has to be discounted in calculating both thresholds and final sizes. Recently, two approaches have been developed. In one, Ball, Mollison and Scalia-Tomba look at stochastic models for a population divided into n subgroups of size m. Instead of the regular approach to take m large, they take m to be small (household size) and consider branching process approximations and coupling methods with $n \to \infty$. They derive thresholds for invasion and final size distributions (see Ball's discussion of MOLLISON, ISHAM and GRENFELL, 1994).

In the second approach, DIEKMANN, DE JONG and METZ (1995) look at a community where every individual has a circle of k acquaintances to whom it has contacts and these k are a random sample from the (infinite) population. The major part of that paper is devoted to characterising final size distributions. They consider the following situation (the notation we use is as in remark 1). The infectious output

of an infected individual (normalised to equal the probability per unit of time of transmission) is assumed to be uniformly distributed among its k acquaintances. Let $a(\tau, x)$ be this output for an individual of 'type' x that was infected τ time-units ago. The expectation for the probability that a specific acquaintance of this particular infected individual is not yet infected by it at time τ is

$$F(\tau) = \int_X e^{-\frac{1}{k}\int_0^\tau a(\sigma,x)\,d\sigma} m(\{dx\})$$

The overall probability of transmission from the particular infective to one of its acquaintances during its entire infectious period is then defined by $Q := 1 - F(\infty)$. To calculate R_0 realise that during the invasion phase any newly infected individual has among its k acquaintances precisely one infected or immune acquaintance (namely the one it was infected by). Hence there are $k-1$ acquaintances left to infect and $R_0 = (k-1)Q$. In the special case described in remark 1, constant infectious output β and an exponentially distributed infectious period with exponent γ, they find

$$R_0 = \frac{\beta \frac{k-1}{k}}{\gamma + \beta/k}$$

One sees that for $k \to \infty$ they obtain the usual mass-action threshold quantity under the given assumptions, $R_0 = \beta/\gamma$. In the general case they recover the Kermack–McKendrick formula for $k \to \infty$

$$\lim_{k\to\infty} R_0 = \int_0^\infty A(\tau)\,d\tau$$

where $A(\tau)$ is as in remark 1.

Received: June 1995. Revised: August 1995.