J. Math. Biol. (1990) 28: 365–382

Journal of
Mathematical
Biology

On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations

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

[1] Centre for Mathematics and Computer Science, P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

[2] Institute of Theoretical Biology, Leiden University, Kaiserstraat 63, 2311 GP Leiden, The Netherlands

Abstract. The expected number of secondary cases produced by a *typical* infected individual during its entire period of infectiousness in a completely susceptible population is mathematically defined as the dominant eigenvalue of a positive linear operator. It is shown that in certain special cases one can easily compute or estimate this eigenvalue. Several examples involving various structuring variables like age, sexual disposition and activity are presented.

Key words: Epidemic models — Heterogeneous populations — Basic reproductive number — Invasion

1. Introduction

Suppose we want to know whether or not a contagious disease can "invade" into a population which is in a steady (at the time scale of disease transmission) demographic state with all individuals susceptible. To decide about this question we first of all *linearize*, i.e. we ignore the fact that the density of susceptibles decreases due to the infection process. It has become common practice in the analysis of the simplest models to consider next the associated *generation process* and to define the *basic reproduction ratio*[1] R_0 as the *expected number of secondary cases* produced, in a completely susceptible population, by a *typical* infected

[1] Quite often this is wrongly called the reproduction 'rate'. Many authors use 'reproductive' instead of 'reproduction', but, as was pointed out to us by M. Gyllenberg, the latter is grammatically more correct. We have followed the advice of I. Nåsell to use 'ratio' rather than 'number' in order to emphasize that R_0 does not even have a quasi-dimension ($R_0 \sim$ cases/case!)

individual during its entire period of infectiousness. The famous *threshold criterion* then states:

the disease can invade if $R_0 > 1$, whereas it cannot if $R_0 < 1$.

It is the aim of this note to demonstrate how these ideas extend to less simple (though probably still highly oversimplified) models involving *heterogeneity* in the population and to explain the meaning of "typical" in the "definition" of R_0 above. Subsequently we shall deal with the actual computation of R_0 in certain special cases, in particular the so-called "separable" or "weighted homogeneous mixing" case.

2. The definition

Let the individuals be characterized by a variable ξ, which we shall call the h-state variable (h for heterogeneity). Let $S = S(\xi)$ denote the density function of susceptibles describing the *steady demographic state in the absence of the disease* (in order to avoid confusion we emphasize that S is not a probability density function; that is, its integral equals the total population size in the steady demographic state, and not one). Define $A(\tau, \xi, \eta)$ to be the expected infectivity of an individual which was infected τ units of time ago, while having h-state η, towards a susceptible which has h-state ξ. The expected number of infections produced during its entire infective life by an individual which was itself infected while having h-state η is then given by

$$\int_\Omega S(\xi) \int_0^\infty A(\tau, \xi, \eta)\, d\tau\, d\xi,$$

where Ω denotes the h-state space, i.e. the domain of definition of ξ. We may call this quantity the next generation factor of η.

Remark. In order to have a unified notation for various cases we write integrals to denote sums whenever Ω is discrete (completely or just with respect to some component of ξ). A precise mathematical justification involves a dominant measure and Radon–Nikodym derivatives.

Since the new cases arise, in general, with h-states different from η, these numbers do not tell us exactly what happens under iteration, i.e. in subsequent generations (although it is clear that the supremum with respect to η yields an upper estimate for R_0).

So we abandon the idea of introducing an infected individual with a particular well-defined h-state and start instead with a "distributed" individual described by a density ϕ. The next-generation operator $K(S)$ defined by

$$(K(S)\phi)(\xi) = S(\xi) \int_\Omega \int_0^\infty A(\tau, \xi, \eta)\, d\tau\, \phi(\eta)\, d\eta \tag{2.1}$$

tells us both how many secondary cases arise from ϕ and how they are distributed over the h-state space. Ignoring the task of writing down conditions

on S and A which guarantee that $K(S)$ is a bounded operator on $L_1(\Omega)$, we note that the next generation factor of ϕ is simply the $L_1(\Omega)$-norm of $K(S)\phi$, i.e.,

$$\int_\Omega S(\xi) \int_\Omega \int_0^\infty A(\tau, \xi, \eta)\, d\tau\, \phi(\eta)\, d\eta\, d\xi$$

(note that we do not have to write absolute value signs since the biological interpretation requires all functions to be positive). If we take the supremum of the next generation factor over all ϕ with $\|\phi\| = 1$ we obtain, by definition, the operator norm of $K(S)$. This yields an upper estimate for R_0 for the same reason as above: the distribution with respect to ξ is changed in the next generation and consequently the factor of ϕ does not predict accurately what happens under iteration.

As a concrete example consider a host-vector model. (For the purpose of exposition we adopt here the strict version of the law of mass action even though this does not necessarily yield a good model in this case, see, e.g., Bailey (1982), Chap. 7.) Taking $\Omega = \{1, 2\}$ and $\int_0^\infty A(\tau, i, j)\, d\tau = a_{ij}$ with $a_{ij} > 0$ if and only if $i \neq j$ we find that $K(S)$ is represented by the matrix

$$\begin{pmatrix} 0 & a_{12}S_1 \\ a_{21}S_2 & 0 \end{pmatrix}$$

and the operator norm is $\max\{a_{12}S_1, a_{21}S_2\}$. These two numbers correspond to vector $\rightarrow$ host and host $\rightarrow$ vector transmission, respectively. No matter which of the two is the larger one, in the next generation it is necessarily the other of the two numbers which is the relevant *factor*.[2] Therefore the operator norm of $K(S)$ is not a good definition of R_0. Since $a_{12}S_1a_{21}S_2$ is the two-generation factor, the *average* next generation factor is

$$\sqrt{a_{12}S_1a_{21}S_2}.$$

How can we define such a quantity in general?

After m generations the magnitude of the infected population is (in the linear approximation) $K(S)^m\phi$ and consequently the per-generation growth factor is $\|K(S)^m\|^{1/m}$. We want to know what happens to the population in the long run, so we let $m \rightarrow \infty$. The so-called *spectral radius* (Schaefer, 1974) $r(K(S))$ is defined by

$$r(K(S)) = \inf_{m \geq 1} \|K(S)^m\|^{1/m} = \lim_{m \rightarrow \infty} \|K(S)^m\|^{1/m}. \tag{2.2}$$

Starting from the zeroth generation ϕ, the mth generation $K(S)^m\phi$ converges to zero for $m \rightarrow \infty$ if $r(K(S)) < 1$ whereas it can be made arbitrarily large by a suitable choice of ϕ and m when $r(K(S)) > 1$. Moreover, the *positivity* guarantees that in the latter case there is not really a restriction on ϕ. Indeed, $K(S)$ is a

[2] One could argue, as MacDonald did (see Bailey 1982, p. 100 and the references given there), that one should consider the average number of cases in the host population arising from one case in the host population via vector cases. From our point of view this amounts to looking *two* generations ahead. Indeed one obtains exactly MacDonald's result if one writes out $a_{12}S_1a_{21}S_2$ in terms of biting rates etc

positive operator (i.e. nonnegative functions are mapped onto nonnegative functions) and one can specify conditions on K (for example compactness) that guarantee that $r(K(S))$ is an eigenvalue (Schaefer 1960, 1974), which we shall call the *dominant eigenvalue* (since $|\lambda| \leqslant r(K(S))$ for all λ in the spectrum of $K(S)$) and denote by ϱ_d. Under minor technical conditions on A and S (see Remark 4 below) one has in addition that

$$K(S)^m\phi \sim c(\phi)\varrho_d^m \phi_d \quad \text{for } m \to \infty, \tag{2.3}$$

where ϕ_d is the corresponding eigenvector (which is positive) and $c(\phi)$ a scalar which is positive whenever ϕ is nonnegative and not identically zero. So after a certain period of transient behaviour each generation is (in an approximation which improves as time proceeds) ϱ_d times as big as the preceding one and distributed over h-state space as described by ϕ_d.

If we rephrase this as: "the *typical* number of secondary cases is ϱ_d", we are ready for the

Definition. $R_0 = r(K(S)) = \varrho_d =$ the dominant eigenvalue of $K(S)$.

With this definition the threshold criterion remains valid, as can be verified as follows. The threshold criterion relates the generation process to the development of the epidemic in real time, both in the linearized version. The linearized real time equation is

$$i(t, \xi) = S(\xi) \int_\Omega \int_0^\infty A(\tau, \xi, \eta) i(t - \tau, \eta)\, d\tau\, d\eta, \tag{2.4}$$

where $i(t, \xi)$ is the rate at which susceptibles with h-state ξ are infected at time t. This equation has a solution of the form $i(t, \xi) = e^{\lambda t}\psi(\xi)$ if and only if ψ is an eigenvector of the operator K_λ defined by

$$(K_\lambda \phi)(\xi) = S(\xi) \int_\Omega \int_0^\infty A(\tau, \xi, \eta)\, e^{-\lambda\tau}\, d\tau\, \phi(\eta)\, d\eta \tag{2.5}$$

with eigenvalue one. Positivity arguments can be used to show that among the set of such λ with *largest real part* there is a real one, which we shall denote by λ_d (and the corresponding eigenvector by ψ_d). Monotonicity arguments then imply that

$$\lambda_d > 0 \Leftrightarrow R_0 > 1 \quad \text{and} \quad \lambda_d < 0 \Leftrightarrow R_0 < 1.$$

(Heijmans 1986, Sects. 4–6, works this out in detail for a different but similar example and gives appropriate references. Hethcote and van Ark (1987) contains a proof for the finite dimensional case.)

Remarks.

(1) Whereas R_0 is a number, λ_d is a rate.

(2) Note that λ_d and ψ_d describe the growth and the h-state distribution in the exponential phase of an epidemic, when the influence of the precise manner in which the epidemic started has died off and the influence of the nonlinearity is not yet perceptible.

(3) In order to guarantee that *any* introduction of infectivity in the population leads to an epidemic when $R_0 > 1$ we need to make an *irreducibility* hypothesis (Schaefer 1974).

(4) Similar parameters determine the asymptotic behaviour in branching processes with general state space. See Jagers and Nerman (1984); Mode (1971).

(5) To obtain a complete model one has to specify the demographic processes, and in particular how per capita birth- and death rates are affected by the disease. If one makes the obvious assumption that the disease leads to a lower (or equal) birth rate and to a higher (or equal) death rate one can use the linearized problem to obtain upper estimates for the nonlinear problem. Thus one can prove, in general, *global* rather than local stability for $R_0 < 1$. Or, in other words, endemic states are impossible when $R_0 < 1$.

(6) Let $\hat{S}$ denote the susceptible population in a steady endemic state. Then necessarily $r(K(\hat{S})) = 1$. See Example 4.3.

(7) We have restricted our attention to the bilinear case. However, replacing $S(\xi)$ in the definition (2.1) of $K(S)$ by $h(S(\xi))$ or $S(\xi)/(1 + \int_\Omega S(\eta)\, d\eta)$ or something similar does not make any essential difference. See Examples 4.1 and 4.2 below. Note that for the invasion problem one will always have an expression involving the (known) function S only. Of course things are different if one wants to characterize endemic states, like in Remark 6 above.

3. Computational aspects: easy special cases

3.1. Separable mixing rate

To compute the dominant eigenvalue of a positive operator is, in general, not an easy task. However, there is one special case in which the task is trivial: when the operator has one-dimensional range. Biologically this corresponds to the situation in which the distribution (over the h-state space Ω) 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). In this case we speak of a *separable mixing rate* or separable infectivity and susceptibility, or (separably) weighted homogeneous mixing.

Assume that

$$\int_0^\infty A(\tau, \xi, \eta)\, d\tau = a(\xi)b(\eta) \tag{3.1}$$

then

$$(K(S)\phi)(\xi) = S(\xi)a(\xi) \int_\Omega b(\eta)\phi(\eta)\, d\eta. \tag{3.2}$$

So there can be but one eigenvector: $S(\xi)a(\xi)$. Since

$$K(S)Sa = \left(\int_\Omega b(\eta)S(\eta)a(\eta)\, d\eta \right) Sa$$

we conclude that

$$R_0 = \varrho_d = \int_\Omega b(\eta)S(\eta)a(\eta)\, d\eta. \tag{3.3}$$

Remarks.

(1) Note that assumption (3.1) is satisfied when

$$A(\tau, \xi, \eta) = a(\xi)B(\tau, \eta) \tag{3.4}$$

but that this is a slightly more restrictive requirement.

(2) A convenient normalization is

$$\int_\Omega S(\xi)a(\xi)\, d\xi = 1. \tag{3.5}$$

Then Sa is the probability density function for h-state at infection while $b(\xi)$ is the total expected number of "offspring" of an individual which was infected while having h-state ξ. This interpretation yields once more that

$$R_0 = \int_\Omega b(\eta)S(\eta)a(\eta)\, d\eta. \tag{3.6}$$

(3) The special case in which a and b differ only by a multiplicative constant, is usually referred to as proportionate mixing (Barbour 1978).

3.2. Separable mixing rate with enhanced infection within each group

A second case in which it is easy to derive an explicit threshold criterion, even if we cannot calculate R_0 explicitly, occurs when individuals preferentially mix with their own kind and otherwise practise weighted homogeneous mixing. If we moreover assume that the h-state stays constant over epidemiological time (but see Example 4.3) then $K(S)$ is of the form

$$(K(S)\phi)(\xi) = S(\xi)\left\{ a(\xi) \int_\Omega b(\eta)\phi(\eta)\, d\eta + c(\xi)\phi(\xi) \right\}, \tag{3.7}$$

where $c(\xi)S(\xi)$ is the number of first generation "offspring" produced "directly" in one's own group.

The eigenvalue problem $K(S)\phi = \varrho\phi$ can be rewritten as

$$\frac{1}{\varrho - c(\xi)S(\xi)} S(\xi)a(\xi) \int_\Omega b(\eta)\phi(\eta)\, d\eta = \phi(\xi). \tag{3.8}$$

Multiplying both sides by $b(\xi)$ and integrating over Ω we obtain the characteristic equation

$$\int_\Omega \frac{b(\xi)S(\xi)a(\xi)}{\varrho - c(\xi)S(\xi)}\, d\xi = 1. \tag{3.9}$$

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

$$\text{(i)} \qquad c(\xi)S(\xi) > 1 \quad \text{for some } \xi \in \Omega, \tag{3.10}$$

or, otherwise,

$$\text{(ii)} \qquad \int_\Omega \frac{b(\xi)S(\xi)a(\xi)}{1 - c(\xi)S(\xi)}\, d\xi > 1.$$

(Of course a more precise formulation of (i) is ess sup $c(\xi)S(\xi) > 1$.) When (i) holds a single just infected individual with h-state ξ will already start a full blown epidemic among its likes. If, on the other hand, $c(\xi)S(\xi) < 1$ for all $\xi \in \Omega$ any epidemic has to be kept going by the additional cross infections among different types. To understand (ii) we distinguish cross infections and direct infections within the own group and argue as follows. As before Sa is, with the normalization (3.5), the probability density function for h-state at *cross* infection. The expected total number of cases, including its own, produced by an individual of h-state ξ through chains of infectives which stay wholly among its likes is $(1 - S(\xi)c(\xi))^{-1}$. Each of these produces an expected number of cross infections equal to $b(\xi)$. So by treating the "clan's" as a kind of individuals we are back to our old separable mixing rate problem and we find

$$\int_\Omega b(\xi) \frac{1}{1 - S(\xi)c(\xi)} S(\xi)a(\xi)\, d\xi \tag{3.11}$$

as the expected offspring number at the clan level. An epidemic occurs if and only if this number exceeds one.

Remark. One of us had derived the result (3.10) in the context of the geographical spread of plant diseases (think of foci within fields). Recently our attention for this special case was revived by Andreasen and Christiansen (1989) (in which they derive the same result in the context of a finite h-state space) and Blythe and Castillo-Chavez (1989). Combinations like (3.7) can also be found in Nold (1980), Hethcote and Yorke (1984), Hyman and Stanley (1988) (where it is called biased mixing), and Jacquez et al. (1988) (where it is called preferred mixing).

3.3. Multigroup separable mixing

An obvious mathematical generalization of a separable mixing rate is to assume that $K(S)$ has a finite dimensional range. In general, however, this does not make biological sense. Therefore we restrict our elaboration to a special example in this category which does allow a biological interpretation.

Let ξ be of the form (i, ξ_i), where i can take the values $1, 2, \ldots, n$ and ξ_i takes values in Ω_i. So $\Omega = \bigcup_{i=1}^{n} \{i\} \times \Omega_i$. Assume that

$$\int_0^\infty A(\tau,(i, \xi_i), (j, \xi_j))\, d\tau = a_i(\xi_i) b_{ij}(\xi_j) \tag{3.12}$$

which one could call a *local* form of weighted homogeneous mixing, since, with the normalization

$$\int_\Omega a_i(\xi_i) S((i, \xi_i))\, d\xi_i = 1, \tag{3.13}$$

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 $a_i(\cdot) S((i, \cdot))$. Then

$$(K(S)\phi)(i, \xi_i) = S(i, \xi_i) a_i(\xi_i) \sum_j \int_{\Omega_j} b_{ij}(\xi_j) \phi(j, \xi_j)\, d\xi_j, \tag{3.14}$$

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

$$\phi(i, \xi_i) = \sigma_i S(i, \xi_i) a_i(\xi_i). \tag{3.15}$$

Substituting (3.15) into (3.14) we deduce that in addition the vector σ should be an eigenvector of the matrix M with entries

$$m_{ij} = \int_{\Omega_j} b_{ij}(\xi_j) S(j, \xi_j) a_j(\xi_j)\, d\xi_j. \tag{3.16}$$

In particular R_0 is the dominant eigenvalue of the matrix M.

4. Examples

4.1. Discrete and static h-state

In this case the operator $K(S)$ is represented by a matrix. We shall first show how this matrix can be derived in the special case of the conventional S-E-I-R compartment models.

Let M be the diagonal matrix of the per capita standard death rates of the various types. After infection individuals enter the exposed class E. From there they make the transition to the infective class I at a rate described by the diagonal matrix Σ where after they are removed at a rate described by the diagonal matrix D. Finally, let $T(S)$ be the *transmission matrix*, i.e. the matrix such that (beware: I denotes the vector of infectives, not the identity matrix)

$$\dot{E} = T(S)I - ME - \Sigma E.$$

We claim that

$$K(S) = T(S)\Sigma(\Sigma + M)^{-1}(D + M)^{-1}. \tag{4.1}$$

The (easy) argument goes as follows. The fraction of the exposed individuals which enters I (before dying) is the diagonal of $\Sigma(\Sigma + M)^{-1}$. The mean time of staying in I is the diagonal of $(D + M)^{-1}$. Within class I the transmission is described by $T(S)$.

If $M = 0$ the expression (4.1) simplifies to

$$K(S) = T(S)D^{-1}. \tag{4.2}$$

(In Sect. 9 of the paper by Jacquez et al. (1988), a special case of this matrix is introduced with $T(S)$ written out in some more detail.) Note that, as to be expected, Σ is irrelevant for the computation of R_0 in case $M = 0$ even though it may, of course, have substantial influence on the magnitude of λ_d.

In the separable case the entries of $T(S)$ are of the form

$$a_i S_i b_j,$$

and according to (3.6) R_0 equals the trace of the matrix $K(S)$. See Hethcote and Yorke (1984) for another derivation of this fact.

4.2. *Sexually transmitted diseases*

4.2.1. Heterosexual transmission only. Let the index 1 refer to females and the index 2 to males. For each sex we distinguish individuals according to some variable ξ_i which is static (the interpretation of ξ_1 may or may not be the same as the interpretation of ξ_2). Adopting the local separable mixing rate assumption and neglecting homosexual transmission we arrive at the matrix

$$M = \begin{pmatrix} 0 & m_{12} \\ m_{21} & 0 \end{pmatrix}, \tag{4.3}$$

where

$$\begin{aligned} m_{12} &= \int_{\Omega_2} b_{12}(\xi_2) S_2(\xi_2) a_2(\xi_2)\, d\xi_2, \\ m_{21} &= \int_{\Omega_1} b_{21}(\xi_1) S_1(\xi_1) a_1(\xi_1)\, d\xi_1. \end{aligned} \tag{4.4}$$

We conclude that

$$R_0 = \sqrt{m_{12} m_{21}}. \tag{4.5}$$

(See Hethcote and Yorke (1984) for a "discrete" version of this result.)

Distinguishing not only males and females, but on top of that hetero-, *bi*- and homosexuals one easily arrives at a six by six matrix whose spectral radius one has to compute to obtain R_0.

4.2.2. Sexual activity. Frequently the variables ξ_i are used to describe sexual activity (in the sense of: propensity to make sexual contacts), and a_i and b_{ji} are taken to be proportional to ξ_i. In the context of the heterosexual transmission

model above we would, more precisely, take

$$a_1(\xi_1) = \frac{\xi_1}{\int_{\Omega_2} \xi_2 S_2(\xi_2)\, d\xi_2}, \qquad (4.6)$$
$$b_{12}(\xi_2) = \beta_{12}\xi_2$$

(with formulas for a_2 and b_{21} obtained by interchanging 1's and 2's; one may argue that $\int_{\Omega_1} \xi_1 S_1(\xi_1)\, d\xi_1 = \int_{\Omega_2} \xi_2 S_2(\xi_2)\, d\xi_2$ is required if ξ is interpreted as the actual number of sexual contacts per unit of time). Thus one arrives at

$$m_{12} = \frac{\beta_{12} \int_{\Omega_2} \xi_2^2 S_2(\xi_2)\, d\xi_2}{\int_{\Omega_1} \xi_1 S_1(\xi_1)\, d\xi_1}$$

and

$$R_0 = \left[\beta_{12}\beta_{21} \frac{\int_{\Omega_1} \xi_1^2 S_1(\xi_1)\, d\xi_1}{\int_{\Omega_1} \xi_1 S_1(\xi_1)\, d\xi_1} \frac{\int_{\Omega_2} \xi_2^2 S_2(\xi_2)\, d\xi_2}{\int_{\Omega_2} \xi_2 S_2(\xi_2)\, d\xi_2} \right]^{1/2}. \qquad (4.7)$$

Recalling that

$$\frac{\int_{\Omega_i} \xi_i^2 S_i(\xi_i)\, d\xi_i}{\int_{\Omega_i} \xi_i S_i(\xi_i)\, d\xi_i} = \text{mean} + \frac{\text{variance}}{\text{mean}} \qquad (4.8)$$

we realize that this result is analogous to a result of Dietz (1980) and identical to formula (5.7) in May and Anderson (1988).

4.2.3. Two is worse than one. Stimulated by work of May and Jose (to appear), as reported in May and Anderson (1988), we now investigate how the presence of some sexually transmitted disease causing ulcers and the like may enhance the possibility of the successful invasion of another sexually transmitted disease like HIV.

Assume that disease d is in an endemic steady state. We want to calculate R_0 for a disease D, assuming that the susceptibility to D is, for individuals having d, v times as large as for individuals without d. What we have in mind is that encounter rates are totally random, but that the success ratio for disease transmission, given that contact takes place, is enlarged by a factor v. Then the proportions of 0 (:= free of d) and + (:= having d) individuals that will be infected by D in the linear initial phase of an epidemic are described by the vector

$$\begin{pmatrix} S_0 \\ vS_+ \end{pmatrix},$$

where S_0 and S_+ are the steady (with respect to d) state population sizes of 0 and + individuals. As we will show below, this vector indeed spans the range of the operator $K(S)$.

Let ζ denote the force of d-infection in the steady state and let γ denote the probability per unit of time that d is cured (whereupon susceptibility to d returns). The dynamics of d are completely described by these parameters ζ and γ. Let μ denote the natural death rate. Since

$$\frac{dS_+}{dt} = \zeta S_0 - \gamma S_+ - \mu S_+$$

we deduce that in steady state

$$\frac{S_+}{S_0} = \frac{\zeta}{\gamma + \mu} \quad \text{or} \quad S_0 = \frac{\gamma + \mu}{\gamma + \mu + \zeta} S, \qquad S_+ = \frac{\zeta}{\gamma + \mu + \zeta} S$$

where S denotes the total population size.

Any individual undergoes, as long as it does not die, transitions between 0 and + according to the matrix of rates

$$G = \begin{pmatrix} -\zeta & \gamma \\ \zeta & -\gamma \end{pmatrix}.$$

We describe the success ratios for D transmission by the matrix

$$p \begin{pmatrix} 1 & w \\ v & vw \end{pmatrix}.$$

Here p is the success ratio when both individuals involved in the contact are free of d and $w > 1$ is the factor by which the success ratio is enlarged when the D infectious individual is suffering from d. Note that we have to satisfy the requirement $pwv \leqslant 1$. We assume that encounters occur independently of the 0+ distinction and that the rate is given by σ/S (i.e., the number of contacts per unit of time is independent of the population size). Finally, we assume that D causes an extra death rate ϱ.

In order to write down $K(S)$ for this example we need an expression for the $A(\tau, i, j)$, $i, j \in \{0, +\}$. In contrast to the previous examples we now have a dynamic h-state. Without going into too much detail of how one could treat dynamic h-states in general, we explain some of the background of our calculations. Recall that $A(\tau, \xi, \eta)$ describes the infectivity, towards a susceptible with h-state ξ, of an individual, say x, that was itself infected τ time units ago while having h-state η. In the time interval $[0, \tau)$ passed since infection the h-state of x has changed to, say, state $\theta \in \Omega$. Assume that the infectivity towards an individual in h-state ξ then depends on θ, but not on the history of h-state transitions by which it reached θ nor on the time elapsed since infection τ (i.e., the influence of h-state on the, otherwise constant, infectivity is only through present h-state). Let $a(\xi, \theta)$ denote this infectivity. The expected infectivity towards ξ's of x at time τ after infection is then

$$A(\tau, \xi, \eta) = \int_\Omega a(\xi, \theta) P(\tau, \theta, \eta)\, d\theta,$$

where $P(\tau, \theta, \eta)$ is the conditional probability that the h-state of x at time τ is θ given that x is still alive at τ and that its h-state at time 0 was η. For $K(S)$ we find

$$(K(S)\phi)(\xi) = S(\xi) \int_\Omega \int_0^\infty \left[\int_\Omega a(\xi, \theta) P(\tau, \theta, \eta)\, d\theta \right] \phi(\eta)\, d\tau\, d\eta.$$

(Note that if the h-state is static, $P(\tau, \theta, \eta)$ is a δ-"function" in η and 'A' and 'a' are identical.) For $P(\tau, \theta, \eta)$ we have to calculate the probability that our individual x is still alive at time τ after infection and that its h-class is now θ. Let μ be a death rate that is independent of the h-state x is in. Then $P(\tau, \theta, \eta) = e^{-\mu\tau} P_1(\tau, \theta, \eta)$. (If there is no other mortality necessarily $\int_\Omega P_1(\tau, \theta, \eta)\, d\theta = 1$, while in general this integral is $\leqslant 1$.)

Let us look at the finite discrete case $\Omega = \{1, \ldots, n\}$. Then $K(S)$, $\mathfrak{a} = (a_{ij})_{1 \leqslant i,j \leqslant n}$, and $\mathfrak{P}_1(\tau) = (P_1(\tau, i, j))_{1 \leqslant i,j \leqslant n}$ are represented by $n \times n$ matrices. Let G describe the possibly, defective, transition probabilities per unit of time between all pairs of h-states (i, j) in Ω. Then $\mathfrak{P}_1(\tau) = e^{G\tau}$ and so $\mathfrak{P}(\tau) = e^{(G-\mu)\tau}$. If we substitute this into the equation for $K(S)$ we find

$$K(S) = \operatorname{diag}(S_1, \ldots, S_n) \int_0^\infty \mathfrak{a} e^{(G-\mu)\tau}\, d\tau.$$

If this operator has a one-dimensional range we can just as in the case of Eq. (2.1) find an explicit expression for R_0.

We now return to the two-disease example, where $\Omega = \{0, +\}$. From the considerations at the beginning of this section we find that

$$\mathfrak{a} = \begin{pmatrix} \dfrac{\sigma p}{S} & \dfrac{\sigma p w}{S} \\ \dfrac{\sigma p v}{S} & \dfrac{\sigma p v w}{S} \end{pmatrix}, \qquad \mathfrak{P} = \int_0^\infty \mathfrak{P}(\tau)\, d\tau = \int_0^\infty e^{(G-(\mu+\varrho))\tau}\, d\tau = (\mu + \varrho - G)^{-1}.$$

So

$$K(S) = \operatorname{diag}(S_0, S_+)\mathfrak{a}\mathfrak{P} = \frac{\sigma p}{S} \begin{pmatrix} S_0 & w S_0 \\ v S_+ & w v S_+ \end{pmatrix} \mathfrak{P}$$

with

$$\mathfrak{P} = \frac{1}{(\mu+\varrho)(\mu+\varrho+\gamma+\zeta)} \begin{pmatrix} \mu+\varrho+\gamma & \gamma \\ \zeta & \mu+\varrho+\zeta \end{pmatrix}.$$

Now note that matrix $\mathfrak{a}$ has a one-dimensional range spanned by $\begin{pmatrix} 1 \\ v \end{pmatrix}$. So the range of $K(S)$ is spanned by

$$\begin{pmatrix} S_0 \\ v S_+ \end{pmatrix} \sim \begin{pmatrix} \gamma + \mu \\ v\zeta \end{pmatrix}.$$

(Note that the crucial point is that we have a product wv in the success ratio in case of a contact between two '+' individuals, any other function of v and w does not correspond to separable mixing or, in other words, does not lead to a

one-dimensional range.) The only eigenvector of $K(S)$ is given by

$$\phi^* = \frac{\sigma p}{S}\begin{pmatrix} S_0 \\ vS_+ \end{pmatrix}.$$

By writing

$$K(S)\phi^* = \frac{\sigma p}{S}\begin{pmatrix} S_0 \\ vS_+ \end{pmatrix}\begin{pmatrix} 1 \\ w \end{pmatrix}^T \mathfrak{P}\phi^* = \phi^*\left[\begin{pmatrix} 1 \\ w \end{pmatrix}^T \mathfrak{P}\phi^*\right]$$

we find

$$R_0 = \frac{\sigma p}{S}\begin{pmatrix} 1 \\ w \end{pmatrix}^T \mathfrak{P}\begin{pmatrix} S_0 \\ vS_+ \end{pmatrix}$$

(where x^T denotes the transpose of a vector x) or, written out in detail,

$$R_0 = \frac{\sigma p\left\{(\gamma+\mu)(\mu+\varrho+\gamma+\zeta w)+\zeta v w\left(\frac{\gamma}{w}+\mu+\varrho+\zeta\right)\right\}}{(\gamma+\mu+\zeta)(\mu+\varrho)(\mu+\varrho+\gamma+\zeta)}.$$

Note that the special case $w = v = 1$ yields

$$R_0 = \frac{\sigma p}{\mu+\varrho}$$

as to be expected (since in that case the $0+$ distinction is totally irrelevant).

Next, let us consider the variant of this model in which the (fixed) sexual activity level figures as another component of the h-state. We now take, for a change, the sexual activity level as a discrete variable. The possible h-states are then $(i, 0)$ and $(i, +)$ with $i = 0, 1, 2, 3, \ldots$. Assuming

$$\frac{dS_{(i,+)}}{dt} = i\zeta S_{(i,0)} - \gamma S_{(i,+)} - \mu S_{(i,+)}$$

we find

$$S_{(i,0)} = \frac{\gamma+\mu}{\gamma+\mu+i\zeta} S_i,$$

$$S_{(i,+)} = \frac{i\zeta}{\gamma+\mu+i\zeta} S_i,$$

where S_i denotes the size of class i. Let

$$\phi_i := \begin{pmatrix} \phi_{(i,0)} \\ \phi_{(i,+)} \end{pmatrix}, \qquad \phi = (\phi_0, \phi_1, \ldots)^T.$$

The operator $K(S)$ is now represented by an infinite matrix acting on ϕ, an infinite sequence of two-vectors. As we assume a fixed sexual activity level for an individual, the h-state dynamics can be considered for each element ϕ_i of this sequence separately. We assume that for the ith two-vector these changes are governed by the matrix

$$G_i = \begin{pmatrix} -i\zeta & \gamma \\ i\zeta & -\gamma \end{pmatrix}.$$

The encounter rate of an $(i, \cdot)$ individual with a $(j, \cdot)$ individual is assumed to be

$$\frac{\sigma ij}{\sum_k kS_k}.$$

The success ratios for D transmission are equal to those in the case with no diversity in sexual activity level treated before, if we write $\mathfrak{P}_i = (\mu + \varrho - G_i)^{-1}$, $\mathfrak{P} := \operatorname{diag}(\mathfrak{P}_0, \mathfrak{P}_1, \ldots)$, and $a = (a_{ij})_{0 \leqslant i,j \leqslant \infty}$ with

$$a_{ij} = \frac{\sigma pij}{\sum_k kS_k} \begin{pmatrix} 1 & w \\ v & wv \end{pmatrix}$$

then $K(S)$ is the infinite matrix

$$K(S) = \operatorname{diag}(S_{(0,0)}, S_{(0,+)}, \ldots) a \mathfrak{P},$$

and

$$K(S)_{ij} = \frac{\sigma pij}{\sum_k kS_k} \begin{pmatrix} S_{(i,0)} & wS_{(i,0)} \\ vS_{(i,+)} & vwS_{(i,+)} \end{pmatrix} \mathfrak{P}_j.$$

So the ith element of the image $K(S)\phi$ of ϕ is given by the two-vector

$$(K(S)\phi)_i = \frac{\sigma pi}{\sum_k kS_k} \begin{pmatrix} S_{(i,0)} & wS_{(i,0)} \\ vS_{(i,+)} & vwS_{(i,+)} \end{pmatrix} \sum_{j=0}^{\infty} [j \mathfrak{P}_j \phi_j].$$

The range of $(K(S))_i$ is spanned by

$$\frac{\sigma pi}{\sum_k kS_k} \begin{pmatrix} S_{(i,0)} \\ vS_{(i,0)} \end{pmatrix} \sim \frac{ip_i}{\gamma + \mu + i\zeta} \begin{pmatrix} \gamma + \mu \\ iv\zeta \end{pmatrix}$$

where $p_i := S_i / \Sigma_k S_k$, the fraction of susceptible individuals with sexual activity level i. As before this leads directly to an expression for R_0,

$$R_0 = \frac{\sigma p}{\sum_k kS_k} \begin{pmatrix} 1 \\ w \end{pmatrix}^T \sum_j j^2 \mathfrak{P}_j \begin{pmatrix} S_{(j,0)} \\ vS_{(j,+)} \end{pmatrix}$$

$$= \frac{\sigma p}{\sum_k kp_k} \sum_j p_j j^2 \frac{(\gamma + \mu)(\mu + \varrho + \gamma + jw\zeta) + j\zeta vw\left(\frac{\gamma}{w} + \mu + \varrho + j\zeta\right)}{(\mu + \varrho)(\mu + \varrho + \gamma + j\zeta)(\gamma + \mu + j\zeta)}.$$

Formula (4.32) in May and Anderson (1988) is the analogue of this expression when one starts from (2.5) in this paper to find the initial growth rate λ_d.

4.3. Age dependence

We now turn our attention to a continuous dynamic h-state variable.

Let $\mathscr{F}(a)$ denote the survival probability as a function of age a in the absence of the disease. Then, at population dynamical equilibrium

$$S(a) = S(0)\mathscr{F}(a). \tag{4.9}$$

Let $\gamma(\tau, a, \alpha)$ be the average infectivity of an infected individual of age α and d-age τ towards a susceptible individual of age a. Then

$$A(\tau, a, \alpha) = \gamma(\tau, a, \alpha + \tau)\frac{\mathscr{F}(\alpha + \tau)}{\mathscr{F}(\alpha)} \tag{4.10}$$

and

$$(K(S)\phi)(a) = S(0)\mathscr{F}(a)\int_0^\infty \int_0^\infty \gamma(\tau, a, \alpha + \tau)\frac{\mathscr{F}(\alpha + \tau)}{\mathscr{F}(\alpha)}\phi(\alpha)\, d\alpha\, d\tau. \tag{4.11}$$

4.3.1. Separable mixing rate. Under the separable mixing rate assumption

$$\gamma(\tau, a, \alpha) = f(a)g(\tau, \alpha) \tag{4.12}$$

we find

$$R_0 = S(0)\int_0^\infty \int_0^\infty g(\tau, \alpha + \tau)\mathscr{F}(\alpha + \tau)f(\alpha)\, d\alpha\, d\tau. \tag{4.13}$$

4.3.2. Endemic steady states. Recalling Remark 6 at the end of Sect. 2 we shall now consider an endemic steady state. Let

$$\lambda(a) = \text{age specific force of infection,} \tag{4.14}$$

i.e. the age specific probability per unit of time of becoming infected. The survival function

$$\mathscr{F}_i(a) = e^{-\int_0^\alpha \lambda(\alpha)\, d\alpha} \tag{4.15}$$

describes the probability of being susceptible for those who did not die. Hence

$$\hat{S}(a) = \hat{S}(0)\mathscr{F}(a)\mathscr{F}_i(a) \tag{4.16}$$

where $\hat{S}$ describes the susceptible population in a steady *endemic* state. The age specific incidence rate is $\lambda(a)\hat{S}(a)$ and consistency now requires that

$$\begin{aligned}\lambda(a) &= \int_0^\infty \int_0^\infty A(\tau, a, \alpha)\lambda(\alpha)\hat{S}(\alpha)\, d\alpha\, d\tau \\ &= \hat{S}(0)\int_0^\infty \int_0^\infty \gamma(\tau, a, \alpha + \tau)\mathscr{F}(\alpha + \tau)\mathscr{F}_i(\alpha)\lambda(\alpha)\, d\alpha\, d\tau\end{aligned} \tag{4.17}$$

which can be considered as a nonlinear (recall (4.15)) integral equation for the (unknown) function λ. Note that linearization at the trivial solution $\lambda \equiv 0$ and

the transformation $\phi \to \mathcal{F}\lambda$ lead us back to the eigenvalue problem for $K(\hat{S})$, as to be expected. If we assume a separable mixing rate (4.12) we find that necessarily

$$\lambda(a) = Qf(a), \tag{4.18}$$

where the scalar Q has to satisfy

$$1 = \hat{S}(0) \int_0^\infty \int_0^\infty g(\tau, \alpha + \tau)\mathcal{F}(\alpha + \tau)\, e^{-Q\int_0^\alpha f(\sigma)\,d\sigma} f(\alpha)\, d\alpha\, d\tau. \tag{4.19}$$

4.3.3. Vaccination. Dietz and Schenzle (1985) consider the effect of vaccination and take

$$\hat{S}(a) = \hat{S}(0)\mathcal{F}(a)\mathcal{F}_v(a)\mathcal{F}_i(a), \tag{4.20}$$

where $\mathcal{F}_v(a)$ denotes the probability that an individual which did not die is immune due to vaccination. The analogue of (4.19) now is

$$1 = \hat{S}(0) \int_0^\infty \int_0^\infty g(\tau, \alpha + \tau)\mathcal{F}(\alpha + \tau)\mathcal{F}_v(\alpha)\, e^{-Q\int_0^\alpha f(\sigma)\,d\sigma} f(\alpha)\, d\alpha\, d\tau \tag{4.21}$$

which alternatively can be written as

$$1 = \hat{S}(0) \int_0^\infty \int_0^\infty g(\tau, \theta)\mathcal{F}(\theta)\mathcal{F}_v(\theta - \tau)\, e^{-Q\int_0^{\theta-\tau} f(\sigma)\,d\sigma} f(\theta - \tau)\, d\tau\, d\theta. \tag{4.22}$$

If we adopt the further assumption that

$$g(\tau, \alpha) = h(\alpha)k(\tau)\mathcal{F}_r(\tau), \tag{4.23}$$

where k describes the infectivity as a function of d-age and $\mathcal{F}_r$ the "removal" from the infected class, we finally arrive at

$$1 = \hat{S}(0) \int_0^\infty h(\theta)\mathcal{F}(\theta) \int_0^\theta k(\tau)\mathcal{F}_r(\tau)\mathcal{F}_v(\theta - \tau)\, e^{-Q\int_0^{\theta-\tau} f(\sigma)\,d\sigma} f(\theta - \tau)\, d\tau\, d\theta, \tag{4.24}$$

which is, apart from the notation, identical to formula (3) in Dietz and Schenzle (1985). These authors introduce yet two other simplifications:

(i) $h = f$, i.e. susceptibles and infectives have the same age dependence in activity level;

(ii) the duration of the disease is short on the time scale of ageing.

Then (4.24) can be approximated by

$$1 = C\hat{S}(0) \int_0^\infty f^2(\theta)\mathcal{F}(\theta)\mathcal{F}_v(\theta)\, e^{-Q\int_0^\theta f(\sigma)\,d\sigma} d\theta, \tag{4.25}$$

where C is a constant (describing the "magnitude" of the total infectivity). One can now use data about the endemic state to estimate f, Q and $\mathcal{F}$ and subsequently calculate whether or not a given $\mathcal{F}_v$ suffices to eradicate the disease. We refer once more to Dietz and Schenzle (1985) for some more details.

4.3.4. Separable mixing rate with enhanced within age group infection. To conclude this subsection we show how to compute the analogue of the threshold condition (3.10) (ii) in the case of age dependence (recall that in deriving (3.10) we assumed that the h-state is constant which it is not if we consider age). Assume that

$$\gamma(\tau, a, \alpha) = f(a)g(\tau, \alpha) + h(\tau, \alpha)\delta(a - \alpha), \tag{4.26}$$

where δ denotes Dirac's delta "function". Then

$$(K(S)\phi)(a) = S(a)\Big\{f(a)\int_0^\infty\int_0^\infty g(\tau, \alpha+\tau)\frac{\mathscr{F}(\alpha+\tau)}{\mathscr{F}(\alpha)}\phi(\alpha)\,d\alpha\,d\tau + \int_0^a h(\tau, a)\frac{\mathscr{F}(a)}{\mathscr{F}(a-\tau)}\phi(a-\tau)\,d\tau\Big\}. \tag{4.27}$$

We define an operator L by

$$(L\psi)(a) = S(a)\int_0^a h(a-\alpha, a)\frac{\mathscr{F}(a)}{\mathscr{F}(\alpha)}\psi(\alpha)\,d\alpha \tag{4.28}$$

and rewrite the eigenvalue problem $K(S)\phi = \varrho\phi$ as

$$\theta(\phi)Sf + L\phi = \varrho\phi, \tag{4.29}$$

where θ is the $\mathbb{C}$-valued mapping defined by

$$\theta(\psi) = \int_0^\infty\int_0^\infty g(\tau, \alpha+\tau)\frac{\mathscr{F}(\alpha+\tau)}{\mathscr{F}(\alpha)}\psi(\alpha)\,d\alpha\,d\tau. \tag{4.30}$$

For ϱ real and sufficiently large we can invert $\varrho I - L$. In fact one has the series expansion

$$(\varrho I - L)^{-1} = \sum_{n=0}^{\infty}\frac{L^{(n)}}{\varrho^{n+1}}. \tag{4.31}$$

Substituting $\phi = (\varrho I - L)^{-1}\theta(\phi)Sf$ in the definition of θ we find the characteristic equation

$$1 = \int_0^\infty\int_0^\infty g(\tau, \alpha+\tau)\frac{\mathscr{F}(\alpha+\tau)}{\mathscr{F}(\alpha)}((\varrho I - L)^{-1}Sf)(\alpha)\,d\alpha\,d\tau. \tag{4.32}$$

Assuming that (4.31) keeps converging up to $\varrho = 1$ (this is the analogue of the assumption $c(\xi)S(\xi) < 1$ for all $\xi \in \Omega$ in Sect. 3) we find that $R_0 > 1$ if and only if

$$\int_0^\infty\int_0^\infty g(\tau, \alpha+\tau)\frac{\mathscr{F}(\alpha+\tau)}{\mathscr{F}(\alpha)}\sum_{n=0}^{\infty}(L^{(n)}Sf)(\alpha)\,d\alpha\,d\tau > 1. \tag{4.33}$$

This condition allows an interpretation similar to the one of (3.11).

References

1. Andreasen, V., Christiansen, F. B.: Persistence of an infectious disease in a subdivided population. Math. Biosci. **96**, 239–253 (1989)

2. Bailey, N. T. J.: The biomathematics of malaria. London: Griffin 1982
3. Barbour, A. D.: MacDonald's model and the transmission of bilharzia. Trans. Roy. Soc. Trop. Med. Hyg. **72**, 6–15 (1978)
4. Blythe, S. P., Castillo-Chavez, C.: Like-with-like preference and sexual mixing models. Math. Biosci. **96**, 221–238 (1989)
5. Dietz, K.: Models for vector-borne parasitic diseases. In: Barigozzi, C. (ed.) Vito Volterra Symposium on Mathematical Models in Biology. (Lect. Notes Biomath., vol. 39, pp. 264–277) Berlin Heidelberg New York: Springer 1980
6. Dietz, K., Schenzle, D.: Proportionate mixing models for age-dependent infection transmission. J. Math. Biol. **22**, 117–120 (1985)
7. Heijmans, H. J. A. M.: The dynamical behaviour of the age-size distribution of a cell population. In: Metz, J. A. J., Diekmann, O. (eds.) The dynamics of physiologically structured populations. (Lect. Notes Biomath., vol. 68, pp. 185–202) Berlin Heidelberg New York: Springer 1986
8. Hethcote, H. W., Yorke, J. A.: Gonorrhea, transmission dynamics and control. (Lect. Notes Biomath., vol. 56) Berlin Heidelberg New York: Springer 1984
9. Hethcote, H. W., van Ark, J. W.: Epidemiological models for heterogeneous populations: proportionate mixing, parameter estimation, and immunization programs. Math. Biosci. **84**, 85–118 (1987)
10. Hyman, J. M., Stanley, E. A.: Using mathematical models to understand the AIDS epidemic. Math. Biosci. **90**, 415–473 (1988)
11. Jacquez, J. A., Simon, C. P., Koopman, J., Sattenspiel, L., Perry, T.: Modeling and analyzing HIV transmission: the effect of contact patterns. Math. Biosci. **92**, 119–199 (1988)
12. Jagers, P., Nerman O.: The growth and composition of branching populations. Adv. Appl. Probab. **16**, 221–259 (1984)
13. May, R. M., Anderson, R. M.: The transmission dynamics of human immunodeficiency virus (HIV). Philo. Trans. R. Soc. Lond., B **321**, 565–607 (1988)
14. Mode, C. J.: Multitype branching processes, theory and applications. New York: Elsevier 1971
15. Nold, A.: Heterogeneity in disease-transmission modeling. Math. Biosci. **52**, 227–240 (1980)
16. Schaefer, H. H.: Some spectral properties of positive linear operators. Pacific J. Math. **10**, 1009–1019 (1960)
17. Schaefer, H. H.: Banach lattices and positive operators. Berlin Heidelberg New York: Springer 1974

Received May 30, 1989; received in revised form November 7, 1989