The Basic Reproduction Ratio for Sexually Transmitted Diseases: I. Theoretical Considerations

O. DIEKMANN
Centrum voor Wiskunde en Informatica, 1098 SJ Amsterdam, The Netherlands, and Instituut voor Theoretische Biologie, 2311 GP Leiden, The Netherlands

K. DIETZ
Institut für Medizinische Biometrie, Universität Tübingen, D-7400 Tübingen 1, Germany

AND

J. A. P. HEESTERBEEK
Centrum voor Wiskunde en Informatica, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands

ABSTRACT

It is shown how one can calculate the basic reproduction ratio R_0 for infectious disease models where an arbitrary but finite number of disease states are recognized and where the phenomena of pair formation and separation are taken into account. Several examples are discussed.

1. INTRODUCTION

Biologically speaking, the basic reproduction ratio R_0 is the expected number of secondary cases caused by one typical infected individual during its entire period of infectiousness, in a population consisting of susceptibles only. Mathematically speaking, one investigates whether or not, starting from a few infected individuals, the disease can invade a susceptible population that is in its demographic steady state [3]. Because the initial number of infectious individuals is low, one can assume that every contact an infectious individual will make is necessarily with a susceptible and that during the initial phase the infectious process will not cause an appreciable decrease in the density of susceptibles. This makes the calculation of R_0 into a linear problem, and it is the reason its determination can be carried out allowing for arbitrary complexity in the description of the transmission dynamics. The main idea is to regard generations of infected individuals and to construct a certain operator that describes the transmission dynamics of

the disease as a discrete process relating subsequent generations. In [3] it was shown what this operator looks like and that R_0, as "defined" above, equals its spectral radius (in most, if not all, cases, R_0 can equivalently be characterized as the dominant eigenvalue of this "next-generation operator").

For sexually transmitted diseases it has been advocated by Dietz and Hadeler [6] that a "realistic" model should take into account the fact that individuals form partnerships for nonnegligible periods of time. During that time period the two partners have contacts only with each other, and in this way they are, momentarily, not in danger of receiving the infection from individuals outside the pair. In [3] the possibility of pair formation and separation was not incorporated. Two assumptions underlying the construction of the next-generation operator fail if we allow individuals to form pairs, and this entails that we cannot incorporate pair formation by a *direct* generalization of the next-generation operator. Implicitly it was assumed that every contact an infected individual has is with a "new" susceptible, which is by definition no longer true in the pair-formation case. Explicitly it was assumed that the only relevant "output" of an individual (i.e., what one has to know of an infected individual to determine its influence on the spread of the infection) was the *expected infectivity A* (where the average is taken over all possible sample paths of disease progress). As a consequence, an age representation for the expected infectivity status of an individual could be used. In other words, in a sufficiently large population, A can be considered as a deterministic function of τ, where τ measures the time that has passed since the individual became infected (for homogeneous populations this is the approach of [7]). In the case of pair formation a second output quantity, survival, comes into play. Of course, the survival of an infected individual is always important because it influences its infectivity and therefore needs to be incorporated in A. Within the context of pair-formation models, however, the survival of the partner has a second influence on the spread of infection: If your partner dies, you yourself become available for new contacts. For models incorporating pair formation, an equivalent age representation of disease state is not always possible.

However, we can make use of the ideas underlying the R_0 calculation in [3]. We can still construct an operator that describes the transmission dynamics as a discrete process on generations of infected individuals. In this paper we show what this operator looks like in the case of pair formation if we recognize an arbitrary but finite number of possible disease states $\{1, \ldots, n\}$ that are passed through sequentially, always starting in state 1. The disease state of an individual determines its current infectivity level and probability of dying. For the pair-formation processes we practically follow the simplest model described in [5] and [6]. The difference is that in [5] and [6] a pair starts, by definition, with a sexual contact. In this paper we take a

"period of courtship," in which the pair is not yet sexually active, into account.

In Section 2 we describe the model assumptions, explain the ingredients that are necessary for determining R_0, and give the algorithm for its calculation. The next-generation operator turns out to be an $m \times m$ matrix, where m is the number of disease states with positive infectivity, and R_0 will be the dominant eigenvalue of this matrix. These results are generalized in Section 3 to include arbitrary heterogeneity (in susceptibility) among the individuals in the population. As an example we treat the characteristics male/female. In section 4 we briefly consider the results of various limit procedures applied to the present pair-formation model. Among other things it is shown how the appropriate models that neglect pair formation can be obtained as limiting cases.

This paper treats a pair-formation analog of a multistage variable infectiousness model developed by Blythe and Anderson [1] and is therefore a generalization of the results in [5], where the cases of one and two disease states were considered. Part II (in preparation) will be concerned with quantitative results.

2. DESCRIPTION OF THE MODEL AND CALCULATION OF R_0

In our model we distinguish two classes of pairs: those who are in a courtship period, and those who are in the sexually active phase. The courtship period is characterized by the absence of sexual contacts. Individuals in a courtship are, just as single individuals, not at risk of either receiving or transmitting the infection. Individuals have sexual contacts only in the sexually active phase of a partnership.

In addition to the courtship label, we recognize the following characteristics of an *infected* individual:

$$\text{Disease state:} \qquad i \in \{1, \ldots, n\}$$

$$\text{Partnership state:} \qquad j \in \{-1, 0, 1, \ldots, n\}.$$

Here -1 signifies that the individual is single (no partner at the moment of observation); 0, that the individual is paired with a susceptible; and $j \in \{1, \ldots, n\}$, that the individual is paired with an infected individual who has disease state j. Together the two characteristics determine the *type* (i, j) of an infected individual. Types of individuals in a courtship period are indicated as $(i, j)_c$. For the moment we assume that all individuals are equally susceptible (so we disregard any heterogeneity other than the single-pair dichotomy).

As we are interested only in R_0, we assume that every new partner of a single infected individual is necessarily susceptible. The consequence of this

is that the only courtship types that we have to consider are $(i,0)_c$, for $i \in \{1, \ldots, n\}$, where the infected individual has a susceptible partner. We assume that all pairs start with a courtship phase.

Let $\Lambda := \{(i,j) \colon 1 \leqslant i \leqslant n, \ -1 \leqslant j \leqslant n\} \cup \{(i,0)_c \colon 1 \leqslant i \leqslant n\}$ be the set of possible types. $|\Lambda| = n(n+3)$, and consequently our type space is $\mathbb{R}^{n(n+3)}$. Let $\Sigma := \{1, 2, \ldots, n(n+3)\}$. We will call Σ the state space of infectives, and the elements of Σ are called *states*. Let $L \colon \Lambda \to \Sigma$ describe the lexicographic ordering on Λ, with the side condition that a courtship type precedes the corresponding sexually active type, that is,

$$L(i,j) < L(i',j') \Leftrightarrow \{i < i'\} \vee \{i = i', j < j'\},$$
$$L(i,-1) < L(i,0)_c < L(i,0), \qquad 1 \leqslant i \leqslant n.$$

We make the following assumptions:

(1) The disease states are passed through in natural order. In particular, a freshly infected individual starts its life (in disease sense) with type $(1,j)$ for some $j \in \{1, \ldots, n\}$.

(2) Given that the infected individual does not die, the time that its disease state is i is exponentially distributed with parameter θ_i (where $\theta_n = 0$; i.e., disease state n is retained until death).

(3) The infectivity in disease state i is described by the probability $p_i \geqslant 0$ that a sexual contact with a susceptible leads to transmission.

(4) μ_0 is the death rate of susceptibles; μ_i is the death rate of an infected individual with disease state i.

(5) Every single individual has a constant probability ρ per unit of time of becoming a member of a courtship pair. The divorce rate is σ_c in the courtship phase and σ in the sexually active phase.

(6) A pair always starts with a courtship phase. The length of this phase is, conditional on the survival and no divorce of the two partners, exponentially distributed with parameter α. By definition the sexually active phase starts with one sexual contact.

(7) During the sexually active phase, the partners have β sexual contacts per unit of time.

The graph in Figure 1 traces the possible changes in the type of an infected individual of fixed disease state i, as long as it does not die. Note that the only two types of individuals that can *cause* an infection are $(i,0)_c$ (first contacts), and $(i,0)$, $i \in \{1, \ldots, n\}$.

Define a matrix $M \colon \mathbb{R}^n \to \mathbb{R}^n$ with elements m_{ij}, $1 \leqslant i, \ j \leqslant n$, as follows:

> m_{ij} is the sum of the expected number of type transitions $(i,0)_c \to (i,1)$, and $(i,0) \to (i,1)$ during the entire remaining life of an individual that just became type $(1,j)$.

$$\sigma + \mu_0 \qquad \sigma + \mu_1 \qquad \sigma + \mu_2 \qquad \sigma + \mu_n$$

$$(i, -1) \xrightarrow{\rho} (i, 0)_c \xrightarrow{(1-p_i)\alpha} (i, 0) \xrightarrow{p_i\beta} (i, 1) \xrightarrow{\theta_1} (i, 2) \xrightarrow{\theta_2} \cdots \xrightarrow{\theta_{n-1}} (i, n)$$

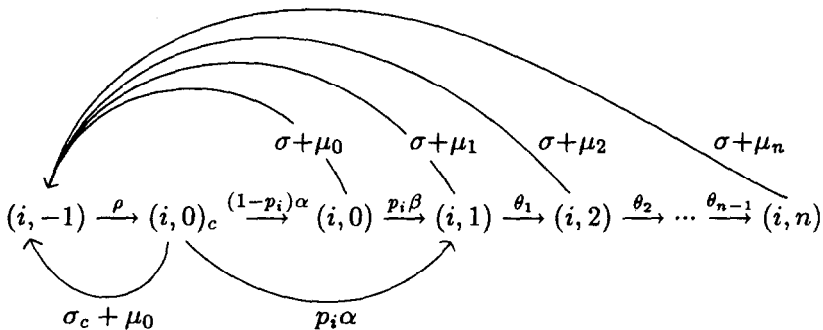$$\sigma_c + \mu_0 \qquad\qquad p_i\alpha$$

FIG. 1. Graph of possible type changes of an infected individual with fixed disease state i.

The matrix M is the next-generation operator, mapping a generation of infectives, distributed with respect to the disease state of the partner at the moment infection took place, onto the next such generation. In other words, M yields the next generation, given the present one, while keeping account of the state at "birth." As was shown in [3], we have to carry out the right averaging over the m_{ij} to arrive at R_0; R_0 is the dominant eigenvalue of the matrix M. In the following we determine the precise nature of M on the basis of the assumptions listed above.

It is convenient to work both with types in Λ (in cases where we use the interpretation to make inferences) and with states in Σ (if we just do straightforward linear algebra), and we will accordingly choose the representation that is the easiest in a given situation.

We regard the changes in disease state and partnership state of an individual as a Markov process on the state space Σ. Let a matrix G: $\mathbb{R}^{n(n+3)} \to \mathbb{R}^{n(n+3)}$ describe the transition probabilities per unit of time between the states; that is, g_{kl} gives the rate of leaving state $l \in \Sigma$ to go to state $k \in \Sigma$ (note that in the probabilistic literature on Markov processes it is usually the other way around), and $g_{ll} = -\sum_{k \neq l} g_{kl}$ the rate of dying. If we let $P(\tau)$: $\mathbb{R}^{n(n+3)} \to \mathbb{R}^{n(n+3)}$ be the matrix containing the probabilities $P_{kl}(\tau)$ of being in state k and alive at time τ after starting in state l at time zero, then we have

$$P(\tau) = e^{G\tau}.$$

The interpretation of the m_{ij} tells us that for their calculation we need to know the probability that a freshly infected individual "born" in state $L(1, j)$ is still alive at time τ after the infection occurred and that its state is

$L(i,0)_c$ or $L(i,0)$ at that time. Then,

$$m_{ij} = p_i\beta \int_0^\infty P_{L(i,0)L(1,j)}(\tau)\, d\tau + p_i\alpha \int_0^\infty P_{L(i,0)_cL(1,j)}(\tau)\, d\tau, \quad (2.1)$$

or, in other words, the expected number of times that an infected individual becomes of type $(i,1)$, given that it is "born" with type $(1,j)$ is

$$m_{ij} = -p_i\beta(G^{-1})_{L(i,0)L(1,j)} - p_i\alpha(G^{-1})_{L(i,0)_cL(1,j)} \quad (2.2)$$

for $1 \leqslant i,\, j \leqslant n$.

The next task is to specify G and calculate the right elements of G^{-1}. The structure of G is determined by the assumption that all infected individuals start their "infected life" with disease state 1 and that their disease state from there on rises from time to time by 1, up to n, as long as the individual does not die "along the way." Exploiting the structure, we can explicitly write down the inverse of G in a very simple way.

Example 1. We work out the case with three disease states, $n = 3$, because this is a prototype for all $n \geqslant 1$. We have $\Sigma = \{1, \ldots, 18\}$ and

$$G = \begin{pmatrix} A_1 & 0 & 0 \\ D_1 & A_2 & 0 \\ 0 & D_2 & A_3 \end{pmatrix},$$

where 0 is the 6×6 zero matrix, $D_j = \mathrm{diag}(\theta_j) = \theta_j Id$, and A_i, $i \in \{1, 2, 3\}$, is given by

$$A_i = \begin{pmatrix} a_1(i) & \mu_0 + \sigma_c & \mu_0 + \sigma & \mu_1 + \sigma & \mu_2 + \sigma & \mu_3 + \sigma \\ \rho & a_2(i) & 0 & 0 & 0 & 0 \\ 0 & (1-p_i)\alpha & a_3(i) & 0 & 0 & 0 \\ 0 & p_i\alpha & p_i\beta & a_4(i) & 0 & 0 \\ 0 & 0 & 0 & \theta_1 & a_5(i) & 0 \\ 0 & 0 & 0 & 0 & \theta_2 & a_6(i) \end{pmatrix},$$

where $a_1(i) = -\mu_i - \theta_i - \rho$; $a_2(i) = -\mu_i - \mu_0 - \theta_i - \sigma_c - \alpha$; $a_3(i) = -\mu_i - \mu_0 - \theta_i - \sigma - p_i\beta$; $a_4(i) = -\mu_i - \mu_1 - \theta_i - \theta_1 - \sigma$; $a_5(i) = -\mu_i - \mu_2 - \theta_i - \theta_2 - \sigma$; $a_6(i) = -\mu_i - \mu_3 - \theta_i - \sigma$. It is easily verified that G^{-1} can

be expressed in the 6×6 matrices that constitute G as follows:

$$
G^{-1} = \begin{pmatrix}
A_1^{-1} & 0 & 0 \\
-D_1 A_2^{-1} A_1^{-1} & A_2^{-1} & 0 \\
D_1 D_2 A_3^{-1} A_2^{-1} A_1^{-1} & -D_2 A_3^{-1} A_2^{-1} & A_3^{-1}
\end{pmatrix}.
$$

The matrix M can now be determined. Let us consider the special case $p_2 = 0$. The assumption that individuals with disease state 2 are not infectious implies both that no individual in state $L(2,0)$, or $L(2,0)_c$, can infect its partner and that the individual itself cannot have been "born" in the state $L(1,2)$. We find that

$$
M = \begin{pmatrix}
m_{11} & 0 & m_{13} \\
0 & 0 & 0 \\
m_{31} & 0 & m_{33}
\end{pmatrix}.
$$

The dominant eigenvalue of this matrix equals the dominant eigenvalue of

$$
M' = \begin{pmatrix}
m_{11} & m_{13} \\
m_{31} & m_{33}
\end{pmatrix}.
$$

In terms of the elements of G^{-1} we can write

$$
m_{11} = -p_1 \beta (G^{-1})_{L(1,0)L(1,1)} - p_1 \alpha (G^{-1})_{L(1,0)_c L(1,1)} = -p_1 \beta (G^{-1})_{3,4}
$$
$$
- p_1 \alpha (G^{-1})_{2,4},
$$

$$
m_{13} = -p_1 \beta (G^{-1})_{L(1,0)L(1,3)} - p_1 \alpha (G^{-1})_{L(1,0)_c L(1,3)} = -p_1 \beta (G^{-1})_{3,6}
$$
$$
- p_1 \alpha (G^{-1})_{2,6},
$$

$$
m_{31} = -p_3 \beta (G^{-1})_{L(3,0)L(1,1)} - p_3 \alpha (G^{-1})_{L(3,0)_c L(1,1)} = -p_3 \beta (G^{-1})_{15,4}
$$
$$
- p_3 \alpha (G^{-1})_{14,4},
$$

$$
m_{33} = -p_3 \beta (G^{-1})_{L(3,0)L(1,3)} - p_3 \alpha (G^{-1})_{L(3,0)_c L(1,3)} = -p_3 \beta (G^{-1})_{15,6}
$$
$$
- p_3 \alpha (G^{-1})_{14,6}.
$$

Finally we find that

$$
R_0 = \frac{(m_{11} + m_{33}) + \sqrt{(m_{11} - m_{33})^2 + 4 m_{31} m_{13}}}{2}.
$$

■

In the general case G will be an $n(n+3) \times n(n+3)$ matrix of the following form (where the A_i, D_i, and 0 are $(n+3) \times (n+3)$ versions of their namesakes from example 1):

$$
G = \begin{pmatrix}
A_1 & 0 & \cdots & \cdots & 0 \\
D_1 & A_2 & \ddots & & \vdots \\
0 & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & D_{n-1} & A_n
\end{pmatrix}.
$$

One checks easily that

$$
G^{-1} = \begin{pmatrix}
A_1^{-1} & 0 & \cdots & \cdots & 0 \\
 & A_2^{-1} & \ddots & & \vdots \\
 & & \ddots & \ddots & \vdots \\
 & B_{rs} & & \ddots & 0 \\
 & & & & A_n^{-1}
\end{pmatrix},
$$

where the (r, s)th matrix below the diagonal, B_{rs}, $r > s$, is given by

$$
B_{rs} = (-1)^{r+s} D_s D_{s+1} \cdots D_{r-1} A_r^{-1} A_{r-1}^{-1} \cdots A_s^{-1}.
$$

Note that we need to know only the inverse of every A_i in order to calculate G^{-1}.

Remark 1. As we have seen in Example 1, the analysis is simplified if some of the p_i's are zero. If m of the p_i's are nonzero, then M' becomes an $m \times m$ matrix.

Example 2. We elaborate our result for the case of one disease state and then compare it with the expression given in [5]. For $n = 1$ we have

$$
G = \begin{pmatrix}
-\mu_1 - \rho & \mu_0 + \sigma_c & \mu_0 + \sigma & \mu_1 + \sigma \\
\rho & -\mu_0 - \mu_1 - \sigma_c - \alpha & 0 & 0 \\
0 & (1-p)\alpha & -\mu_0 - \mu_1 - \sigma - p\beta & 0 \\
0 & p\alpha & p\beta & -2\mu_1 - \sigma
\end{pmatrix},
$$

and R_0 is given by

$$p\beta \int_0^\infty P_{P(1,0)L(1,1)}(\tau)\, d\tau + p\alpha \int_0^\infty P_{L(1,0)_cL(1,1)}(\tau)\, d\tau = -p\beta(G^{-1})_{3,4}$$
$$-p\alpha(G^{-1})_{2,4}.$$

Explicitly we find

$$R_0 = \frac{p\beta\alpha(\mu_1+\sigma)(1-p)\rho + p\rho\alpha(\mu_1+\mu_0+\sigma+p\beta)(\mu_1+\sigma)}{\mu_1(x+y)}, \quad (2.3)$$

with

$$x = (\mu_0+\mu_1+\sigma+p\beta)(2\mu_1+\sigma)(\mu_0+\mu_1+\rho+\sigma_c+\alpha),$$
$$y = \rho\alpha[\,p\mu_0+(2-p)\mu_1+\sigma+p\beta\,].$$

In [5] the courtship period is infinitely short. If we let the rate α of entering the sexually active phase tend to infinity (i.e., the average length of a courtship period tends to zero), we get from (2.3)

$$R_0 = \frac{p\rho(\mu_1+\sigma)(\mu_0+\mu_1+\sigma+\beta)}{\mu_1(\mu_0+\mu_1+\sigma+p\beta)(2\mu_1+\rho+\sigma)+\mu_1(1-p)\rho(\mu_1-\mu_0)}, \quad (2.4)$$

which is exactly expression (12) from [5] (with appropriate renaming of parameters). ∎

Remark 2. If we choose $\alpha = \beta$ and $\sigma_c = \sigma$ and we ''lump'' the types $(i,0)_c$ and $(i,0)$, for each $i \in \{1,\ldots,n\}$, we are in a situation similar to the one described in [5] but with the difference that a pair does not necessarily start with a sexual contact. Let us consider the case $n = 1$. Then G is a 3×3 matrix given by

$$G = \begin{pmatrix} -\mu_1-\rho & \mu_0+\sigma & \mu_1+\sigma \\ \rho & -\mu_0-\mu_1-\sigma-p\beta & 0 \\ 0 & p\beta & -2\mu_1-\sigma \end{pmatrix},$$

and $R_0 = -p\beta(G^{-1})_{2,3}$ or, explicitly,

$$R_0 = \frac{p\beta\rho(\mu_1+\sigma)}{\mu_1(\mu_0+\mu_1+\sigma+\rho+p\beta)(2\mu_1+\sigma)+\rho p\beta\mu_1}. \quad (2.5)$$

3. INCORPORATING HETEROGENEITY IN SUSCEPTIBILITY

If we want to incorporate heterogeneity among the individuals in the population we have to specify not only the characteristics (called h-state) of an individual itself but also those of its current partner (if the individual is not single) because the h-state of the partner can influence the death rate and in this way the probability that the original individual becomes single. The characteristics can take discrete or continuous values, and the h-state of an individual can be constant in time or dynamic. Among the most important characteristics to be incorporated in the context of sexually transmitted diseases are age, sexual activity level, male/female, homo-/bi-/heterosexual, and behavioral traits such as condom use.

Let, in general, a variable ξ represent the heterogeneity characteristics of an individual; ξ is assumed to take values in some set Ω. The type of a sexually active individual in a pair is now represented by

$$\left(i, j; \xi_i, \xi_j\right), \qquad i \in \{1, \ldots, n\}; \; j \in \{0, \ldots, n\}; \; \xi_i, \xi_j \in \Omega,$$

while $(i, -1; \xi_i)$ denotes a single infected individual, and $(i, 0; \xi_i, \xi_0)_c$ describes the relevant types of infected individuals in the courtship phase of a partnership. Suppose we have an individual, say X, that was infected by an individual with disease state j and h-state ν. Suppose X itself had h-state η at the moment of infection. Then X was "born" with type $(1, j; \eta, \nu)$. As time progresses, assuming that X stays alive, X will become separated from its original partner, the h-state of X will change to, say, θ, and X will form a pair with a susceptible with h-state, say, ξ. Analogously to the case without heterogeneity in Section 2, we want to evaluate the expected number of type transitions

$$\left(i, 0; \theta, \xi\right)_c \rightarrow \left(i, 1; \theta, \xi\right), \qquad \left(i, 0; \theta, \xi\right) \rightarrow \left(i, 1; \theta, \xi\right),$$

that is, partner infections, of our individual X during its entire remaining life. In analogy with the notation of Section 2 we denote this number by $m_{ij}(\xi, \theta; \eta, \nu)$. Let the current infectivity of X toward its partner be described by the probability $p_i(\xi, \theta)$ and the rate of entering the sexually active phase by $\alpha(\xi, \theta)$, and let $\beta(\xi, \theta)$ be the number of sexual contacts per unit of time within the sexually active phase of a partnership. Generalizing the expressions from Section 2, we find

$$m_{ij}(\xi, \theta; \eta, \nu) = p_i(\xi, \theta)\beta(\xi, \theta) \int_0^\infty P_{L(i,0)L(1,j)}(\tau, \xi, \theta; \eta, \nu) \, d\tau$$

$$+ p_i(\xi, \theta)\alpha(\xi, \theta) \int_0^\infty P_{L(i,0)_c L(1,j)}(\tau, \xi, \theta; \eta, \nu) \, d\tau,$$

where $P_{L(i,0)L(1,j)}(\tau, \xi, \theta; \eta, \nu)$ denotes the probability that at time τ after X became infected as type $(1, j; \eta, \nu)$ it is still alive and currently has type $(i, 0; \theta, \xi)$. The analogous quantity with index c has a similar interpretation.

Let $\phi = \phi(j; \eta, \nu)$ be the distribution of just infected individuals over the space $\{1, \ldots, n\} \times \Omega \times \Omega$. We call this a *generation*. The next generation consists of all individuals that are infected by the members of this generation, ϕ. When one individual "born" with type $(j; \eta, \nu)$ infects $m_{ij}(\xi, \theta; \eta, \nu)$ partners of h-state ξ while having disease state i and h-state θ, we obtain, by summing with respect to j, η, and ν, for the next generation

$$(K\phi)(i; \xi, \theta) = \sum_{j=1}^{n} \int_{\Omega \times \Omega} m_{ij}(\xi, \theta; \eta, \nu)\phi(j; \eta, \nu)\, d\eta\, d\nu \quad (3.1)$$

cases that are "born" with type $(i; \xi, \theta)$. Formula (3.1) defines the next-generation operator K, which tells us both how many secondary cases arise and how they are distributed with respect to type at "birth."

We regard the next-generation operator K as an operator mapping $L_1(\{1, \ldots, n\} \times \Omega \times \Omega)$ into itself. As shown in [3], R_0 is the spectral radius of the operator K.

Remark 3. It could prove to be no more than an academic exercise to work at this level of generality because it will be rather involved to determine analytically the probabilities P in the case of a dynamic continuous h-state such as *age*. Instead of solving a coupled set of ODEs, which is basically what happens in Section 2, one has to solve a coupled system of PDEs. See [8] for a different approach in the case of age as h-state, but with much more restrictive assumptions.

Example 3. We discuss the simplest example. Let $\Omega = \{1, 2\}$, where 1 represents females and 2 represents males, and take only heterosexual contacts into account. The next-generation operator K is in this case a $2n \times 2n$ matrix of form

$$K = \begin{pmatrix} 0 & K_1 \\ K_2 & 0 \end{pmatrix},$$

where $K_1 = (m_{ij}(1, 2; 2, 1))_{1 \le i, j \le n}$ and $K_2 = (m_{ij}(2, 1; 1, 2))_{1 \le i, j \le n}$. For the spectral radius $r(A)$ of an operator A, we have that $r(A^k) = r(A)^k$, $k \ge 1$; furthermore, if B is another operator, then $r(AB) = r(BA)$. Since

$$K^2 = \begin{pmatrix} K_1 K_2 & 0 \\ 0 & K_2 K_1 \end{pmatrix},$$

we find

$$R_0 = \sqrt{r(K_1 K_2)} \, .$$

Note that, because we look only at heterosexual contacts, the probabilities $P_{L(i,0)L(1,j)}(\tau)$ and $P_{L(i,0)_cL(1,j)}(\tau)$ can be calculated in a way that is completely analogous to the example in Section 2, with the only differences being that death rates, infectivities in each disease state, and the rates of change in disease states are allowed to depend on the h-state of the individuals in the pair. Let us treat the case of $n = 1$ in somewhat more detail. Let the male parameter set be given by $\{\mu_1, \mu_0, p, \beta, \sigma, \rho, \alpha\}$ and the female set by $\{\mu_1', \mu_0', p', \beta, \sigma', \rho', \alpha\}$, where p is the probability that a male infects a female. Note that there will be consistency requirements involving the pair-formation parameters.

The transition matrix G is given by

$$G = \begin{pmatrix} G^{12} & 0 \\ 0 & G^{21} \end{pmatrix},$$

where G^{12} and G^{21}, which describe how the types of a female or male individual, respectively change, are essentially the matrix from Example 2 with appropriate placing of accents. We find

$$K_1 = m_{11}(1,2;2,1) = -p\beta(G^{21})_{3,4}^{-1} - p\alpha(G^{21})_{2,4}^{-1},$$

$$K_2 = m_{11}(2,1;1,2) = -p'\beta(G^{12})_{3,4}^{-1} - p'\alpha(G^{12})_{2,4}^{-1},$$

and $R_0 = \sqrt{K_1 K_2}$.   ■

For arbitrary heterogeneity one can derive an explicit formula for R_0 in the very special case that the next-generation operator K has a one-dimensional range. If we assume

$$m_{ij}(\xi,\theta;\eta,\nu) = a_i(\xi,\theta)b_j(\eta,\nu), \tag{3.3}$$

the only eigenvalue of K is

$$R_0 = \sum_{j=1}^{n} \int_{\Omega \times \Omega} b_j(\eta,\nu) a_j(\eta,\nu) \, d\eta \, d\nu.$$

In [3], assumption (3.3) is called *separable infectivity and susceptibility*, or *separable mixing*. If the functions a and b are equal up to a multiplicative constant, then the assumption is known as *proportionate mixing* [3].

Somewhat less restrictive than separable mixing is the assumption

$$m_{ij}(\xi,\theta;\eta,\nu) = a_i(\xi,\theta)b_{ij}(\eta,\nu),\qquad(3.4)$$

which leads to a next-generation operator with n-dimensional range, where n is the number of disease states. We then have [3] that the eigenvalues of K are equal to those of an $n \times n$ matrix $E = (e_{ij})$ with

$$e_{ij} = \int_{\Omega \times \Omega} b_{ij}(\eta,\nu)a_j(\eta,\nu)\,d\eta\,d\nu.$$

R_0 is then the dominant eigenvalue of E. Assumption (3.4) is called *local separable infectivity and susceptibility*, or local separable mixing, in [3].

4. VARIOUS LIMIT PROCEDURES

In this short section we show how various limit procedures can lead to interesting expressions for R_0. We restrict ourselves to the case of one disease state. The R_0 for this case is explicitly given in Equation (2.3) in Example 2. First we "collapse" the sexually active period of a partnership to a point event. We write $\beta = k\sigma + O(1)$ and let $\sigma \to \infty$. Then $1 + k$ is the average number of sexual contacts during one partnership with a sexually active phase, and (2.3) simplifies to

$$R_0 = \frac{\rho \alpha p(1+k)}{\mu_1(1+pk)(\mu_0 + \mu_1 + \rho + \sigma_c + \alpha)}.\qquad(4.1)$$

The interpretation of (4.1) is as follows: α is the rate of becoming sexually active, given that one is in the courtship phase; $\rho/[\mu_1(\mu_0 + \mu_1 + \rho + \sigma_c + \alpha)]$ is the expected time that an infected individual will have a susceptible partner, that is, the expected time spent in courtship (the product of this term with α gives the expected number of sexual partners); $p(1+k)/(1+pk)$ is the success ratio per sexual partner (this is identical to the formula on page 405 of [4]). Note that (4.1) also covers the case where β remains bounded (simply put $k = 0$, this means that there are only "first contacts").

To completely eliminate pair formation from our model we still have to let $\alpha \to \infty$ in (4.1); in other words, we have to let the length of the courtship period tend to zero. We then find

$$R_0 = \rho p(1+k)/\mu_1(1+pk).\qquad(4.2)$$

This gives $R_0 = p\rho/\mu_1$ for $k = 0$, which can be found immediately from the appropriate non-pair-formation model by looking at the interpretation of the parameters.

If, instead of α, we let $\rho \to \infty$ in (4.1), we are in the situation where the individual is constantly in the courtship phase,

$$R_0 = \alpha p(1 + k)/\mu_1(1 + pk).$$

Remark 4. In the case without pair formation, the formal route to R_0 would be to specify the infectivity A as a function of disease age τ and calculate [2, 7] $R_0 = \int_0^\infty A(\tau) \, d\tau$. Under our assumptions, listed in Section 2, $A(\tau)$ has a special form $A(\tau) = p\rho e^{-\mu_1 \tau}$, and this leads once more to (4.2) with $k = 0$. In the case of n possible disease states, $A(\tau)$ involves an expectation for an infected individual to have a certain disease state. The limit procedure to eliminate the pair formation completely is essentially the same as above: Let $\sigma \to \infty$, $\alpha \to \infty$, and take $k = 0$.

In the heterogeneous case we conjecture that a similar limit argument collapses the spectral radius of the next-generation operator (3.1) into the spectral radius of the operator

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

from [3], but with a special form for the infectivity kernel A. As an illustration we look at the case where we recognize male and female individuals and allow only heterosexual contacts (Example 3). In the situation without pair formation we have [9]

$$R_0 = \sqrt{pp'\rho\rho'/\mu_1\mu_1'} \,. \tag{4.3}$$

R_0 for the pair formation case, from Example 3, transforms into (4.3) if we let σ, σ', and α tend to infinity.

In Part II of this paper (in preparation), the theory developed here will be used to derive quantitative results that are relevant in the context of HIV/AIDS. Among other things we will discuss there the influence on R_0 of (i) the pair-formation parameters, (ii) changes in behavior, and (iii) testing for seropositivity.

REFERENCES

1 S. P. Blythe and R. M. Anderson, Variable infectiousness in HIV transmission models, *IMA J. Math. Appl. Med. Biol.* 5:181–200 (1988).
2 O. Diekmann, Modeling epidemics in structured populations, *Proceedings of the 11th Dundee Conference on the Theory of Ordinary and Partial Differential Equations*, Pitman Research Notes in Mathematics, 1990.
3 O. Diekmann, J. A. P. Heesterbeek and J. A. J. Metz, On the definition and the calculation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations, *J. Math. Biol.* 28:365–382 (1990).
4 K. Dietz, On the transmission dynamics of HIV, *Math. Biosci.* 90:397–414 (1988).
5 K. Dietz, The role of pair formation in the transmission dynamics of HIV, in *Sesquicentennial Invited Paper Sessions*, M. H. Gail and N. L. Johnson, Eds., Am. Statistical Association, Alexandria, VA, 1989, pp. 609–621.
6 K. Dietz and K. P. Hadeler, Epidemiological models for sexually transmitted diseases, *J. Math. Biol.* 26:1–25 (1988).
7 W. O. Kermack and A. G. McKendrick, Contributions to the mathematical theory of epidemics, *Proc. Roy. Soc. Lond. A* 115:700–721 (1927).
8 H. Knolle, Age preference in sexual choice and the basic reproduction number of HIV/AIDS, *Biometr. J.* 32:243–256 (1990).
9 R. M. May and R. M. Anderson, The transmission dynamics of human immunodeficiency virus (HIV), *Phil. Trans. Roy. Soc. Lond. B* 321:565–607 (1988).