MATHEMATICAL BIOSCIENCES
AND ENGINEERING
Volume **4**, Number **2**, April **2007**

http://www.mbejournal.org/

pp. **159–175**

A FINAL SIZE RELATION FOR EPIDEMIC MODELS

Julien Arino

Department of Mathematics, University of Manitoba
Winnipeg, Manitoba R3T 2N2, Canada

Fred Brauer

Department of Mathematics, University of British Columbia
Vancouver, BC V6T 1Z2, Canada

P. van den Driessche

Department of Mathematics and Statistics, University of Victoria
Victoria, B.C. V8W 3P4, Canada

James Watmough

Department of Mathematics and Statistics, University of New Brunswick
Fredericton, N.B. E3B 5A3, Canada

Jianhong Wu
Department of Mathematics and Statistics, York University
Toronto, Ontario M3J 1P3, Canada

(Communicated by Zhilan Feng)

Abstract. A final size relation is derived for a general class of epidemic models, including models with multiple susceptible classes. The derivation depends on an explicit formula for the basic reproduction number of a general class of disease transmission models, which is extended to calculate the basic reproduction number in models with vertical transmission. Applications are given to specific models for influenza and SARS.

1. **Introduction.** The purpose of this paper is to obtain a relation between the basic reproduction number and the final size of an epidemic in a general deteministic model of disease transmission.

The basic reproduction number, a central concept in the study of the spread of communicable diseases [1], is defined as the spectral radius of the next generation operator [7, 8], and if the model is formulated as a system of ordinary differential equations the basic reproduction number is the spectral radius of a matrix whose entries are determined by the model parameters [16]. We concentrate mainly on the very common case that this matrix has rank 1, and in Section 2 we obtain an explicit formula for a general class of models with horizontal disease transmission only. While this representation has been mentioned in [7, p.107], it has not been used for explicit calculations. In Section 3 we incorporate vertical disease transmission into the model. In this case the matrix may have rank 1, so that this formula is

2000 *Mathematics Subject Classification.* 92D30.

Key words and phrases. epidemic models, basic reproduction number, final size relation, vertical transmission.

applicable, but we also describe some examples in which the matrix has rank higher than 1.

In Sections 4 and 5 we show that for a general class of epidemic models the infection will die out and the representation formula for the basic reproduction number can be used to derive an explicit relation between the basic reproduction number and the final size of the epidemic. This final size relation generalizes the one originally given by Kermack and Mckendrick [10] to models in which there are multiple susceptible classes, such as models including vaccination. A final size relation for models with multiple susceptible classes has previously been established in [15, Theorem 2.2] but without reference to the basic reproduction number. It has recently been shown [13] that the final size relation of Kermack and McKendrick [10] holds for models with multiple infective stages and models in which the durations of stages are arbitrarily distributed. In Section 6 we extend the epidemic results to models with a continuing inflow of infectives. In Sections 7 and 8 we give examples for specific disease models to illustrate applications of the formulae and a heterogeneous mixing epidemic model. We end with some conclusions in Section 9.

2. **The basic reproduction number.** Consider a general disease transmission model in which $x \in R^n$ represents the set of infected compartments, $y \in R^m$ represents the set of susceptible compartments, and $z \in R^k$ represents the set of compartments removed from disease either by immunity or by recovery with immunity. While y and z correspond to disease-free compartments, we separate them because only y contributes to disease transmission. The vectors x, y, z are functions of time t, and we use $'$ to denote differentiation with respect to t.

We let D be an $m \times m$ diagonal matrix whose diagonal entries $\sigma_i > 0$ are the relative susceptibilities of the corresponding susceptible class; we take $\sigma_1 = 1$ without loss of generality. Note that if $m = 1$, then D is the scalar 1. We let Π be an $n \times m$ matrix with the property that the (i, j) entry represents the fraction of the j^{th} susceptible compartment that goes into the i^{th} infective compartment on becoming infected. We also let b be an n-dimensional row vector of relative horizontal transmissions. This vector is multiplied by a scalar factor representing infectivity. For general incidence this factor is a function of total population size and/or infective population size, and we write it as $\beta(x, y, z)$. For mass action incidence $\beta(x, y, z)$ is a constant β.

We assume here that there is no vertical transmission, so that there is no recruitment term in the infected compartments. However, we will return to the case of vertical transmission in Section 3. The disease model can be represented by the system

$$\begin{aligned} x' &= \Pi D y \beta(x, y, z) b x - V x \\ y' &= g(x, y, z) - D y \beta(x, y, z) b x \\ z' &= h(x, y, z) + W x, \end{aligned} \tag{1}$$

with non-negative initial conditions such that at least one component of $x(0)$ is positive. This form assumes that there is no transfer out of the class z, that is, that there is no temporary immunity after recovery. However, the results of this section are applicable to models with temporary immunity. Here, the $n \times n$ matrix V describes the transitions between infected states as well as removals from infected states through death and recovery. For any non-negative vector x, the components

of the vector Vx represent the net rate of decrease of each infected compartment. Since this rate cannot be positive if the compartment is empty, it follows that the off-diagonal entries of V must be negative or zero. Similarly, the sum of the components of the vector Vx, which represents the net rate of decrease in infected individuals due to death and recovery, must be non-negative for every non-negative vector x. It is shown in [16] that V is a non-singular M-matrix. This implies that the eigenvalues of V all have positive real part, and V^{-1} is a matrix with non-negative entries [3].

The $k \times n$ matrix W has the property that the (i, j) entry represents the rate at which members of the j^{th} disease compartment go into the i^{th} removed compartment on recovery. The function $g(x, y, z)$, assumed continuous, represents recruitment of uninfected members through birth or immigration as well as deaths of uninfected members, and the function $h(x, y, z)$, also assumed continuous, represents the flow into and out from the system of members immune to infection through natural immunity or inoculation against infection.

The disease-free set $\{(x, y, z)|x = 0, y \geq 0, z \geq 0\}$ is invariant. Suppose that a point $(0, y_0, z_0)$ is a locally stable equilibrium of the system without disease

$$\begin{aligned} y' &= g(0, y, z) \\ z' &= h(0, y, z) \end{aligned}$$

in the sense that solutions that start close to $(0, y_0, z_0)$ remain close to $(0, y_0, z_0)$. Such a point is referred to as a disease-free equilibrium. The community matrix of the system without disease at this equilibrium is

$$J_{yz} = \begin{bmatrix} g_y(0, y_0, z_0) & g_z(0, y_0, z_0) \\ h_y(0, y_0, z_0) & h_z(0, y_0, z_0) \end{bmatrix},$$

and this assumption implies that all the eigenvalues of J_{yz} have negative or zero real parts.

The point $(0, y_0, z_0)$ is also an equilibrium of the system (1). We define

$$F_h = \Pi D y_0 \beta(0, y_0, z_0) b.$$

If all eigenvalues of $F_h - V$ have negative real parts, then this point is also locally stable. If, in addition, all eigenvalues of J_{yz} have negative real parts, this equilibrium is locally asymptotically stable. If some eigenvalues of J_{yz} have zero real parts, a case that arises in epidemic models where demographic effects are not included, the local centre manifold of the equilibrium is contained in the disease-free set. Thus solutions initially near the equilibrium remain near the equilibrium and approach the disease-free set asymptotically.

According to the theory of [7, 16], the basic reproduction number $\mathcal{R}_0$ is the spectral radius of the matrix $F_h V^{-1}$. We recall that the spectral radius $\rho(A)$ of a matrix A is defined as the maximum modulus of an eigenvalue of A, and that a non-negative matrix has a real eigenvalue equal to its spectral radius [3]. We remark also that the spectral radius of $F_h V^{-1}$ has absolute value less than 1 if and only if all eigenvalues of the matrix $F_h - V$ have negative real part [16].

Since F_h is the product of the column vector $\Pi D y_0$ and the row vector b it has rank 1. Therefore $(n - 1)$ of the eigenvalues of $F_h V^{-1}$ are zero, and the spectral radius is the remaining eigenvalue. This eigenvalue is the trace of the matrix $\Pi D y_0 \beta(0, y_0, z_0) b V^{-1}$, and it is easy to verify that this is equal to the scalar $\beta(0, y_0, z_0) b V^{-1} \Pi D y_0$. The basic reproduction number is calculated with respect to a disease-free equilibrium $(0, y_0, z_0)$. This gives the following result.

THEOREM 2.1. *The basic reproduction number $\mathcal{R}_0$ for the model* (1) *at a disease-free equilibrium $(0, y_0, z_0)$ is given by*

$$\mathcal{R}_0 = \beta(0, y_0, z_0)bV^{-1}\Pi Dy_0.$$

The disease-free equilibrium is (locally) asymptotically stable if $\mathcal{R}_0 < 1$ and is unstable if $\mathcal{R}_0 > 1$.

3. **Vertical disease transmission.** Vertical disease transmission is the transmission of disease from infected members of the population to some of their offspring at birth. Such transmission occurs in many diseases, including malaria, AIDS, Chagas' disease, cholera, and dengue fever. The standard reference on vertical transmission is [6], and a more recent contribution is [11].

We assume that births are distributed proportionally among all the compartments, so that the rate of births to parents in a disease compartment i has the form $x_i\varphi(x, y, z)$ with φ assumed continuous. We assume that p_{ij} is the fraction of infected births to a member of the infected compartment x_j born into the infected compartment x_i, and we define the matrix $P = [p_{ij}]$. Then an extension of the model (1) to include vertical transmission is

$$\begin{aligned} x' &= \varphi(x, y, z)Px + \Pi Dy\beta(x, y, z)bx - Vx \\ y' &= g(x, y, z) - Dy\beta(x, y, z)bx \\ z' &= h(x, y, z) + Wx, \end{aligned}$$

with non-negative initial conditions such that at least one component of $x(0)$ is positive.

Now the matrix F is the sum of two matrices representing horizontal and vertical transmission respectively, $F = F_h + F_v$, with F_h as before and

$$F_v = \varphi(0, y_0, z_0)P.$$

The horizontal basic reproduction number is, as before,

$$\mathcal{R}_h = \rho(F_hV^{-1}) = tr(F_hV^{-1}),$$

and we define the vertical basic reproduction number

$$\mathcal{R}_v = \rho(F_vV^{-1}).$$

If $F_h + F_v$ has rank 1, which is true, for example, if all new infections, both horizontally and vertically transmitted, come from a source in one infected compartment, or if all new infections go into the same infected compartment, then

$$\begin{aligned} \mathcal{R}_0 &= tr(FV^{-1}) = tr((F_v + F_h)V^{-1}) \\ &= tr(F_hV^{-1}) + tr(F_vV^{-1}) = \mathcal{R}_h + \mathcal{R}_v. \end{aligned} \tag{2}$$

Intuitively, we would expect this to be true in general. However, if F_v has rank greater than 1, then R_v is not necessarily equal to the trace of F_vV^{-1}, and if F_h+F_v has rank greater than 1 then $\mathcal{R}_0$ is not necessarily equal to the trace of FV^{-1} even if F_h and F_v both have rank 1. We now give some examples illustrating that (2) is not always true. In our examples we assume for simplicity that P is a diagonal matrix, and this choice appears to be biologically reasonable. It is possible to carry out the calculations with a general P.

Some vertical transmission examples. In disease transmission models it has become standard to use E to denote the class of exposed (infected but not yet infective) members. However, epidemiologists also use the term latent to describe such individuals. This terminology is less ambiguous because it makes it clear that these individuals will develop infection while the term exposed could be interpreted to include the possibility of contact with an infective without developing infection. Accordingly, we use L to denote the number of latent (infected but not yet fully infective) members, and we use L in place of E in identifications of models. Thus we speak of $SLIR$ models rather than $SEIR$ models. Latent members may have some infectivity, usually less than infective members. Letting S, I, R denote the number of susceptible, infective, and removed members respectively, we consider some $SLIR$ models that include vertical transmission. For simplicity, we assume mass action incidence so that $\beta(x, y, z)$ is a constant β, but it is easy to extend the examples to general incidence.

We consider an $SLIR$ model with a birth rate $N\phi(N)$ divided proportionally among the classes in which infected births may arise from parents in both L and I, and in which each newborn infected is in the same compartment as the infected parent. We let μ be the natural death rate and we assume the population has a carrying capacity K in the absence of disease with $\phi(K) = \mu, \phi'(K) < 0$. We let $1/\kappa$ the mean latent period, $1/\alpha$ the mean infective period, p_1 the fraction of infected births to a parent in L, and p_2 the fraction of infected births to a parent in I. We assume also that a fraction f of infectives recovers while the complementary fraction $1 - f$ dies of disease. Finally, we assume that there is some infectivity in the latent class, multiplying the infectivity in the infective class by a factor ε. This leads to the model generalizing that of [7, Exercise 2.2] by the inclusion of infectivity in the latent class,

$$\begin{aligned} S' &= (N - p_1 L - p_2 I)\varphi(N) - \mu S - \beta S(I + \varepsilon L) \\ L' &= p_1 L\varphi(N) + \beta S(I + \varepsilon L) - (\kappa + \mu)L \\ I' &= p_2 I\varphi(N) + \kappa L - (\alpha + \mu)I \\ R' &= f\alpha I - \mu R \end{aligned}$$

with $N = S + L + I + R$. There is a disease-free equilibrium with $L = I = R = 0$, and at the disease-free equilibrium $N = S = S_0 = K$.

In terms of our notation, $m = 1, n = 2$, D is the scalar 1, and

$$b = [\varepsilon, 1], \quad \Pi = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad P = \begin{bmatrix} p_1 & 0 \\ 0 & p_2 \end{bmatrix}.$$

Also,

$$F = \begin{bmatrix} \varepsilon S_0\beta + \mu p_1 & S_0\beta \\ 0 & \mu p_2 \end{bmatrix}.$$

We separate F into horizontal and vertical parts,

$$F_h = \begin{bmatrix} \varepsilon S_0\beta & S_0\beta \\ 0 & 0 \end{bmatrix}, \qquad F_v = \begin{bmatrix} \mu p_1 & 0 \\ 0 & \mu p_2 \end{bmatrix}.$$

We have

$$V = \begin{bmatrix} \kappa + \mu & 0 \\ -\kappa & \alpha + \mu \end{bmatrix}, \quad V^{-1} = \begin{bmatrix} \dfrac{1}{\kappa + \mu} & 0 \\ \dfrac{\kappa}{(\kappa + \mu)(\alpha + \mu)} & \dfrac{1}{\alpha + \mu} \end{bmatrix}.$$

Since F_h has rank 1, Theorem 2.1 gives

$$\mathcal{R}_h = \frac{S_0\beta\kappa}{(\kappa+\mu)(\alpha+\mu)} + \frac{\varepsilon S_0\beta}{\kappa+\mu}.$$

Since the eigenvalues of F_vV^{-1} are $\lambda_1 = \mu p_1/(\kappa+\mu)$ and $\lambda_2 = \mu p_2/(\alpha+\mu)$,

$$\mathcal{R}_v = \max\left(\frac{\mu p_1}{\kappa+\mu}, \frac{\mu p_2}{\alpha+\mu}\right).$$

The basic reproduction number $\mathcal{R}_0$ is the larger eigenvalue of the matrix

$$FV^{-1} = \begin{bmatrix} \dfrac{\varepsilon S_0\beta}{\kappa+\mu} + \dfrac{S_0\beta\kappa}{(\kappa+\mu)(\alpha+\mu)} + \dfrac{\mu p_1}{\kappa+\mu} & \dfrac{S_0\beta}{\alpha+\mu} \\ \dfrac{\kappa\mu p_2}{(\kappa+\mu)(\alpha+\mu)} & \dfrac{\mu p_2}{\alpha+\mu} \end{bmatrix} = \begin{bmatrix} \mathcal{R}_h + \lambda_1 & \dfrac{S_0\beta}{\alpha+\mu} \\ \dfrac{\kappa}{\kappa+\mu}\lambda_2 & \lambda_2 \end{bmatrix}.$$

If $p_2 = 0$ (no vertical transmission in the infective class) or if $p_1 = 0$ and $\varepsilon = 0$ (no infectivity in the latent class and no vertical transmission in the latent class) then $F = F_h + F_v$ has rank 1 and (2) is valid.

The spectral radius of a 2×2 non-negative matrix is at least as large as the maximum of the diagonal elements. Thus

$$\mathcal{R}_0 \geq \max(\mathcal{R}_h + \lambda_1, \lambda_2).$$

The second diagonal entry of FV^{-1} corresponds to births in I from mothers in I, and the first diagonal element corresponds to the sum of secondary infections, both horizontal and vertical, due to an index case in L. Thus, if $\lambda_1 \geq \lambda_2$, that is, if vertical transmission from L to L is equal to or exceeds vertical transmission from I to I, then $\mathcal{R}_0 \geq \mathcal{R}_h + \mathcal{R}_v$. However, if $\lambda_1 < \lambda_2$ then it remains true that $\mathcal{R}_0 \geq \mathcal{R}_h + \lambda_1$, but as we show below it is not necessarily true that $\mathcal{R}_0 \geq \mathcal{R}_h + \mathcal{R}_v$.

The characteristic polynomial of FV^{-1} is

$$g(\lambda) = \lambda^2 - (\lambda_1 + \lambda_2 + \mathcal{R}_h)\lambda + \lambda_2\left[\lambda_1 + (\mathcal{R}_h - \frac{\kappa}{\kappa+\mu}\frac{S_0\beta}{\alpha+\mu})\right].$$

If $g(\mathcal{R}_h + \mathcal{R}_v) < 0$, then, since $g'(\mathcal{R}_h + \mathcal{R}_v) > 0$, it follows that $\mathcal{R}_0 > \mathcal{R}_h + \mathcal{R}_v$. Likewise, $g(\mathcal{R}_h + \mathcal{R}_v) = 0$ implies $\mathcal{R}_0 = \mathcal{R}_h + \mathcal{R}_v$, and $g(\mathcal{R}_h + \mathcal{R}_v) > 0$ implies $\mathcal{R}_0 < \mathcal{R}_h + \mathcal{R}_v$. We have

$$g(\mathcal{R}_h + \mathcal{R}_v) = (\mathcal{R}_v - \lambda_1)(\mathcal{R}_h + \mathcal{R}_v) + \lambda_2\left[\lambda_1 - \mathcal{R}_v - \frac{\kappa}{\kappa+\mu}\frac{S_0\beta}{\alpha+\mu}\right].$$

If $\lambda_1 \geq \lambda_2$, so that $\mathcal{R}_v = \lambda_1$, then

$$g(\mathcal{R}_h + \mathcal{R}_v) = -\frac{\kappa}{\kappa+\mu}\lambda_2\frac{S_0\beta}{\alpha+\mu}.$$

Thus $g(\mathcal{R}_h + \mathcal{R}_v) = 0$ if $\lambda_2 = 0$, that is, if $p_2 = 0$, and $g(\mathcal{R}_h + \mathcal{R}_v) < 0$ if $p_2 > 0$. This implies that

$$\mathcal{R}_0 > \mathcal{R}_h + \mathcal{R}_v$$

if $p_2 > 0$. In this case, $\mathcal{R}_v$ counts only one of the two infected compartments in which there are births, but the births not counted here do contribute to $\mathcal{R}_0$.

If $\lambda_1 \leq \lambda_2$, so that $\mathcal{R}_v = \lambda_2$,

$$g(\mathcal{R}_v + \mathcal{R}_h) = \frac{\varepsilon S_0\beta}{\kappa+\mu}\mathcal{R}_v - \frac{\mu p_1}{\kappa+\mu}\mathcal{R}_h.$$

Thus, if

$$\varepsilon S_0\beta\mathcal{R}_v > \mu p_1\mathcal{R}_h,$$

then $\mathcal{R}_0 < \mathcal{R}_h + \mathcal{R}_v$. This is true, for example, if $\varepsilon > 0$ and $p_1 = 0$. Note that here F_h and F_v have rank 1 but $F_h + F_v$ has rank 2. In this situation there are some births in the infective class who do not go through a latent stage in which they would contribute secondary infections, and this is the reason for the inequality.

To summarize these calculations, if $\varepsilon = 0$ and $p_1 = 0$, so that there are no infections arising from the latent class, or if $p_2 = 0$, so that there are no new infections arising from births in the infective class, then the basic reproduction number is equal to the sum of the horizontal and vertical reproduction numbers. If there are infected newborns in both latent and infective classes, with the latent class contribution at least as large, then $\mathcal{R}_0 > \mathcal{R}_h + \mathcal{R}_v$. On the other hand, it is possible to have $\mathcal{R}_0 < \mathcal{R}_h + \mathcal{R}_v$ if the contribution to new infections arising from births in the infective class is large compared to the contribution coming from the latent class.

4. **Epidemic models.** In the special case of an epidemic, we ignore demographic effects, and may assume that $g(x, y, z) = h(x, y, z) = 0$ in (1). In particular, the possibility of vertical disease transmission does not arise. Also, the matrix V no longer includes natural deaths. Then the epidemic model can be represented by the system

$$\begin{aligned} x' &= \Pi D y \beta(x, y, z) b x - V x \\ y' &= -D y \beta(x, y, z) b x \\ z' &= W x \end{aligned} \tag{3}$$

with non-negative initial conditions such that at least one component of $x(0)$ is positive. The equation for z enters into the analysis of the first two equations of (3) only if the incidence is more general than mass action.

For the model (3) every point $(0, y_0, z_0)$ is an equilibrium. The calculation of the basic reproduction number with respect to any equilibrium $(0, y_0, z_0)$ is the same as for the model (1) with demographic effects, and Theorem 2.1 applies to the model (3).

For epidemic models we first show that the number of members in each infected compartment tends to zero as $t \to \infty$. In analyzing the system (3) we adopt the conventions that for an arbitrary continuous function $w(t)$ with non-negative components,

$$w_\infty = \lim_{t \to \infty} w(t), \quad \hat{w} = \int_0^\infty w(t) dt.$$

Addition of the equations in (3) gives

$$(x + \Pi y)' = -V x. \tag{4}$$

Integration of (4) with respect to t from 0 to ∞ gives

$$(x(0) - x_\infty) + \Pi(y(0) - y_\infty) = V \hat{x}. \tag{5}$$

The left side of (5) is finite because the components of $x(0), y(0), x_\infty$ and y_∞ are bounded by the initial total population size. Therefore the right side is also finite and because V is non-singular, $\hat{x} < \infty$. Since each component of x is a smooth non-negative function, $x_\infty = 0$, and

$$\hat{x} = V^{-1} \Pi(y(0) - y_\infty) + V^{-1} x(0). \tag{6}$$

Since the components of $V\hat{x}$ are the numbers of infected individuals leaving each infected compartment over the course of the epidemic, the total number of cases of disease in the epidemic is the sum of the components of

$$V\hat{x} = \Pi(y(0) - y_\infty) + x(0).$$

5. **The final size relations.** The final size relations are relations involving the basic reproduction number and the number of members of the population that remain in each disease-free compartment over the course of the epidemic. If $m = 1$ the relations are explicit in $\mathcal{R}_0$, while if $m > 1$ they are given in terms of a vector Γ to be defined shortly with $\mathcal{R}_0 = \Gamma y_0$. For mass action incidence (β constant), the final size relations are equalities. For general incidence, there are always upper and lower bounds for $\beta(x, y, z)$ and these bounds give upper and lower bounds for the quantities $\ln[y_i(0)/y_i(\infty)]$. These bounds yield the final size relations as inequalities. In this section, we use mass action incidence and treat β as constant and give some remarks at the end of this section on the interpretation of the final size inequalities as approximate equalities in the general case.

In terms of components, the equations for y in (3) take the form

$$y_i' = -\sigma_i \beta b x y_i, \quad i = 1, 2, \ldots, m. \tag{7}$$

If $y_i(0) = 0$, then $y_i(t) = 0$ for $t \geq 0$, while if $y_i(0) > 0$, then $y_i(t) > 0$ for $t > 0$. Next, we show that $y_i(\infty) > 0$. Division of (7) by y_i and integration of these scalar equations gives

$$\ln\left(\frac{y_i(0)}{y_i(t)}\right) = \sigma_i \beta b \int_0^t x(s)ds \leq \sigma_i \beta b \hat{x}.$$

Since the right side of this inequality is finite for $0 \leq t < \infty$, the left side remains finite and $y_i(t) > 0$. We let $t \to \infty$, and obtain, using (6),

$$\begin{aligned} \ln\left(\frac{y_i(0)}{y_i(\infty)}\right) &= \sigma_i \beta b \hat{x} \\ &= \sigma_i \beta b V^{-1} \Pi(y(0) - y_\infty) + \sigma_i \beta b V^{-1} x(0). \end{aligned} \tag{8}$$

This implies that $y_i(\infty) > 0$ for $i = 1, 2, \ldots, m$.

If we define the m-dimensional row vector

$$\Gamma = [\Gamma_1, \Gamma_2, \cdots \Gamma_m] = \beta b V^{-1} \Pi D,$$

then, from Theorem 2.1,

$$\mathcal{R}_0 = \Gamma y(0)$$

and we may rewrite (8) as

$$\ln\left(\frac{y_i(0)}{y_i(\infty)}\right) = \sigma_i \Gamma D^{-1}(y(0) - y_\infty) + \sigma_i \beta b V^{-1} x(0), \quad i = 1, 2, \ldots, m.$$

Then

$$\frac{1}{\sigma_i} \ln\left(\frac{y_i(0)}{y_i(\infty)}\right) = \Gamma D^{-1}(y(0) - y_\infty) + \beta b V^{-1} x(0) = \frac{1}{\sigma_1} \ln\left(\frac{y_1(0)}{y_1(\infty)}\right) \tag{9}$$

for $i = 1, 2, \cdots, m$. Thus

$$y_i(\infty) = y_i(0) \left(\frac{y_1(\infty)}{y_1(0)}\right)^{\sigma_i/\sigma_1},$$

and we may express $y_i(\infty)$ in terms of $y_1(\infty)$ for $i = 1, \cdots, m$, then substitute into (9) with $i = 1$ and solve to determine $y_1(\infty)$.

We summarize our results for the epidemic model.

THEOREM 5.1. *Consider the epidemic model* (3). *Then* $x_\infty = 0$ *and the final size relation is given by* (9) *if* $y_i(0) > 0$.

The multi-dimensional form is related to a result in [15]. If $m = 1$, that is, if there is only one susceptible class, so that y, D and $\Gamma = \mathcal{R}_0/y(0)$ are scalars and $D = 1, \sigma_1 = 1$, (9) takes the form

$$\ln\left(\frac{y(0)}{y_\infty}\right) = \frac{\mathcal{R}_0}{y(0)}[y(0) - y_\infty] + \beta b V^{-1} x(0). \tag{10}$$

This is the well-known Kermack-McKendrick form [10]; it has also been used with the initial term $\beta b V^{-1} x(0)$ approximated by zero; see for example [7, Sec. 1.3]. We note that (10) holds both for $\mathcal{R}_0 < 1$ and $\mathcal{R}_0 > 1$; the equation

$$\ln\left(\frac{y(0)}{y(t)}\right) = \frac{\mathcal{R}_0}{y(0)}[y(0) - y(t)] + \beta b V^{-1} x(0)$$

describes the orbit of the solution of the model in the phase space.

For a contact rate that is more general than mass action incidence β is a non-increasing function $\beta(N)$ of total population size N, so that

$$\beta(K) \le \beta(N(t)) \le \beta(N_\infty)$$

on the solution curve, where K is the initial total population size and $N_\infty > 0$ is the limiting total population size as $t \to \infty$, with $K \ge N_\infty$. More generally, if infectives withdraw partially from contact, β may be a function of both total population size and infective population size [4]. Then instead of (9), obtained with mass action incidence, we obtain a set of inequalities

$$\ln\left(\frac{y_i(0)}{y_i(\infty)}\right) \ge \sigma_i \Gamma D^{-1}(y(0) - y_\infty) + \sigma_i \beta(K) b V^{-1} x(0), \quad i = 1, 2, \ldots, m,$$

giving lower bounds for $\ln(y_i(0)/y_i(\infty))$. In addition, there are upper bounds involving the vectors $\beta(N_\infty)b$ and Γ. It is possible to prove that if the disease mortality rate is small, so that the contact rate does not change much over the course of the epidemic, then for realistic transmission terms the differences between the upper and lower bounds are small so that $y_i(\infty)$ can be approximated by treating the final size relations as if they were equalities.

Immunity against re-infection is often only temporary. It is not difficult to formulate models of $SLIRS$ type and to calculate the basic reproduction number. However, for epidemic models with loss of immunity there can be an endemic equilibrium even without demographic effects and it is not necessarily true that the number of infectives always tends to zero. Thus, there can not be a final size relation for such models.

6. **Immigration in epidemic models.** Normally an epidemic is triggered by the introduction of an infective into a wholly susceptible population, and the classical models for epidemics have incorporated this in the initial conditions. However, especially in the modern era with frequent global travel, it may be more realistic to assume a continuing flow of infectives into the population. Such models have been

studied in [5]. We model this by replacing (3) by

$$\begin{aligned} x' &= \Pi D y \beta(x,y,z) b x - V x + x^* \\ y' &= -D y \beta(x,y,z) b x \\ z' &= W x, \end{aligned} \tag{11}$$

with the vector x^* of infected immigrants having non-negative components with at least one positive component, and with non-negative initial values for $x(0)$. We assume that $\hat{x}^*$, representing the total number of infected immigrants during the epidemic, is finite.

The calculation of $\mathcal{R}_0$ in Section 2 can not be applied because the model (11) does not have a disease-free equilibrium. Nevertheless, we may carry out the same calculation formally and use its result in developing the final size relation. There is a threshold behavior, as shown for a special case in [5].

The derivation of the final size relation parallels that of Section 5. The equation (4) is replaced by

$$(x + \Pi y)' = -Vx + x^*,$$

and (5) becomes

$$(\hat{x}^* + x(0) - x_\infty) + \Pi(y(0) - y_\infty) = V\hat{x}. \tag{12}$$

Because of the assumption $\hat{x}^* < \infty$, the left side of (12) is finite, and from this we conclude as in Section 4 that $\hat{x} < \infty$. Since each component of x is a smooth non-negative function, $x_\infty = 0$, and

$$\hat{x} = V^{-1}\Pi(y(0) - y_\infty) + V^{-1}[x(0) + \hat{x}^*].$$

Thus $x(0)$ is replaced by $x(0) + \hat{x}^*$ in (9). In other words, in the final size relation the flow of infectives into the population during the epidemic is added to the initial infective population. A convenient way to assure that $\hat{x}^*$ remains finite is to cut off the flow of infected immigrants when the epidemic has passed. However, numerical simulations indicate that continuing the flow indefinitely does not significantly alter the shape of the epidemic curve.

7. **Examples.** In this section we describe several epidemic model examples to illustrate the calculation of the basic reproduction number and the final size relation. The calculation of the basic reproduction number requires only matrix multiplications and the inversion of the matrix V. Since V is usually in block lower triangular form, this matrix inversion can be carried out relatively simply. As in Section 3 we assume mass action incidence so that $\beta(x,y,z)$ is constant, but there is no difficulty in extending the results to general incidence.

7.1. **An influenza model.** As a first example, we consider a model for influenza described in [2] based on the disease properties described in [12]. In this model, which we call an $SLIAR$ model, there is a latent stage in which there is some reduced infectivity (represented by a factor $\varepsilon < 1$). A fraction p of the members of the latent stage go to an infective stage, while the remainder go to an asymptomatic stage in which there is some reduced infectivity (represented by a factor $\delta < 1$). There may be some deaths due to disease in the infective stage; we assume that a fraction f of infectives recover. The mean latent, infective, and asymptomatic

periods are assumed to be $1/\kappa, 1/\alpha$ and $1/\eta$, respectively. The model may be described by the system

$$\begin{aligned} S' &= -S\beta[I + \varepsilon L + \delta A] \\ L' &= S\beta[I + \varepsilon L + \delta A] - \kappa L \\ I' &= p\kappa L - \alpha I \\ A' &= (1-p)\kappa L - \eta A \\ R' &= f\alpha I + \eta A \\ N' &= -(1-f)\alpha I \end{aligned} \tag{13}$$

with initial conditions

$$S(0) = S_0, \quad I(0) = I_0, \quad L(0) = A(0) = R(0) = 0,$$

where $S_0 + I_0 = K$. We consider the disease-free equilibrium

$$S = S_0, \quad L = I = A = R = 0.$$

In terms of our notation, $m = 1, n = 3$, D is the scalar 1,

$$b = [\varepsilon, 1, \delta], \quad F = \begin{bmatrix} \varepsilon S_0\beta & S_0\beta & \delta S_0\beta \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \Pi = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Also,

$$V = \begin{bmatrix} \kappa & 0 & 0 \\ -p\kappa & \alpha & 0 \\ -(1-p)\kappa & 0 & \eta \end{bmatrix}, \quad V^{-1} = \begin{bmatrix} 1/\kappa & 0 & 0 \\ p/\alpha & 1/\alpha & 0 \\ (1-p)/\eta & 0 & 1/\eta \end{bmatrix}.$$

Then from Theorem 2.1,

$$\mathcal{R}_0 = S_0\beta\left[\frac{\varepsilon}{\kappa} + \frac{p}{\alpha} + \frac{\delta(1-p)}{\eta}\right],$$

and from Theorem 5.1,

$$\ln\frac{S_0}{S_\infty} = \mathcal{R}_0\frac{S_0 - S_\infty}{S_0} + \frac{\beta I_0}{\alpha}.$$

More generally, if the initial conditions were

$$L(0) = L_0, \quad I(0) = I_0, \quad A(0) = A_0,$$

the initial term $\beta I_0/\alpha$ would be replaced in the final size relation by

$$\beta\left[\frac{\varepsilon}{\kappa} + \frac{p}{\alpha} + \frac{\delta(1-p)}{\eta}\right]L_0 + \frac{\beta\delta A_0}{\eta} + \frac{\beta I_0}{\alpha}.$$

A refinement of the model (13), as suggested in [12] includes the possibility that the contact rate β may be a non-increasing function $\beta(N)$ of total population size and also withdrawal of a fraction q of infectives from contact, so that the average contact rate of infectives is multiplied by a factor $1 - q$. Then

$$\Psi(N, I) = \frac{N\beta(N)}{N - qI}$$

is the replacement for β in the model equations. In our notation,

$$\beta b = [\epsilon\Psi, (1-q)\Psi, \delta\Psi]$$

and at the equilibrium $S = S_0, L = I = A = R = 0, \Psi = \beta(S_0)$. Thus, at this equilibrium

$$b = [\epsilon, (1-q), \delta].$$

This gives

$$\mathcal{R}_0 = S_0\beta(S_0)\left[\frac{\varepsilon}{\kappa} + \frac{p(1-q)}{\alpha} + \frac{\delta(1-p)}{\eta}\right],$$

and from Theorem 5.1,

$$\ln\frac{S_0}{S_\infty} \geq \mathcal{R}_0\frac{S_0 - S_\infty}{S_0} + \frac{(1-q)\beta(S_0)I_0}{\alpha}.$$

7.2. **A SARS model.** A second example, based on a model for SARS [9] includes quarantine of latent members and isolation of diagnosed infectives. Members of the latent compartment are moved to a quarantined compartment Q at rate γ_1 while members of the quarantined class go directly to an isolated compartment J when they become infective. In addition, infectives are isolated at rate γ_2. Infectivity is assumed multiplied by factors $\varepsilon_L, \varepsilon_Q, \varepsilon_J$, each less than 1, respectively, in the classes L, Q, J. The mean latent and quarantine periods are $1/\kappa_1, 1/\kappa_2$ respectively. The mean infective and isolated periods are $1/\alpha_1, 1/\alpha_2$ respectively. The quarantine and isolation rates are γ_1, γ_2 respectively, and the fractions that recover in the infective and isolated compartments are f_1, f_2 respectively. The model is

$$\begin{aligned}
S' &= -\beta S\Lambda \\
L' &= \beta S\Lambda - (\gamma_1 + \kappa_1)L \\
Q' &= \gamma_1 L - \kappa_2 Q \\
I' &= \kappa_1 L - (\gamma_2 + \alpha_1)I \\
J' &= \kappa_2 Q + \gamma_2 I - \alpha_2 J \\
R' &= f_1\alpha_1 I + f_2\alpha_2 J,
\end{aligned}$$

with

$$\Lambda = \varepsilon_L L + \varepsilon_Q Q + I + \varepsilon_J J,$$

and initial conditions

$$S(0) = S_0, \quad I(0) = I_0, \quad L(0) = Q(0) = J(0) = R(0) = 0.$$

In terms of our notation, $m = 1, n = 4$, D is the scalar 1,

$$b = [\varepsilon_L, \varepsilon_Q, 1, \varepsilon_J], \quad \Pi = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

and

$$V = \begin{bmatrix} \gamma_1 + \kappa_1 & 0 & 0 & 0 \\ -\gamma_1 & \kappa_2 & 0 & 0 \\ -\kappa_1 & 0 & \gamma_2 + \alpha_1 & 0 \\ 0 & -\kappa_2 & -\gamma_2 & \alpha_2 \end{bmatrix},$$

$$V^{-1} = \begin{bmatrix} \frac{1}{\gamma_1+\kappa_1} & 0 & 0 & 0 \\ \frac{\gamma_1}{\kappa_2(\gamma_1+\kappa_1)} & \frac{1}{\kappa_2} & 0 & 0 \\ \frac{\kappa_1}{(\gamma_2+\alpha_1)(\gamma_1+\kappa_1)} & 0 & \frac{1}{\gamma_2+\alpha_1} & 0 \\ \frac{\kappa_1\gamma_2}{\alpha_2(\gamma_2+\alpha_1)(\gamma_1+\kappa_1)} + \frac{\gamma_1}{\alpha_2(\gamma_1+\kappa_1)} & \frac{1}{\alpha_2} & \frac{\gamma_2}{\alpha_2(\gamma_2+\alpha_1)} & \frac{1}{\alpha_2} \end{bmatrix}.$$

Since this model includes quarantine and isolation arranged prior to the beginning of the epidemic, we call the corresponding reproduction number the control reproduction number, denoted by $\mathcal{R}_c$ [9] rather than the basic reproduction number. The basic reproduction number would correspond to the choices $\gamma_1 = \gamma_2 = 0$ so that there are no quarantined or isolated compartments. We calculate from Theorem 2.1

$$\mathcal{R}_c = \frac{S_0\beta}{\gamma_1+\kappa_1}\left[\varepsilon_L + \frac{\kappa_1}{\gamma_2+\alpha_1} + \frac{\varepsilon_Q\gamma_1}{\kappa_2} + \varepsilon_J\left(\frac{\kappa_1\gamma_2}{\alpha_2(\gamma_2+\alpha_1)} + \frac{\gamma_1}{\alpha_2}\right)\right].$$

This is similar to the control reproduction number calculated in [9], (where demographic terms are included and standard incidence is used). From Theorem 5.1 the final size relation is

$$\ln\frac{S_0}{S_\infty} = \mathcal{R}_c\frac{S_0-S_\infty}{S_0} + \beta I_0\left[\frac{1}{\gamma_2+\alpha_1} + \frac{\varepsilon_J\gamma_2}{\alpha_2(\gamma_2+\alpha_1)}\right].$$

7.3. **A vaccination model.** As a final example, we consider an $SLIR$ model in which a fraction γ of susceptibles have been vaccinated before the beginning of an epidemic with a vaccine that multiplies susceptibility by a factor $\sigma_S \le 1$ and infectivity by a factor $\sigma_I \le 1$. This model is a simplified version of an influenza model analyzed in [2]. We assume that the mean latent periods for unvaccinated and vaccinated individuals are $1/\kappa, 1/\kappa_T$ respectively and that the mean infective periods for unvaccinated and vaccinated individuals are $1/\alpha, 1/\alpha_T$ respectively. The recovery fractions are f for unvaccinated infectives and f_T for vaccinated infectives. We let S, S_T be the number of unvaccinated and vaccinated susceptibles respectively, L, L_T be the number of unvaccinated and vaccinated latent members respectively, and I, I_T the number of unvaccinated and vaccinated infectives respectively. The model is described by the system

$$\begin{aligned} S' &= -\beta S[I+\sigma_I I_T] \\ S_T' &= -\sigma_S\beta S[I+\sigma_I I_T] \\ L' &= \beta S_T[I+\sigma_I I_T] - \kappa L \\ L_T' &= \sigma_S\beta S_T[I+\sigma_I I_T] - \kappa_T L_T \\ I' &= \kappa L - \alpha I \\ I_T' &= \kappa_T L_T - \alpha_T I_T \\ R' &= f\alpha I + f_T\alpha_T I, \end{aligned}$$

with

$$S(0) = (1-\gamma)S_0, \quad S_T(0) = \gamma S_0, \quad I(0) = I_0, \quad L(0) = L_T(0) = I_T(0) = R(0) = 0.$$

In terms of our notation, $m = 2, n = 4$,

$$b = [0, 0, 1, \sigma_I] \quad D = \begin{bmatrix} 1 & 0 \\ 0 & \sigma_S \end{bmatrix} \quad \Pi = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

$$V = \begin{bmatrix} \kappa & 0 & 0 & 0 \\ 0 & \kappa_T & 0 & 0 \\ -\kappa & 0 & \alpha & 0 \\ 0 & -\kappa_T & 0 & \alpha_T \end{bmatrix} \quad V^{-1} = \begin{bmatrix} \frac{1}{\kappa} & 0 & 0 & 0 \\ 0 & \frac{1}{\kappa_T} & 0 & 0 \\ \frac{1}{\alpha} & 0 & \frac{1}{\alpha} & 0 \\ 0 & \frac{1}{\alpha_T} & 0 & \frac{1}{\alpha_T} \end{bmatrix}.$$

Then we calculate

$$\Gamma = \left[\frac{\beta}{\alpha}, \frac{\sigma_I \sigma_S \beta}{\alpha_T}\right], \quad \mathcal{R}_c = S_0 \beta \left[\frac{1-\gamma}{\alpha} + \frac{\sigma_I \sigma_S \gamma}{\alpha_T}\right],$$

and from (9) the final size relation is

$$\begin{aligned} \ln\left(\frac{(1-\gamma)S_0}{S_\infty}\right) &= \frac{\beta}{\alpha}[(1-\gamma)S_0 - S_\infty] + \frac{\sigma_I \beta}{\alpha_T}[\gamma S_0 - S_{T\infty}] + \frac{\beta I_0}{\alpha}, \\ S_{T\infty} &= \gamma S_0 \left(\frac{S_\infty}{(1-\gamma)S_0}\right)^{\sigma_S}. \end{aligned}$$

We have formulated an influenza model based on this example in [2] to compare the effects of different management strategies in attempting to manage a threatened pandemic influenza.

8. **A simple heterogeneous mixing epidemic model.** The assumption of homogeneous mixing in epidemic models is usually quite unrealistic. Frequently there are superspreaders, who make many contacts and are instrumental in spreading disease. To model this heterogeneity in mixing we may assume that the population is divided into subgroups with different activity levels. A simple compartmental model may capture some of the essential properties of an epidemic model without going to a full network model.

Consider two subpopulations of sizes N_1, N_2 respectively, each divided into susceptibles, infectives, and removed members with subscripts to identify the subpopulation. Suppose that group i members make a_i contacts in unit time and that the fraction of contacts made by a member of group i that is with a member of group j is $p_{ij}, i, j = 1, 2$. Then

$$p_{11} + p_{12} = p_{21} + p_{22} = 1.$$

The total number of contacts made by members of group 1 with members of group 2 is $a_1 p_{12} N_1$ and because this must equal the total number of contacts by members of group 2 with members of group 1, we have a balance relation

$$\frac{p_{12} a_1}{N_2} = \frac{p_{21} a_2}{N_1}.$$

Suppose the mean infective periods in the two groups are $1/\alpha_1, 1/\alpha_2$ and the recovery fractions in the two groups are f_1, f_2. Then the two-group SIR epidemic

model is

$$\begin{aligned} S_1' &= -[p_{11}a_1\frac{S_1I_1}{N_1} + p_{12}a_1\frac{S_1I_2}{N_2}] \\ I_1' &= [p_{11}a_1\frac{S_1I_1}{N_1} + p_{12}a_1\frac{S_1I_2}{N_2}] - \alpha_1 I_1 \\ R_1' &= f_1\alpha_1 I_1 \\ S_2' &= -[p_{21}a_2\frac{S_2I_1}{N_1} + p_{22}a_2\frac{S_2I_2}{N_2}] \\ I_2' &= [p_{21}a_2\frac{S_2I_1}{N_1} + p_{22}a_2\frac{S_2I_2}{N_2}] - \alpha_2 I_2 \\ R_2' &= f_2\alpha_2 I_2. \end{aligned} \tag{14}$$

We further assume proportionate mixing between groups, that is, that the number of contacts between groups is proportional to the relative activity levels. In other words, mixing is random but constrained by the activity levels [14]. Then

$$p_{ij} = \frac{a_j N_j}{a_1 N_1 + a_2 N_2},$$

and we may write

$$p_{11} = p_{21} = p_1, \quad p_{12} = p_{22} = p_2,$$

with $p_1 + p_2 = 1$. In other words, proportionate mixing means that each group makes a fraction p_j of its contacts with group j for $j = 1, 2$.

The model (14) becomes

$$\begin{aligned} S_1' &= -[p_1a_1\frac{S_1I_1}{N_1} + p_2a_1\frac{S_1I_2}{N_2}] \\ I_1' &= [p_1a_1\frac{S_1I_1}{N_1} + p_2a_1\frac{S_1I_2}{N_2}] - \alpha_1 I_1 \\ R_1' &= f_1\alpha_1 I_1 \\ S_2' &= -[p_1a_2\frac{S_2I_1}{N_1} + p_2a_2\frac{S_2I_2}{N_2}] \\ I_2' &= [p_1a_2\frac{S_2I_1}{N_1} + p_2a_2\frac{S_2I_2}{N_2}] - \alpha_2 I_2 \\ R_2' &= f_2\alpha_2 I_2. \end{aligned} \tag{15}$$

We now write

$$x = \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}, \quad y = \begin{bmatrix} S_1 \\ S_2 \end{bmatrix}, \quad z = \begin{bmatrix} R_1 \\ R_2 \end{bmatrix}.$$

We put the system (15) into a form

$$\begin{aligned} x' &= \Pi DQy\beta bx - Vx \\ y' &= -\Pi DQy\beta bx \\ z' &= Wx, \end{aligned} \tag{16}$$

with Π and D identity matrices and

$$Q = \begin{bmatrix} \frac{a_1}{N_1} & 0 \\ 0 & \frac{a_2}{N_2} \end{bmatrix}, \quad \beta b = [p_1, p_2],$$

and

$$V = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix}, \quad V^{-1} = \begin{bmatrix} \frac{1}{\alpha_1} & 0 \\ 0 & \frac{1}{\alpha_2} \end{bmatrix}.$$

Then, as in Section 2,

$$\mathcal{R}_0 = tr(Qy\beta bV^{-1}) = \beta bV^{-1}Qy,$$

with the variable matrix Q calculated at $t = 0$ and the vector y calculated at a disease-free equilibrium (since $F = Qy\beta b$ has rank 1). Using

$$p_1 = \frac{a_1N_1}{a_1N_1 + a_2N_2}, \quad p_2 = \frac{a_2N_2}{a_1N_1 + a_2N_2},$$

this gives

$$\mathcal{R}_0 = \frac{a_1^2N_1}{\alpha_1(a_1N_1 + a_2N_2)} + \frac{a_2^2N_1}{\alpha_2(a_1N_1 + a_2N_2)}.$$

The general formulation of a heterogeneous mixing epidemic model with an arbitrary number of activity levels and proportionate mixing has the form

$$\begin{aligned} x' &= \Pi DQy\beta(x, y, z)bx - Vx \\ y' &= -DQy\beta(x, y, z)bx \\ z' &= Wx, \end{aligned}$$

with Q the $m \times m$ diagonal matrix whose j, j entry is the fraction of contacts p_j that each group (including group j) makes with group j. To calculate the basic reproduction number we must evaluate Q, y and $\beta(x, y, z)$ at $(0, y_0, z_0)$ and we obtain

$$\mathcal{R}_0 = tr(DQy_0\beta(0, y_0, z_0)bV^{-1}) = \beta(0, y_0, z_0)bV^{-1}DQy_0.$$

The final size relation may also be calculated as in Section 5, but is an equality only if the matrix Q is constant. This means that the population sizes N_1, N_2 must be constant, that is, there are no disease deaths. However, if the number of disease deaths is small, we conjecture that the final size relation is an approximate inequality.

If the mixing is not proportionate, it is not possible to write the model in the form (16). It is known [7] that the final size relation may then take a different form. However, proportionate mixing seems to be a plausible assumption for an infection spread by random contacts.

9. **Conclusions.** The basic reproduction number defined in [7, 8] is characterized as the spectral radius of a matrix that in many applications has rank 1. We have used this to express the basic reproduction number explicitly as a product of matrices, thus simplifying the calculation.

We have shown how to calculate the number of members of each susceptible compartment that escape infection over the course of the epidemic and have illustrated our results with models for influenza and SARS with disease control measures as well as for models with vaccination and heterogeneous mixing. For general incidence, the final size relations are inequalities, but they become equalities for mass action incidence. This suggests that mass action is the natural setting for final size equations, and points to the importance of results to the effect that the final size relations are approximate equalities if disease mortality is small.

The results in this paper do not include diseases transmitted by a vector, and it would be of interest to see if our approach can be adapted to obtain analogous results for vector-borne diseases. As the vectors in epidemic diseases often have short life spans, study of an epidemic transmitted by a vector may require inclusion of demographic effects in the vector population.

Acknowledgements. This work was supported by MITACS (Mathematics of Information Technology and Complex Systems (MITACS) and NSERC (The National Science and Engineering Research Council of Canada).

REFERENCES

[1] R.M. Anderson & R.M. May, Infectious Diseases of Humans: Dynamics and Control, Oxford University Press, Oxford (1991)
[2] J. Arino, F. Brauer, P. van den Driessche, J. Watmough & J. Wu, Simple models for containment of a pandemic, J. Roy. Soc. Interface **3**, 453–457 (2006).
[3] A. Berman & R.J. Plemmons, Nonnegative Matrices in the Mathematical Sciences, Classics in Applied Mathematics **9**, SIAM, Philadelphia (1994).
[4] F. Brauer, Some simple epidemic models, Math. Biosc. & Eng., **3**, 1–15 (2006).
[5] F. Brauer & P. van den Driessche, Models for transmission of disease with immigration of infectives, Math. Biosc. **171**, 143–154 (2001).
[6] S. Busenberg & K.L. Cooke, Vertically Transmitted Diseases: Models and Dynamics, Biomathematics **23**, Springer-Verlag, Berlin-Heidelberg-New York (1993).
[7] O. Diekmann & J.A.P. Heesterbeek, Mathematical Epidemiology of Infectious Diseases, Wiley, Chichester (2000).
[8] O. Diekmann, J.A.P. Heesterbeek & J.A.J. Metz, On the definition and the computation of the basic reproductive ratio $\mathcal{R}_0$ in models for infectious diseases in heterogeneous populations, J. Math. Biol. **28**, 365–382 (1990).
[9] A. Gumel, S. Ruan, T. Day, J. Watmough, F. Brauer, P. van den Driessche, D. Gabrielson, C. Bowman, M.E. Alexander, S. Ardal, J. Wu, & B.M. Sahai, Modelling strategies for controlling SARS outbreaks, Proc. Roy. Soc. London B **271**, 2223–2232 (2004).
[10] W.O. Kermack & A.G. McKendrick, A contribution to the mathematical theory of epidemics, Proc. Royal Soc. London **115**, 700–721 (1927).
[11] M.Y. Li, H.L. Smith & L. Wang, Global dynamics of an SEIR epidemic model with vertical transmission, SIAM J. Appl. Math. **62**, 58–69 (2001).
[12] I.M. Longini, M.E. Halloran, A. Nizam, & Y. Yang, Containing pandemic influenza with antiviral agents, Am. J. Epidem. **159**, 623–633 (2004).
[13] J. Ma & D.J.D. Earn, Generality of the final size formula for an epidemic of a newly invading infectious disease, Bull. Math. Biol. **68**, 679–702 (2006).
[14] A. Nold, Heterogeneity in disease transmission modeling, Math. Biosc. **52**, 227–240 (1980).
[15] L. Rass & J. Radcliffe, Spatial Deterministic Epidemics, Mathematical Surveys and Monographs **103**, AMS, Providence, R.I. (2003).
[16] P. van den Driessche & J. Watmough, Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission, Math. Biosc. **180**, 29–48 (2002).

Received on May 8, 2006. Accepted on October 16, 2006.

E-mail address: arinoj@cc.umanitoba.ca
E-mail address: brauer@math.ubc.ca
E-mail address: pvdd@math.uvic.ca
E-mail address: watmough@unb.ca
E-mail address: wujh@mathstat.yorku.ca