ELSEVIER

Mathematical Biosciences 180 (2002) 29–48

Mathematical Biosciences

www.elsevier.com/locate/mbs

Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission

P. van den Driessche [a,1], James Watmough [b,*,2]

[a] *Department of Mathematics and Statistics, University of Victoria, Victoria, BC, Canada V8W 3P4*

[b] *Department of Mathematics and Statistics, University of New Brunswick, Fredericton, NB, Canada E3B 5A3*

Received 26 April 2001; received in revised form 27 June 2001; accepted 27 June 2001

Dedicated to the memory of John Jacquez

Abstract

A precise definition of the basic reproduction number, $\mathscr{R}_0$, is presented for a general compartmental disease transmission model based on a system of ordinary differential equations. It is shown that, if $\mathscr{R}_0 < 1$, then the disease free equilibrium is locally asymptotically stable; whereas if $\mathscr{R}_0 > 1$, then it is unstable. Thus, $\mathscr{R}_0$ is a threshold parameter for the model. An analysis of the local centre manifold yields a simple criterion for the existence and stability of super- and sub-threshold endemic equilibria for $\mathscr{R}_0$ near one. This criterion, together with the definition of $\mathscr{R}_0$, is illustrated by treatment, multigroup, staged progression, multistrain and vector–host models and can be applied to more complex models. The results are significant for disease control.

Keywords: Basic reproduction number; Sub-threshold equilibrium; Disease transmission model; Disease control

1. Introduction

One of the most important concerns about any infectious disease is its ability to invade a population. Many epidemiological models have a disease free equilibrium (DFE) at which the

* Corresponding author. Tel.: +1-506 458 7323; fax: +1-506 453 4705.

E-mail addresses: pvdd@math.uvic.ca (P. van den Driessche), watmough@unb.ca (J. Watmough).

URL: http://www.math.unb.ca/~watmough.

[1] Research supported in part by an NSERC Research Grant, the University of Victoria Committee on faculty research and travel and MITACS.

[2] Research supported by an NSERC Postdoctoral Fellowship tenured at the University of Victoria.

0025-5564/02/$ - see front matter
PII: S0025-5564(02)00108-6

population remains in the absence of disease. These models usually have a threshold parameter, known as the basic reproduction number, $\mathscr{R}_0$, such that if $\mathscr{R}_0 < 1$, then the DFE is locally asymptotically stable, and the disease cannot invade the population, but if $\mathscr{R}_0 > 1$, then the DFE is unstable and invasion is always possible (see the survey paper by Hethcote [1]). Diekmann et al. [2] define $\mathscr{R}_0$ as the spectral radius of the next generation matrix. We write down in detail a general compartmental disease transmission model suited to heterogeneous populations that can be modelled by a system of ordinary differential equations. We derive an expression for the next generation matrix for this model and examine the threshold $\mathscr{R}_0 = 1$ in detail.

The model is suited to a heterogeneous population in which the vital and epidemiological parameters for an individual may depend on such factors as the stage of the disease, spatial position, age or behaviour. However, we assume that the population can be broken into homogeneous subpopulations, or compartments, such that individuals in a given compartment are indistinguishable from one another. That is, the parameters may vary from compartment to compartment, but are identical for all individuals within a given compartment. We also assume that the parameters do not depend on the length of time an individual has spent in a compartment. The model is based on a system of ordinary equations describing the evolution of the number of individuals in each compartment.

In addition to showing that $\mathscr{R}_0$ is a threshold parameter for the local stability of the DFE, we apply centre manifold theory to determine the existence and stability of endemic equilibria near the threshold. We show that some models may have unstable endemic equilibria near the DFE for $\mathscr{R}_0 < 1$. This suggests that even though the DFE is locally stable, the disease may persist.

The model is developed in Section 2. The basic reproduction number is defined and shown to be a threshold parameter in Section 3, and the definition is illustrated by several examples in Section 4. The analysis of the centre manifold is presented in Section 5. The epidemiological ramifications of the results are presented in Section 6.

2. A general compartmental epidemic model for a heterogeneous population

Consider a heterogeneous population whose individuals are distinguishable by age, behaviour, spatial position and/or stage of disease, but can be grouped into n homogeneous compartments. A general epidemic model for such a population is developed in this section. Let $x = (x_1, \ldots, x_n)^t$, with each $x_i \geqslant 0$, be the number of individuals in each compartment. For clarity we sort the compartments so that the first m compartments correspond to infected individuals. The distinction between infected and uninfected compartments must be determined from the epidemiological interpretation of the model and cannot be deduced from the structure of the equations alone, as we shall discuss below. It is plausible that more than one interpretation is possible for some models. A simple epidemic model illustrating this is given in Section 4.1. The basic reproduction number can not be determined from the structure of the mathematical model alone, but depends on the definition of infected and uninfected compartments. We define $\mathbf{X}_s$ to be the set of all disease free states. That is

$$\mathbf{X}_s = \{x \geqslant 0 | x_i = 0,\ i = 1, \ldots, m\}.$$

In order to compute $\mathscr{R}_0$, it is important to distinguish new infections from all other changes in population. Let $\mathscr{F}_i(x)$ be the rate of appearance of new infections in compartment i, $\mathscr{V}_i^+(x)$ be the rate of transfer of individuals into compartment i by all other means, and $\mathscr{V}_i^-(x)$ be the rate of transfer of individuals out of compartment i. It is assumed that each function is continuously differentiable at least twice in each variable. The disease transmission model consists of non-negative initial conditions together with the following system of equations:

$$\dot{x}_i = f_i(x) = \mathscr{F}_i(x) - \mathscr{V}_i(x), \quad i = 1, \ldots, n, \tag{1}$$

where $\mathscr{V}_i = \mathscr{V}_i^- - \mathscr{V}_i^+$ and the functions satisfy assumptions (A1)–(A5) described below. Since each function represents a directed transfer of individuals, they are all non-negative. Thus,

(A1) if $x \geqslant 0$, then $\mathscr{F}_i, \mathscr{V}_i^+, \mathscr{V}_i^- \geqslant 0$ for $i = 1, \ldots, n$.

If a compartment is empty, then there can be no transfer of individuals out of the compartment by death, infection, nor any other means. Thus,

(A2) if $x_i = 0$ then $\mathscr{V}_i^- = 0$. In particular, if $x \in \mathbf{X}_s$ then $\mathscr{V}_i^- = 0$ for $i = 1, \ldots, m$.

Consider the disease transmission model given by (1) with $f_i(x)$, $i = 1, \ldots, n$, satisfying conditions (A1) and (A2). If $x_i = 0$, then $f_i(x) \geqslant 0$ and hence, the non-negative cone ($x_i \geqslant 0$, $i = 1, \ldots, n$) is forward invariant. By Theorems 1.1.8 and 1.1.9 of Wiggins [3, p. 37] for each non-negative initial condition there is a unique, non-negative solution.

The next condition arises from the simple fact that the incidence of infection for uninfected compartments is zero.

(A3) $\mathscr{F}_i = 0$ if $i > m$.

To ensure that the disease free subspace is invariant, we assume that if the population is free of disease then the population will remain free of disease. That is, there is no (density independent) immigration of infectives. This condition is stated as follows:

(A4) if $x \in \mathbf{X}_s$ then $\mathscr{F}_i(x) = 0$ and $\mathscr{V}_i^+(x) = 0$ for $i = 1, \ldots, m$.

The remaining condition is based on the derivatives of f near a DFE. For our purposes, we define a DFE of (1) to be a (locally asymptotically) stable equilibrium solution of the disease free model, i.e., (1) restricted to $\mathbf{X}_s$. Note that we need not assume that the model has a unique DFE. Consider a population near the DFE x_0. If the population remains near the DFE (i.e., if the introduction of a few infective individuals does not result in an epidemic) then the population will return to the DFE according to the linearized system

$$\dot{x} = Df(x_0)(x - x_0), \tag{2}$$

where $Df(x_0)$ is the derivative $[\partial f_i/\partial x_j]$ evaluated at the DFE, x_0 (i.e., the Jacobian matrix). Here, and in what follows, some derivatives are one sided, since x_0 is on the domain boundary. We restrict our attention to systems in which the DFE is stable in the absence of new infection. That is,

(A5) If $\mathscr{F}(x)$ is set to zero, then all eigenvalues of $Df(x_0)$ have negative real parts.

The conditions listed above allow us to partition the matrix $Df(x_0)$ as shown by the following lemma.

Lemma 1. *If x_0 is a DFE of* (1) *and $f_i(x)$ satisfies* (A1)–(A5), *then the derivatives $D\mathscr{F}(x_0)$ and $D\mathscr{V}(x_0)$ are partitioned as*

$$D\mathscr{F}(x_0) = \begin{pmatrix} F & 0 \\ 0 & 0 \end{pmatrix}, \quad D\mathscr{V}(x_0) = \begin{pmatrix} V & 0 \\ J_3 & J_4 \end{pmatrix},$$

where F and V are the $m \times m$ matrices defined by

$$F = \left[\frac{\partial \mathscr{F}_i}{\partial x_j}(x_0)\right] \quad \text{and} \quad V = \left[\frac{\partial \mathscr{V}_i}{\partial x_j}(x_0)\right] \quad \text{with } 1 \leqslant i,\ j \leqslant m.$$

Further, F is non-negative, V is a non-singular M-matrix and all eigenvalues of J_4 have positive real part.

Proof. Let $x_0 \in \mathbf{X}_s$ be a DFE. By (A3) and (A4), $(\partial \mathscr{F}_i/\partial x_j)(x_0) = 0$ if either $i > m$ or $j > m$. Similarly, by (A2) and (A4), if $x \in \mathbf{X}_s$ then $\mathscr{V}_i(x) = 0$ for $i \leqslant m$. Hence, $(\partial \mathscr{V}_i/\partial x_j)(x_0) = 0$ for $i \leqslant m$ and $j > m$. This shows the stated partition and zero blocks. The non-negativity of F follows from (A1) and (A4).

Let $\{e_j\}$ be the Euclidean basis vectors. That is, e_j is the jth column of the $n \times n$ identity matrix. Then, for $j = 1, \ldots, m$,

$$\left(\frac{\partial \mathscr{V}_i}{\partial x_j}\right)(x_0) = \lim_{h \to 0^+} \left(\frac{\mathscr{V}_i(x_0 + he_j) - \mathscr{V}_i(x_0)}{h}\right).$$

To show that V is a non-singular M-matrix, note that if x_0 is a DFE, then by (A2) and (A4), $\mathscr{V}_i(x_0) = 0$ for $i = 1, \ldots, m$, and if $i \neq j$, then the ith component of $x_0 + he_j = 0$ and $\mathscr{V}_i(x_0 + he_j) \leqslant 0$, by (A1) and (A2). Hence, $\partial \mathscr{V}_i/\partial x_j \leq 0$ for $i \leq m$ and $j \neq i$ and V has the Z sign pattern (see Appendix A). Additionally, by (A5), all eigenvalues of V have positive real parts. These two conditions imply that V is a non-singular M-matrix [4, p. 135 (G_{20})]. Condition (A5) also implies that the eigenvalues of J_4 have positive real part. □

3. The basic reproduction number

The basic reproduction number, denoted $\mathscr{R}_0$, is 'the expected number of secondary cases produced, in a completely susceptible population, by a typical infective individual' [2]; see also [5, p. 17]. If $\mathscr{R}_0 < 1$, then on average an infected individual produces less than one new infected individual over the course of its infectious period, and the infection cannot grow. Conversely, if $\mathscr{R}_0 > 1$, then each infected individual produces, on average, more than one new infection, and the disease can invade the population. For the case of a single infected compartment, $\mathscr{R}_0$ is simply the product of the infection rate and the mean duration of the infection. However, for more complicated models with several infected compartments this simple heuristic definition of $\mathscr{R}_0$ is

insufficient. A more general basic reproduction number can be defined as the number of new infections produced by a typical infective individual in a population at a DFE.

To determine the fate of a 'typical' infective individual introduced into the population, we consider the dynamics of the linearized system (2) with reinfection turned off. That is, the system

$$\dot{x} = -D\mathscr{V}(x_0)(x - x_0). \tag{3}$$

By (A5), the DFE is locally asymptotically stable in this system. Thus, (3) can be used to determine the fate of a small number of infected individuals introduced to a disease free population. Let $\psi_i(0)$ be the number of infected individuals initially in compartment i and let $\psi(t) = (\psi_1(t), \ldots, \psi_m(t))^t$ be the number of these initially infected individuals remaining in the infected compartments after t time units. That is the vector ψ is the first m components of x. The partitioning of $D\mathscr{V}(x_0)$ implies that $\psi(t)$ satisfies $\psi'(t) = -V\psi(t)$, which has the unique solution $\psi(t) = \mathrm{e}^{-Vt}\psi(0)$. By Lemma 1, V is a non-singular M-matrix and is, therefore, invertible and all of its eigenvalues have positive real parts. Thus, integrating $F\psi(t)$ from zero to infinity gives the expected number of new infections produced by the initially infected individuals as the vector $FV^{-1}\psi(0)$. Since F is non-negative and V is a non-singular M-matrix, V^{-1} is non-negative [4, p. 137 (N_{38})], as is FV^{-1}.

To interpret the entries of FV^{-1} and develop a meaningful definition of $\mathscr{R}_0$, consider the fate of an infected individual introduced into compartment k of a disease free population. The (j,k) entry of V^{-1} is the average length of time this individual spends in compartment j during its lifetime, assuming that the population remains near the DFE and barring reinfection. The (i,j) entry of F is the rate at which infected individuals in compartment j produce new infections in compartment i. Hence, the (i,k) entry of the product FV^{-1} is the expected number of new infections in compartment i produced by the infected individual originally introduced into compartment k. Following Diekmann et al. [2], we call FV^{-1} the next generation matrix for the model and set

$$\mathscr{R}_0 = \rho(FV^{-1}), \tag{4}$$

where $\rho(A)$ denotes the spectral radius of a matrix A.

The DFE, x_0, is locally asymptotically stable if all the eigenvalues of the matrix $Df(x_0)$ have negative real parts and unstable if any eigenvalue of $Df(x_0)$ has a positive real part. By Lemma 1, the eigenvalues of $Df(x_0)$ can be partitioned into two sets corresponding to the infected and uninfected compartments. These two sets are the eigenvalues of $F - V$ and those of $-J_4$. Again by Lemma 1, the eigenvalues of $-J_4$ all have negative real part, thus the stability of the DFE is determined by the eigenvalues of $F - V$. The following theorem states that $\mathscr{R}_0$ is a threshold parameter for the stability of the DFE.

Theorem 2. *Consider the disease transmission model given by* (1) *with* $f(x)$ *satisfying conditions* (A1)–(A5). *If* x_0 *is a DFE of the model, then* x_0 *is locally asymptotically stable if* $\mathscr{R}_0 < 1$, *but unstable if* $\mathscr{R}_0 > 1$, *where* $\mathscr{R}_0$ *is defined by* (4).

Proof. Let $J_1 = F - V$. Since V is a non-singular M-matrix and F is non-negative, $-J_1 = V - F$ has the Z sign pattern (see Appendix A). Thus,

$$s(J_1) < 0 \iff -J_1 \text{ is a non-singular M-matrix},$$

where $s(J_1)$ denotes the maximum real part of all the eigenvalues of the matrix J_1 (the spectral abscissa of J_1). Since FV^{-1} is non-negative, $-J_1V^{-1} = I - FV^{-1}$ also has the Z sign pattern. Applying Lemma 5 of Appendix A, with $H = V$ and $B = -J_1 = V - F$, we have

$$-J_1 \text{ is a non-singular M-matrix} \iff I - FV^{-1} \text{ is a non-singular M-matrix.}$$

Finally, since FV^{-1} is non-negative, all eigenvalues of FV^{-1} have magnitude less than or equal to $\rho(FV^{-1})$. Thus,

$$I - FV^{-1} \text{ is a non-singular M-matrix, } \iff \rho(FV^{-1}) < 1.$$

Hence, $s(J_1) < 0$ if and only if $\mathscr{R}_0 < 1$.

Similarly, it follows that

$$\begin{aligned} s(J_1) = 0 &\iff -J_1 \text{ is a singular M-matrix,} \\ &\iff I - FV^{-1} \text{ is a singular M-matrix,} \\ &\iff \rho(FV^{-1}) = 1. \end{aligned}$$

The second equivalence follows from Lemma 6 of Appendix A, with $H = V$ and $K = F$. The remainder of the equivalences follow as with the non-singular case. Hence, $s(J_1) = 0$ if and only if $\mathscr{R}_0 = 1$. It follows that $s(J_1) > 0$ if and only if $\mathscr{R}_0 > 1$. □

A similar result can be found in the recent book by Diekmann and Heesterbeek [6, Theorem 6.13]. This result is known for the special case in which J_1 is irreducible and V is a positive diagonal matrix [7–10]. The special case in which V has positive diagonal and negative subdiagonal elements is proven in Hyman et al. [11, Appendix B]; however, our approach is much simpler (see Section 4.3).

4. Examples

4.1. Treatment model

The decomposition of $f(x)$ into the components $\mathscr{F}$ and $\mathscr{V}$ is illustrated using a simple treatment model. The model is based on the tuberculosis model of Castillo-Chavez and Feng [12, Eq. (1.1)], but also includes treatment failure used in their more elaborate two-strain model [12, Eq. (2.1)]. A similar tuberculosis model with two treated compartments is proposed by Blower et al. [13]. The population is divided into four compartments, namely, individuals susceptible to tuberculosis (S), exposed individuals (E), infectious individuals (I) and treated individuals (T). The dynamics are illustrated in Fig. 1. Susceptible and treated individuals enter the exposed compartment at rates $\beta_1 I/N$ and $\beta_2 I/N$, respectively, where $N = E + I + S + T$. Exposed individuals progress to the infectious compartment at the rate v. All newborns are susceptible, and all individuals die at the rate $d > 0$. Thus, the core of the model is an SEI model using standard incidence. The treatment rates are r_1 for exposed individuals and r_2 for infectious individuals. However, only a fraction q of the treatments of infectious individuals are successful. Unsuccessfully treated infectious individuals re-enter the exposed compartment ($p = 1 - q$). The disease

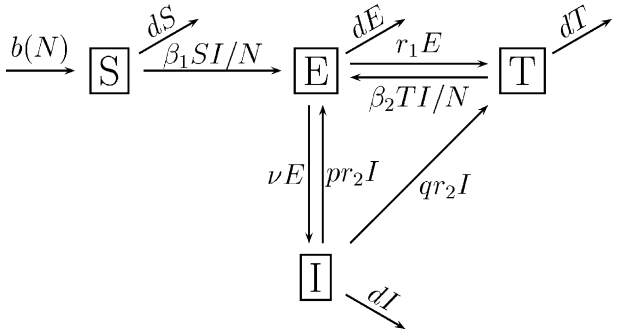

Fig. 1. Progression of infection from susceptible (S) individuals through the exposed (E), infected (I), and treated (T) compartments for the treatment model of (5a)–(5d).

transmission model consists of the following differential equations together with non-negative initial conditions:

$$\dot{E} = \beta_1 SI/N + \beta_2 TI/N - (d + v + r_1)E + pr_2 I, \tag{5a}$$

$$\dot{I} = vE - (d + r_2)I, \tag{5b}$$

$$\dot{S} = b(N) - dS - \beta_1 SI/N, \tag{5c}$$

$$\dot{T} = -dT + r_1 E + qr_2 I - \beta_2 TI/N. \tag{5d}$$

Progression from E to I and failure of treatment are not considered to be new infections, but rather the progression of an infected individual through the various compartments. Hence,

$$\mathscr{F} = \begin{pmatrix} \beta_1 SI/N + \beta_2 TI/N \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathscr{V} = \begin{pmatrix} (d + v + r_1)E - pr_2 I \\ -vE + (d + r_2)I \\ -b(N) + dS + \beta_1 SI/N \\ dT - r_1 E - qr_2 I + \beta_2 TI/N \end{pmatrix}. \tag{6}$$

The infected compartments are E and I, giving $m = 2$. An equilibrium solution with $E = I = 0$ has the form $x_0 = (0, 0, S_0, 0)^t$, where S_0 is any positive solution of $b(S_0) = dS_0$. This will be a DFE if and only if $b'(S_0) < d$. Without loss of generality, assume $S_0 = 1$ is a DFE. Then,

$$F = \begin{pmatrix} 0 & \beta_1 \\ 0 & 0 \end{pmatrix}, \quad V = \begin{pmatrix} d + v + r_1 & -pr_2 \\ -v & d + r_2 \end{pmatrix},$$

giving

$$V^{-1} = \frac{1}{(d + v + r_1)(d + r_2) - vpr_2} \begin{pmatrix} d + r_2 & pr_2 \\ v & d + v + r_1 \end{pmatrix}$$

and $\mathscr{R}_0 = \beta_1 v/((d + v + r_1)(d + r_2) - vpr_2)$. A heuristic derivation of the $(2, 1)$ entry of V^{-1} and $\mathscr{R}_0$ are as follows: a fraction $h_1 = v/(d + v + r_1)$ of exposed individuals progress to compartment I, a fraction $h_2 = pr_2/(d + r_2)$ of infectious individuals re-enter compartment E. Hence, a fraction h_1 of exposed individuals pass through compartment I at least once, a fraction $h_1^2 h_2$ pass through

at least twice, and a fraction $h_1^k h_2^{k-1}$ pass through at least k times, spending an average of $\tau = 1/(d+r_2)$ time units in compartment I on each pass. Thus, an individual introduced into compartment E spends, on average, $\tau(h_1 + h_1^2 h_2 + \cdots) = \tau h_1/(1-h_1h_2) = v/((d+v+r_1)(d+r_2) - vpr_2)$ time units in compartment I over its expected lifetime. Multiplying this by β_1 gives $\mathscr{R}_0$.

The model without treatment ($r_1 = r_2 = 0$) is an SEI model with $\mathscr{R}_0 = \beta_1 v/(d(d+v))$. The interpretation of $\mathscr{R}_0$ for this case is simpler. Only a fraction $v/(d+v)$ of exposed individuals progress from compartment E to compartment I, and individuals entering compartment I spend, on average, $1/d$ time units there.

Although conditions (A1)–(A5) do not restrict the decomposition of $f_i(x)$ to a single choice for $\mathscr{F}_i$, only one such choice is epidemiologically correct. Different choices for the function $\mathscr{F}$ lead to different values for the spectral radius of FV^{-1}, as shown in Table 1. In column (a), treatment failure is considered to be a new infection and in column (b), both treatment failure and progression to infectiousness are considered new infections. In each case the condition $\rho(FV^{-1}) < 1$ yields the same portion of parameter space. Thus, $\rho(FV^{-1})$ is a threshold parameter in both cases. The difference between the numbers lies in the epidemiological interpretation rather than the mathematical analysis. For example, in column (a), the infection rate is $\beta_1 + pr_2$ and an exposed individual is expected to spend $v/((d+v+r_1)(d+r_2))$ time units in compartment I. However, this reasoning is biologically flawed since treatment failure does not give rise to a *newly infected* individual.

Table 1
Decomposition of f leading to alternative thresholds

	(a)	(b)
$\mathscr{F}$	$\begin{pmatrix} \beta_1 SI/N + \beta_2 TI/N + pr_2 I \\ 0 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} \beta_1 SI/N + \beta_2 TI/N + pr_2 I \\ vE \\ 0 \\ 0 \end{pmatrix}$
$\mathscr{V}$	$\begin{pmatrix} (d+v+r_1)E \\ -vE + (d+r_2)I \\ -b(N) + dS + \beta_1 SI/N \\ dT - r_1 E - qr_2 I + \beta_2 TI/N \end{pmatrix}$	$\begin{pmatrix} (d+v+r_1)E \\ (d+r_2)I \\ -b(N) + dS + \beta_1 SI/N \\ dT - r_1 E - qr_2 I + \beta_2 TI/N \end{pmatrix}$
F	$\begin{pmatrix} 0 & \beta_1 + pr_2 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & \beta_1 + pr_2 \\ v & 0 \end{pmatrix}$
V	$\begin{pmatrix} d+v+r_1 & 0 \\ -v & d+r_2 \end{pmatrix}$	$\begin{pmatrix} d+v+r_1 & 0 \\ 0 & d+r_2 \end{pmatrix}$
$\rho(FV^{-1})$	$\dfrac{\beta_1 v + pr_2 v}{(d+v+r_1)(d+r_2)}$	$\sqrt{\dfrac{\beta_1 v + pr_2 v}{(d+v+r_1)(d+r_2)}}$

4.2. Multigroup model

In the epidemiological literature, the term 'multigroup' usually refers to the division of a heterogeneous population into several homogeneous groups based on individual behaviour (e.g., [14]). Each group is then subdivided into epidemiological compartments. The majority of multigroup models in the literature are used for sexually transmitted diseases, such as HIV/AIDS or gonorrhea, where behaviour is an important factor in the probability of contracting the disease [7,8,14,15]. As an example, we use an m-group SIRS-vaccination model of Hethcote [7,14] with a generalized incidence term. The sample model includes several SI multigroup models of HIV/AIDS as special cases [8,15]. The model equations are as follows:

$$\dot{I}_i = \sum_{j=1}^{m} \beta_{ij}(x) S_i I_j - (d_i + \gamma_i + \epsilon_i) I_i, \tag{7a}$$

$$\dot{S}_i = (1 - p_i) b_i - (d_i + \theta_i) S_i + \sigma_i R_i - \sum_{j=1}^{m} \beta_{ij}(x) S_i I_j, \tag{7b}$$

$$\dot{R}_i = p_i b_i + \gamma_i I_i + \theta_i S_i - (d_i + \sigma_i) R_i, \tag{7c}$$

for $i = 1, \ldots, m$, where $x = (I_1, \ldots, I_m, S_1, \ldots, S_m, R_1, \ldots, R_m)^t$. Susceptible and removed individuals die at the rate $d_i > 0$, whereas infected individuals die at the faster rate $d_i + \epsilon_i$. Infected individuals recover with temporary immunity from re-infection at the rate γ_i, and immunity lasts an expected $1/\sigma_i$ time units. All newborns are susceptible, and a constant fraction b_i are born into each group. A fraction p_i of newborns are vaccinated at birth. Thereafter, susceptible individuals are vaccinated at the rate θ_i. The incidence, $\beta_{ij}(x)$ depends on individual behaviour, which determines the amount of mixing between the different groups (see, e.g., Jacquez et al. [16]).

The DFE for this model is

$$x_0 = (0, \ldots, 0, S_1^0, \ldots, S_m^0, R_1^0, \ldots, R_m^0)^t,$$

where

$$S_i^0 = \frac{b_i(d_i(1 - p_i) + \sigma_i)}{d_i(d_i + \theta_i + \sigma_i)},$$

$$R_i^0 = \frac{b_i(\theta_i + d_i p_i)}{d_i(d_i + \theta_i + \sigma_i)}.$$

Linearizing (7a) about $x = x_0$ gives

$$F = \left[S_i^0 \beta_{ij}(x_0)\right]$$

and

$$V = [(d_i + \gamma_i + \epsilon_i)\delta_{ij}],$$

where δ_{ij} is one if $i = j$, but zero otherwise. Thus,

$$FV^{-1} = \left[S_i^0 \beta_{ij}(x_0)/(d_i + \gamma_i + \epsilon_i)\right].$$

For the special case with β_{ij} separable, that is, $\beta_{ij}(x) = \alpha_i(x)\lambda_j(x)$, F has rank one, and the basic reproduction number is

$$\mathscr{R}_0 = \sum_{i=1}^{m} \frac{S_i^0 \alpha_i(x_0)\lambda_i(x_0)}{d_i + \gamma_i + \epsilon_i}. \tag{8}$$

That is, the basic reproduction number of the disease is the sum of the 'reproduction numbers' for each group.

4.3. Staged progression model

The staged progression model [11, Section 3 and Appendix B] has a single uninfected compartment, and infected individuals progress through several stages of the disease with changing infectivity. The model is applicable to many diseases, particularly HIV/AIDS, where transmission probabilities vary as the viral load in an infected individual changes. The model equations are as follows (see Fig. 2):

$$\dot{I}_1 = \sum_{k=1}^{m-1} \beta_k S I_k/N - (v_1 + d_1)I_1, \tag{9a}$$

$$\dot{I}_i = v_{i-1}I_{i-1} - (v_i + d_i)I_i, \quad i = 2, \ldots, m-1, \tag{9b}$$

$$\dot{I}_m = v_{m-1}I_{m-1} - d_m I_m, \tag{9c}$$

$$\dot{S} = b - bS - \sum_{k=1}^{m-1} \beta_k S I_k/N. \tag{9d}$$

The model assumes standard incidence, death rates $d_i > 0$ in each infectious stage, and the final stage has a zero infectivity due to morbidity. Infected individuals spend, on average, $1/v_i$ time units in stage i. The unique DFE has $I_i = 0$, $i = 1, \ldots, m$ and $S = 1$. For simplicity, define $v_m = 0$. Then $F = [F_{ij}]$ and $V = [V_{ij}]$, where

$$F_{ij} = \begin{cases} \beta_j & i = 1,\ j \leqslant m-1, \\ 0 & \text{otherwise}, \end{cases} \tag{10}$$

$$V_{ij} = \begin{cases} v_i + d_i & j = i, \\ -v_j & i = 1 + j, \\ 0 & \text{otherwise}. \end{cases} \tag{11}$$

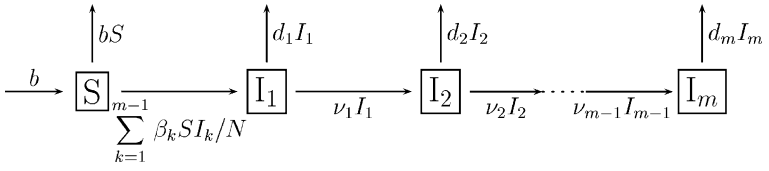

Fig. 2. Progression diagram for the staged progression model of (9a)–(9d).

Let a_{ij} be the (i,j) entry of V^{-1}. Then

$$a_{ij} = \begin{cases} 0 & i < j, \\ 1/(v_i + d_i) & i = j, \\ \dfrac{\prod_{k=j}^{i-1} v_k}{\prod_{k=j}^{i}(v_k + d_k)} & j < i. \end{cases} \tag{12}$$

Thus,

$$\mathscr{R}_0 = \frac{\beta_1}{v_1 + d_1} + \frac{\beta_2 v_1}{(v_1 + d_1)(v_2 + d_2)} + \frac{\beta_3 v_1 v_2}{(v_1 + d_1)(v_2 + d_2)(v_3 + d_3)} + \cdots + \frac{\beta_{m-1} v_1 \ldots v_{m-2}}{(v_1 + d_1)\ldots(v_{m-1} + d_{m-1})}. \tag{13}$$

The ith term in $\mathscr{R}_0$ represents the number of new infections produced by a typical individual during the time it spends in the ith infectious stage. More specifically, $v_{i-1}/(v_{i-1} + d_{i-1})$ is the fraction of individuals reaching stage $i-1$ that progress to stage i, and $1/(v_i + d_i)$ is the average time an individual entering stage i spends in stage i. Hence, the ith term in $\mathscr{R}_0$ is the product of the infectivity of individuals in stage i, the fraction of initially infected individuals surviving at least to stage i, and the average infectious period of an individual in stage i.

4.4. Multistrain model

The recent emergence of resistant viral and bacterial strains, and the effect of treatment on their proliferation is becoming increasingly important [12,13]. One framework for studying such systems is the multistrain model shown in Fig. 3, which is a caricature of the more detailed treatment model of Castillo-Chavez and Feng [12, Section 2] for tuberculosis and the coupled two-strain vector–host model of Feng and Velasco-Hernández [17] for Dengue fever. The model has only a single susceptible compartment, but has two infectious compartments corresponding to the two infectious agents. Each strain is modelled as a simple SIS system. However, strain one may 'super-infect' an individual infected with strain two, giving rise to a new infection in compartment

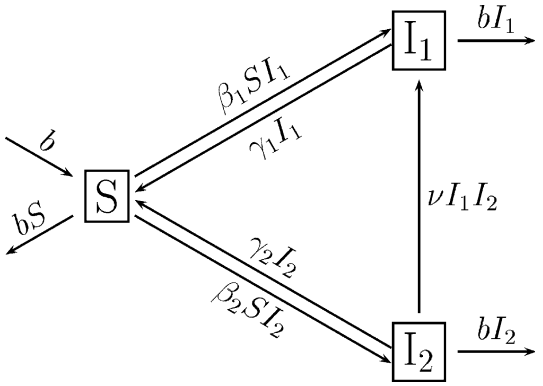

Fig. 3. Progression diagram for the multistrain model of (14a)–(14c).

I_1. The parameter $v > 0$ is the contact rate for the super-infection. The model equations are as follows:

$$\dot{I}_1 = \beta_1 I_1 S - (b + \gamma_1) I_1 + v I_1 I_2, \tag{14a}$$

$$\dot{I}_2 = \beta_2 I_2 S - (b + \gamma_2) I_2 - v I_1 I_2, \tag{14b}$$

$$\dot{S} = b - bS + \gamma_1 I_1 + \gamma_2 I_2 - (\beta_1 I_1 + \beta_2 I_2) S. \tag{14c}$$

For simplicity we have scaled the birth and death rates to $b > 0$. Hence, the DFE is $x_0 = (0, 0, 1)^t$, and

$$F = \begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{pmatrix}, \quad V = \begin{pmatrix} b + \gamma_1 & 0 \\ 0 & b + \gamma_2 \end{pmatrix}, \tag{15}$$

with V non-singular as required. The next generation matrix, FV^{-1}, has the two eigenvalues

$$\mathscr{R}_i = \frac{\beta_i}{b + \gamma_i}, \quad i = 1, 2. \tag{16}$$

In this example, $J_1 = F - V$ is reducible and (14a) and (14b) decouple near the DFE. The two eigenvalues correspond to the reproduction numbers for each strain. The basic reproduction number for the system is the maximum of the two. That is,

$$\mathscr{R}_0 = \max_{i \in \{1,2\}} \mathscr{R}_i. \tag{17}$$

An alternate interpretation of this model is that I_1 is the sole infected compartment and that I_2 is an uninfected compartment. The strain two equilibrium is $(0, 1 - (b + \gamma_2)/\beta_2, (b + \gamma_2)/\beta_2)$. Linearizing about this equilibrium gives $F = \beta_1 (b + \gamma_2)/\beta_2 + v(1 - (b + \gamma_2)/\beta_2)$, and $V = b + \gamma_1$. Thus,

$$\mathscr{R}_{12} = \frac{\mathscr{R}_1}{\mathscr{R}_2} + \frac{v}{b + \gamma_1} \left(1 - \frac{1}{\mathscr{R}_2} \right) \tag{18}$$

is the reproduction number for strain one near the strain two equilibrium. The interesting case is, of course, if $\mathscr{R}_2 > 1 > \mathscr{R}_1$, but $\mathscr{R}_{12} > 1$. That is, strain two can invade the DFE, but strain one cannot, and yet strain one can invade the strain two equilibrium. This can occur if v is sufficiently large.

4.5. Vector–host model

The general framework developed in Section 2 includes vector–host models. As an example, consider the following simplification of the two-strain, vector–host model proposed by Feng and Velasco-Hernández [17] for Dengue fever. The model couples a simple SIS model for the hosts with an SI model for the vectors. The four compartments correspond to infected hosts (I), infected vectors (V), susceptible hosts (S) and susceptible vectors (M). Hosts are infected by contacts with infected vectors, and vectors are in turn infected by contacts with infected hosts. These infection rates are given by the two terms $\beta_s SV$ and $\beta_m MI$. The model is written as follows (see Fig. 4):

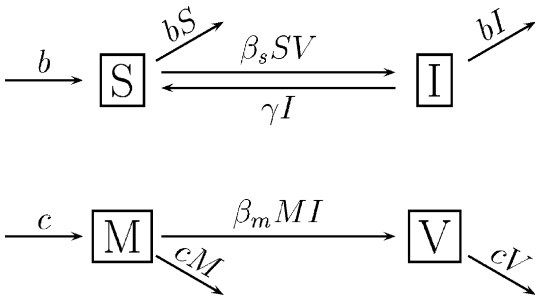

Fig. 4. Progression diagram for the vector–host model of (19a)–(19d).

$$\dot{I} = \beta_s SV - (b+\gamma)I, \tag{19a}$$

$$\dot{V} = \beta_m MI - cV, \tag{19b}$$

$$\dot{S} = b - bS + \gamma I - \beta_s SV, \tag{19c}$$

$$\dot{M} = c - cM - \beta_m MI. \tag{19d}$$

The birth and death rates have been scaled to $b > 0$ for the host and $c > 0$ for the vector. Thus, the DFE is $x_0 = (0, 0, 1, 1)^t$,

$$F = \begin{pmatrix} 0 & \beta_s \\ \beta_m & 0 \end{pmatrix}, \quad V = \begin{pmatrix} b+\gamma & 0 \\ 0 & c \end{pmatrix}, \tag{20}$$

with V non-singular, and the basic reproduction number is

$$\mathscr{R}_0 = \sqrt{\frac{\beta_s \beta_m}{c(b+\gamma)}}. \tag{21}$$

Near the DFE, each infected host produces β_m/c new infected vectors over its expected infectious period, and each infected vector produces $\beta_s/(b+\gamma)$ new infected hosts over its expected infectious period. The square root arises from the two 'generations' required for an infected vector or host to 'reproduce' itself.

5. The existence of sub-threshold equilibria

5.1. Analysis of the centre manifold near $x = x_0$, $\mathscr{R}_0 = 1$

In this section we consider the nature of the equilibrium solutions of the disease transmission model near the bifurcation point $x = x_0$, $\mathscr{R}_0 = 1$. Since $\mathscr{R}_0$ is often inconvenient to use directly as a

bifurcation parameter, we introduce a bifurcation parameter μ. Let μ be a bifurcation parameter such that $\mathscr{R}_0 < 1$ for $\mu < 0$ and $\mathscr{R}_0 > 1$ for $\mu > 0$ and such that x_0 is a DFE for all values of μ. Consider the system

$$\dot{x} = f(x, \mu), \tag{22}$$

where f is as described in Section 2, with the further restriction that f is continuously differentiable at least twice in both x and μ. The DFE is the line (x_0, μ) and the local stability of the DFE changes at the point $(x_0, 0)$. We use results of centre manifold theory (see e.g., [3]) to show that there are non-trivial (endemic) equilibria near the bifurcation point $(x_0, 0)$. Before stating these results we introduce some notation and collect a few facts.

We use the notation $D_x f(x_0, 0)$ for the partial derivative of f with respect to x evaluated at the point $x = x_0$, $\mu = 0$. Assume that the zero eigenvalue of $D_x f(x_0, 0)$ is simple and let v and w be the corresponding left and right nullvectors chosen such that $vw = 1$. By Lemma 1 and Theorem 2, all other eigenvalues of $D_x f(x_0, 0)$ have negative real parts. Let

$$a = \frac{v}{2} D_{xx} f(x_0, 0) w^2 = \frac{1}{2} \sum_{i,j,k=1}^{n} v_i w_j w_k \frac{\partial^2 f_i}{\partial x_j \partial x_k}(x_0, 0), \tag{23}$$

$$b = v D_{x\mu} f(x_0, 0) w = \sum_{i,j=1}^{n} v_i w_j \frac{\partial^2 f_i}{\partial x_j \partial \mu}(x_0, 0). \tag{24}$$

We show below that the sign of a determines the nature of the endemic equilibria near the bifurcation point. First, however, we note that the expression for a can be written in a different form using results of the previous sections.

Lemma 3. *If $f(x, \mu)$ is continuously dierentiable at least twice in both x and μ and conditions* (A1)–(A5) *are satisfied, and 0 is a simple eigenvalue of $D_x f(x_0, 0)$, then in the nullvectors of $D_x f(x_0, 0)$, $v_i \geqslant 0$ and $w_i \geqslant 0$ for $i = 1, \ldots, m$, $v_i = 0$ for $i = m + 1, \ldots, n$, and*

$$a = \sum_{i,j,k=1}^{m} v_i w_j w_k \left(\frac{1}{2} \frac{\partial^2 f_i}{\partial x_j \partial x_k}(x_0, 0) + \sum_{l=m+1}^{n} \alpha_{lk} \frac{\partial^2 f_i}{\partial x_j \partial x_l}(x_0, 0) \right), \tag{25}$$

with $[\alpha_{lk}]$, $l = m + 1, \ldots, n$, $k = 1, \ldots, m$, denoting the $(l - m, k)$ entry of $-J_4^{-1} J_3$ where J_3 and J_4 are the lower blocks of $D_x f(x_0, 0) = D(\mathscr{F}(x_0) - \mathscr{V}(x_0))|_{\mathscr{R}_0 = 1}$ defined in Lemma 1.

Proof. By Lemma 1 and Theorem 2, the first m components of v and w are the left and right null vectors of J_1. Since J_1 is essentially non-negative (i.e., $-J_1$ has the Z sign pattern), v and w can be chosen such that $v_i \geqslant 0$ and $w_i \geqslant 0$ for $i = 1, \ldots, m$ [4]. Further, since the eigenvalues of J_4 all have positive real parts, J_4^{-1} exists and the remaining components of v must be zero. Hence, from the definition of α_{lk},

$$w_l = \sum_{k=1}^{m} \alpha_{lk} w_k, \quad l = m + 1, \ldots, n. \tag{26}$$

With these facts, (23) leads to (25) as follows:

$$
\begin{aligned}
a &= \frac{1}{2}\sum_{i=1}^{m} v_i \sum_{j,k=1}^{n} w_j w_k \frac{\partial^2 f_i}{\partial x_j \partial x_k}(x_0, 0) \\
&= \frac{1}{2}\sum_{i=1}^{m} v_i \left(\sum_{j,k=1}^{m} w_j w_k \frac{\partial^2 f_i}{\partial x_j \partial x_k} + 2\sum_{j=1}^{m}\sum_{l=m+1}^{n} w_j w_l \frac{\partial^2 f_i}{\partial x_j \partial x_l} \right)\Bigg|_{(x_0,0)} \\
&= \frac{1}{2}\sum_{i=1}^{m} v_i \left(\sum_{j,k=1}^{m} w_j w_k \frac{\partial^2 f_i}{\partial x_j \partial x_k} + 2\sum_{j=1}^{m}\sum_{l=m+1}^{n} w_j \sum_{k=1}^{m} \alpha_{lk} w_k \frac{\partial^2 f_i}{\partial x_j \partial x_l} \right)\Bigg|_{(x_0,0)} \\
&= \sum_{i,j,k=1}^{m} v_i w_j w_k \left(\frac{1}{2}\frac{\partial^2 f_i}{\partial x_j \partial x_k} + \sum_{l=m+1}^{n} \alpha_{lk} \frac{\partial^2 f_i}{\partial x_j \partial x_l} \right)\Bigg|_{(x_0,0)}.
\end{aligned}
$$

For the second step, the second partial derivatives with respect to the uninfected compartments are zero by (A2)–(A4) (the details are similar to those in the proof of Lemma 1). □

Since the first m components of v and w are non-negative, the sign of a is determined by the signs of the partial derivatives and of α_{lk}. In many applications, the first set of partial derivatives are negative. Hence the sign of a is determined by the mixed partial derivatives involving both infected and uninfected compartments and α_{lk}.

Theorem 4. *Consider the disease transmission model defined by* (22) *with the function* $f(x, \mu)$ *satisfying the conditions* (A1)–(A5) *of Section* 2 *and the parameter* μ *as described above. Assume that the zero eigenvalue of* $D_x f(x_0, 0)$ *is simple. Let a and b be as defined by* (23) *and* (24) *and assume that* $b \neq 0$. *Then, there exists* $\delta > 0$ *such that*

(*i*) *if* $a < 0$, *then there are locally asymptotically stable endemic equilibria near* x_0 *for* $0 < \mu < \delta$ *and*
(*ii*) *if* $a > 0$, *then there are unstable endemic equilibria near* x_0 *for* $-\delta < \mu < 0$.

Proof. Centre manifold theory [3, Theorem 2.1.1] states that there exists a local centre manifold parameterized by u and μ of the form

$$W^c = \{(x, \mu) | x = x_0 + uw + z(u, \mu)\}, \tag{27}$$

where $z(u, \mu)$ is orthogonal to w and is second order in both u and μ. Further, the centre manifold, W^c, is invariant under (22). That is,

$$\dot{u}w + \frac{\mathrm{d}z}{\mathrm{d}t} = f(x_0 + uw + z(u, \mu), \mu). \tag{28}$$

Premultiplying (28) by v leads to the equation

$$\dot{u} = vf(x_0 + uw + z(u, \mu), \mu), \tag{29}$$

since $vz = 0$ for all (u, μ). Centre manifold theory further states [3, Theorem 2.1.2] that the behaviour of solutions of (22) near the bifurcation point $(x_0, 0)$ is governed by (29). The right-hand side of (29) can be expanded in a Taylor series as follows:

$$\dot{u} = vf(x_0,0) + vD_\mu f(x_0,0)\mu + vD_x f(x_0,0)(uw+z) + \frac{v}{2}D_{\mu\mu}f(x_0,0)\mu^2 + vD_{x\mu}f(x_0,0)\mu(uw+z) + \frac{v}{2}D_{xx}f(x_0,0)(uw+z)^2 + \mathcal{O}(3). \tag{30}$$

The notation $\mathcal{O}(3)$ is used to denote terms of third order and higher in u and μ. Since $f(x_0,\mu) = 0$ for all μ, the first, second and fourth terms in the expansion are zero, and since v is a left null vector of $D_x f(x_0,0)$, the third term vanishes. Hence, all remaining terms involving z are higher order, and

$$\dot{u} = au^2 + bu\mu + \mathcal{O}(3), \tag{31}$$

where a and b are defined by (23) and (24).

For $\delta > 0$ sufficiently small, there are non-zero, steady state solutions of (31) near the line $u = -b\mu/a$ for $|\mu| < \delta$. Since we have chosen μ so that the DFE is stable for $\mu < 0$, a local stability analysis of (31) shows that b must be positive. Further these non-zero solutions are stable if $a < 0$ and unstable if $a > 0$. Since the first m components of w are non-negative, it follows that the endemic solutions of (22) corresponding to these non-zero solutions of (31) are feasible (i.e., the components of x are non-negative) only if either $\mu > 0$ and $a < 0$ or if $\mu < 0$ and $a > 0$. □

In summary, the nature of the bifurcation at $\mathcal{R}_0 = 1$ is given by the sign of a. If either a or b are zero, then higher order terms in the Taylor series must be considered. If a is negative, then a branch of super-threshold endemic equilibria exists, and the bifurcation is supercritical. If $a > 0$, then there are unstable sub-threshold endemic equilibria, and the bifurcation is subcritical. These cases are often referred to as a forward bifurcation and a backward bifurcation respectively.

5.2. Examples

5.2.1. Treatment models

The result of Theorem 4 can be applied to the tuberculosis example of Section 4.1, since J_1 has a simple zero eigenvalue when $\mathcal{R}_0 = 1$. All second derivatives of f_i in (23) are zero at the DFE except the following:

$$\frac{\partial^2 f_1}{\partial E \partial I} = -\beta_1, \quad \frac{\partial^2 f_1}{\partial I^2} = -2\beta_1, \quad \frac{\partial^2 f_1}{\partial I \partial T} = \beta_2 - \beta_1.$$

Hence,

$$a = -\beta_1 v_1 w_2 (w_1 + w_2 + (1 - \beta_2/\beta_1)w_4).$$

Computation shows that the eigenvectors v and w can be chosen so that each component of w is positive and v_1 is also positive. Since biologically $\beta_2 < \beta_1$, it follows that $a < 0$. Hence, by Theorem 4 the DFE is locally asymptotically stable if $\mathcal{R}_0$ is slightly less than one (i.e., $\mu < 0$), and if $\mathcal{R}_0$ is slightly greater than one then the DFE is unstable and there is a locally asymptotically stable positive equilibrium near the DFE. The positivity of the endemic equilibrium follows from the positivity of 'infected' components (w_1 and w_2) of the right null vector. This vector gives the direction of the invasion when the DFE is unstable.

Castillo-Chavez et al. [18] propose the addition of a second infection term, $\beta_3 EI/N$, to f_2 and the negative of that term to f_1. Thus, progression from the exposed to the infected compartments is not linear, but is increased by exogenous re-infection. This change does not alter the DFE or $\mathscr{R}_0$. However, with this term

$$a = -\beta_1 v_1 w_2 (w_1 + w_2 + (1 - \beta_2/\beta_1) w_4) + \beta_3 w_1 w_2 (v_2 - v_1).$$

Calculation shows that $v_2 - v_1 > 0$. Hence, the direction of the bifurcation changes if β_3 is sufficiently large. If β_3 is such that $a > 0$, then there exists an unstable sub-threshold endemic equilibrium near the DFE. The significance of this unstable equilibrium is not trivial. It implies that, although the DFE is locally stable, perturbations above a small threshold can grow. Further, if $\mathscr{R}_0 > 1$, then the analysis of the centre manifold tells us not only that the DFE is unstable, but that there is no non-zero stable equilibrium near the DFE, and thus a small invasion will grow rapidly and to significant proportions even for $\mathscr{R}_0$ near one. The importance of this backward bifurcation for disease control is discussed in Section 6.

5.2.2. Multigroup model

Next, consider the multigroup model of Section 4.2. The $3m \times 3m$ Jacobian matrix $D_x f(x_0, 0)$ can be partitioned into blocks corresponding to I, S and R compartments as follows:

$$D_x f(x_0, 0) = \begin{pmatrix} \left[S_i^0 \beta_{ij}(x_0) - (d_i + \gamma_i + \epsilon_i)\delta_{ij}\right] & 0 & 0 \\ -\left[S_i^0 \beta_{ij}(x_0)\right] & -\left[(d_i + \theta_i)\delta_{ij}\right] & \left[\sigma_i \delta_{ij}\right] \\ \left[\gamma_i \delta_{ij}\right] & \left[\theta_i \delta_{ij}\right] & -\left[(d_i + \sigma_i)\delta_{ij}\right] \end{pmatrix}.$$

The upper left block is J_1 evaluated at $\mathscr{R}_0 = 1$, and the four lower right blocks comprise J_4. Note that $-J_4$ is a non-singular M-matrix, and therefore $s(J_4) < 0$. Let $w_i^S = w_{m+i}$ and $w_i^R = w_{2m+i}$, for $i = 1, \ldots, m$. Then,

$$w_i^S = -\frac{(d_i + \sigma_i)(d_i + \epsilon_i) + d_i \gamma_i}{d_i (d_i + \theta_i + \sigma_i)} w_i, \tag{32}$$

$$w_i^R = -\frac{\theta_i (d_i + \epsilon_i) - d_i \gamma_i}{d_i (d_i + \theta_i + \sigma_i)} w_i. \tag{33}$$

Applying (25) with the second partial derivatives for this model leads to

$$a = a_0 + \sum_{i,j,k=1}^{m} v_i w_j S_i^0 \left(w_k \frac{\partial \beta_{ij}}{\partial I_k}(x_0) + w_k^S \frac{\partial \beta_{ij}}{\partial S_k}(x_0) + w_k^R \frac{\partial \beta_{ij}}{\partial R_k}(x_0) \right), \tag{34}$$

where

$$a_0 = -\sum_{k=1}^{m} \frac{v_k w_k^2 (d_k + \gamma_k + \epsilon_k)(d_k \gamma_k + (d_k + \sigma_k)(d_k + \epsilon_k))}{b_k (d_k (1 - p_k) + \sigma_k)} < 0. \tag{35}$$

Consider now two cases of interest. First, if β_{ij} is constant, then $a = a_0 < 0$ and the bifurcation is in the forward direction. Second, if $\beta_{ij}(x) = \beta_{ij}(N_1, \ldots, N_m)$ where $N_i = I_i + S_i + R_i$, then $w_k + w_k^S + w_k^R = -\epsilon_k w_k / d_k$ and (34) leads to

$$a = a_0 - \sum_{i,j,k=1}^{m} \frac{v_i w_j w_k \epsilon_k S_i^0}{d_k} \frac{\partial \beta_{ij}}{\partial N_k}(x_0). \tag{36}$$

The results of several models [7,8] can be generalized using

$$\beta_{ij}(x) = \frac{\lambda_{ij}}{N_i} + \frac{\Lambda_{ij}}{\sum_{l=1}^{m} r_l N_l}. \tag{37}$$

For this model,

$$\begin{aligned} a = &- \sum_{k=1}^{m} \frac{d_k v_k w_k^2 (d_k + \gamma_k + \epsilon_k)(\epsilon_k p_k + \gamma_k + d_k + \sigma_k)}{b_k (d_k (1 - p_k) + \sigma_k)} \\ &+ \sum_{i,k=1}^{m} \frac{\left(\sum_{j=1}^{m} \Lambda_{kj} w_j \right) S_k^0 v_k r_i (\epsilon_i w_i b_k - \epsilon_k w_k b_i)}{b_k d_i \left(\sum_{j=1}^{m} r_j b_j / d_j \right)^2}. \end{aligned} \tag{38}$$

In the case studied by Hethcote and Van Ark [7], $\beta_{ij}(x) = \lambda_{ij}/N_i$. That is, $\Lambda_{ij} = 0$ in (37), and, by (38), $a < 0$ and the bifurcation is always in the forward direction. Huang et al. [8] used this model with $p_i = \theta_i = \gamma_i = \sigma_i = 0$, $[\lambda_{ij}]$ diagonal and $[\Lambda_{ij}]$ irreducible and found that backward bifurcation is possible. Our results remove these restrictions.

6. Discussion

The analysis presented herein can be applied to a large class of compartmental epidemic models that possess a DFE. The basic reproduction number, $\mathscr{R}_0$ (given by (4)), is a threshold parameter for these models. Moreover, the local analysis of the centre manifold yields a second parameter, a (given by (23) or (25)), whose sign indicates the existence and stability of a branch of endemic equilibria near the threshold $\mathscr{R}_0 = 1$. The stability of these equilibria is important for disease control, as there are large differences in the solutions of the system between the two cases $a < 0$ and $a > 0$. For the forward bifurcation ($a < 0$), there are stable super-threshold endemic equilibria near the DFE. Thus, reducing $\mathscr{R}_0$ through one lowers the incidence of the disease until it is eliminated as $\mathscr{R}_0$ passes below one. For a backward bifurcation ($a > 0$), there are unstable sub-threshold endemic equilibria near the DFE. The unstable sub-threshold endemic equilibria indicate that the DFE is stable only to very small perturbations, and that even a small perturbation can result in an epidemic. Further, as $\mathscr{R}_0$ increases through the threshold, there is a catastrophic increase in disease incidence. The lack of a local super-threshold endemic equilibrium suggests the existence of a non-local endemic equilibrium with a relatively large fraction of infected individuals, or a periodic solution. Backward bifurcations have been studied in models for HIV/AIDS [8,15], tuberculosis [18] and for BRSV [19].

Throughout the analysis, we have assumed that a well defined DFE exists. However, some models may be cast in terms of fractions so that there is an equilibrium distribution of individuals over the compartments even though the total population size is not constant. In this case the analysis can be applied to the fractions of individuals in each compartment to yield a threshold parameter (see, e.g., [20,21]). This threshold is not the basic reproduction number, since it is a

threshold for the fraction rather than the number of infected individuals, but the analysis for both the threshold condition and the direction of the bifurcation is similar.

Acknowledgements

The authors thank Dr. Herbert W. Hethcote for suggestions and for a preprint of [1] and Dr. Charles R. Johnson for helpful comments on the material in Appendix A.

Appendix A

Let $s(A)$ be the maximum real part of the eigenvalues of A (the spectral abscissa), and let $\rho(A)$ be the maximum modulus of the eigenvalues of A (the spectral radius). In Section 3, we make use of several results from the theory of M-matrices. A matrix $B = [b_{ij}]$ has the Z sign pattern if $b_{ij} \leqslant 0$ for all $i \neq j$. If $B = sI - P$, where I is the identity matrix, P is non-negative ($P \geqslant 0$ entrywise), and $s > \rho(P)$, then B is a non-singular M-matrix; if $s = \rho(P)$, then B is a singular M-matrix. There are many definitions of M-matrices equivalent to the above. For example, if a matrix B has the Z sign pattern and $s(B) > 0$, then B is a non-singular M-matrix [4, p. 135 (G_{20})].

Lemma 5. *Let H be a non-singular M-matrix and suppose B and BH^{-1} have the Z sign pattern. Then B is a non-singular M-matrix if and only if BH^{-1} is a non-singular M-matrix.*

The forward implication is stated in a slightly different form as Exercise 6b of Horn and Johnson [22, p. 127] and the reverse implication is stated in Berman and Plemmons [4, p. 159 (5.2)].

In general, this lemma does not hold if B a singular M-matrix. It can be shown to hold if B is singular and irreducible. However, this is not sufficient for our needs. More specifically, our proof of Theorem 4 makes use of the following lemma.

Lemma 6. *Let H be a non-singular M-matrix and suppose $K \geqslant 0$. Then,*

(i) *$(H - K)$ is a non-singular M-matrix if and only if $(H - K)H^{-1}$ is a non-singular M-matrix.*
(ii) *$(H - K)$ is a singular M-matrix if and only if $(H - K)H^{-1}$ is a singular M-matrix.*

Proof. Let $B = H - K$. Then both B and $BH^{-1} = I - KH^{-1}$ have the Z sign pattern. (Recall that $H^{-1} \geqslant 0$ since H is a non-singular M-matrix.) Hence, Lemma 5 implies statement (i). A separate continuity argument can be constructed for each implication in the singular case. □

References

[1] H.W. Hethcote, The mathematics of infectious diseases, SIAM Rev. 42 (2000) 599.
[2] O. Diekmann, J.A.P. Heesterbeek, J.A.J. Metz, On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations, J. Math. Biol. 28 (1990) 365.

[3] S. Wiggins, Introduction to Applied Nonlinear Dynamical Systems and Chaos, Springer, Berlin, 1990.
[4] A. Berman, R.J. Plemmons, Nonnegative Matrices in the Mathematical Sciences, Academic Press, New York, 1970.
[5] R.M. Anderson, R.M. May, Infectious Diseases of Humans, Oxford University, Oxford, 1991.
[6] O. Diekmann, J.A.P. Heesterbeek, Mathematical Epidemiology of Infectious Diseases: Model Building, Analyis and Interpretation, Wiley, New York, 1999.
[7] H.W. Hethcote, J.W.V. Ark, Epidemiological models for heterogeneous populations: proportionate mixing, parameter estimation, and immunization programs, Math. Biosci. 84 (1987) 85.
[8] W. Huang, K.L. Cooke, C. Castillo-Chavez, Stability and bifurcation for a multiple-group model for the dynamics of HIV/AIDS transmission, SIAM J. Appl. Math. 52 (3) (1992) 835.
[9] C. Castillo-Chavez, Z. Feng, W. Huang, On the computation of R_0 and its role in global stability, in: C. Castillo-Chavez, with S. Blower, P. van den Driessche, D. Kirschner, A.-A. Yakubu (Eds.), Mathematical Approaches for Emerging and Reemerging Infectious Diseases: An Introduction, Springer, 2002, p. 229.
[10] A. Nold, Heterogeneity in disease-transmission modeling, Math. Biosci. 52 (1980) 227.
[11] J.M. Hyman, J. Li, E.A. Stanley, The differential infectivity and staged progression models for the transmission of HIV, Math. Biosci. 155 (1999) 77.
[12] C. Castillo-Chavez, Z. Feng, To treat of not to treat: the case of tuberculosis, J. Math. Biol. 35 (1997) 629.
[13] S.M. Blower, P.M. Small, P.C. Hopewell, Control strategies for tuberculosis epidemics: new models for old problems, Science 273 (1996) 497.
[14] H.W. Hethcote, An immunization model for a heterogeneous population, Theor. Populat. Biol. 14 (1978) 338.
[15] J. Dushoff, W. Huang, C. Castillo-Chavez, Backwards bifurcations and catastrophe in simple models of fatal diseases, J. Math. Biol. 36 (1998) 227.
[16] J.A. Jacquez, C.P. Simon, J. Koopman, L. Sattenspiel, T. Perry, Modelling and analyzing HIV transmission: the effect of contact patterns, Math. Biosci. 92 (1988) 119.
[17] Z. Feng, J.X. Velasco-Hernández, Competitive exclusion in a vector–host model for the Dengue fever, J. Math. Biol. 35 (1997) 523.
[18] C. Castillo-Chavez, Z. Feng, A.F. Capurro, A model for TB with exogenous reinfection, Theor. Populat. Biol. 57 (2000) 235.
[19] D. Greenhalgh, O. Diekmann, M.C.M. de Jong, Subcritical endemic steady states in mathematical models for animal infections with incomplete immunity, Math. Biosci. 165 (2000) 1.
[20] S. Busenberg, P. van den Driessche, Disease transmission in multigroup populations of variable size, in: O. Arino, D. Axelrod, M. Kimmel, M. Langlais (Eds.), Mathematical Population Dynamics: Analysis of Heterogeneity, Theory of Epidemics, vol. 1, Wuerz, 1995, p. 15.
[21] X. Lin, H.W. Hethcote, P. van den Driessche, An epidemiological model for HIV/AIDS with proportional recruitment, Math. Biosci. 118 (1993) 181.
[22] R.A. Horn, C.R. Johnson, Topics in Matrix Analysis, Cambridge University, Cambridge, 1991.