ELSEVIER

Mathematical Biosciences 181 (2003) 85–106

Mathematical Biosciences

www.elsevier.com/locate/mbs

The critical vaccination fraction for heterogeneous epidemic models

Andrew N. Hill *, Ira M. Longini Jr.

Department of Biostatistics, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA

Received 16 October 2001; received in revised form 16 April 2002; accepted 27 June 2002

Abstract

Given a population with m heterogeneous subgroups, a method is developed for determining minimal vaccine allocations to prevent an epidemic by setting the reproduction number to 1. The framework is sufficiently general to apply to several epidemic situations, such as SIR, SEIR and SIS models with vital dynamics. The reproduction number is the largest eigenvalue of the linearized system round the local point of equilibrium of the model. Using the Perron–Frobenius theorem, an exact method for generating solutions is given and the threshold surface of critical vaccine allocations is shown to be a compact, connected subset of a regular $(m-1)$-dimensional manifold. Populations with two subgroups are examined in full. The threshold curves are either hyperbolas or straight lines. Explicit conditions are given as to when threshold elimination is achievable by vaccinating just one or two groups in a multi-group population and expressions for the critical coverage are derived. Specific reference is made to an influenza A model. Separable or proportionate mixing is also treated. Conditions are conjectured for convexity of the threshold surface and the problem of minimizing the amount of vaccine used while remaining on the threshold surface is discussed.

Keywords: Reproduction number; Vaccine efficacy; Next generation matrix; Perron–Frobenius; Regular manifold; Separable mixing

1. Introduction

Vaccination remains the only means for controlling many infectious diseases [1]. In heterogeneous populations, we would like to know what fraction of each subgroup should be vaccinated

* Corresponding author. Tel.: +1-404 727 5065; fax: +1-404 727 1370.
E-mail address: anhill@sph.emory.edu (A.N. Hill).

0025-5564/03/$ - see front matter
PII: S0025-5564(02)00129-3

with a vaccine with specific characteristics to eliminate the infectious disease from the whole population. The critical vaccination fractions are those fractions of each subpopulation that should be vaccinated to just achieve elimination. Knowledge of the critical vaccination fractions provides a starting point for disease elimination. These fractions can provide epidemiologists with information about the best deployment of limited quantities of vaccine to contain the infectious disease.

Mathematically, we find the critical vaccination fractions by solving for those fractions for which the reproduction number is 1. There has been a great deal of research carried out on the basic reproduction number and the reproduction number, going back to Bernoulli in 1766 [2]. We develop our concepts based heavily on the concept of the next generation matrix as introduced by Diekmann et al. [3]. We use the control of influenza as our motivating example. In this application, we ask the following question: *What minimal fraction of each age group should be vaccinated to eliminate the possibility of an influenza epidemic in the whole population*? [4].

Heterogeneous groups have previously been examined from various perspectives by several authors. Agur et al. [5] look at the two group case in detail and consider a weighted optimization problem. Hethcote and van Ark [6] treat the multi-group situation for SIRS systems under the proportionate mixing assumption. Cairns [7] examines optimal policies for heterogeneous models in discrete time, continuous time and a deterministic model with natural births and deaths. He also formulates optimal policies for these models. Becker and Starczak [8] look at household models and assume proportionate mixing when treating heterogeneity amongst individuals. Britton [9] has given bounds for the reproduction number for heterogeneous models. He assumes no knowledge of the next generation matrix and only observation of the final attack rates.

In this paper, we derive explicit conditions which govern when the reproduction number is 1 and investigate the properties of the threshold surface of critical vaccine allocations. In Section 2, the reproductive number is identified as the spectral radius of a matrix related to the next generation matrix. In Section 3, a way of generating exact solutions is described, and the threshold surface is seen to be compact and connected. From a different point of view, the threshold surface is shown to be a subset of the solution space to a certain determinantal equation from which it may be deduced that the threshold surface is a regular manifold. These results all rely crucially on the Perron–Frobenius theory for positive matrices. Explicit threshold curve equations are given in Section 4 when the population consists of just two groups. For more than two groups, the feasibility of vaccinating just one group is discussed in Section 5, and of vaccinating two groups in Section 6. A model for influenza A is used as an example. Separable mixing is analysed in Section 7. A conjecture is made on the convexity of the threshold surface in Section 8. Finally, an optimization application is presented in Section 9. Proofs of all the theorems are in the Appendix A.

2. The next generation matrix

We define the next generation matrix for a variety of infectious disease processes. The next generation matrix is applied in the case where there are no infected individuals in the population, a fixed point for most epidemic systems of equations. We partition the population into m mutually exclusive mixing groups. Let R_{ij} be the expected number of secondary infections in unvaccinated people in subgroup i resulting from a single randomly selected unvaccinated infectious person in mixing group j, where $i, j \in \{1, \ldots, m\}$. Then, the next generation matrix, R, is the $m \times m$ matrix

$$\begin{bmatrix} R_{11} & \cdots & R_{1m} \\ \vdots & \vdots & \vdots \\ R_{m1} & \cdots & R_{mm} \end{bmatrix}.$$

The basic reproduction number R_0 is defined to be the spectral radius of R, i.e., the largest magnitude eigenvalue [3]. If $R_0 > 1$, the epidemic process grows away from zero infectives.

Now assume that we randomly vaccinate the fraction f_i people in group i. The vaccine allocation in the whole population is given by $\mathbf{f} = [f_1, \ldots, f_m]^{\mathrm{T}}$. Without loss of generality, we define vaccine efficacy for susceptibility as $\mathrm{VE_S} = 1 - \theta$ and vaccine efficacy for infectiousness as $\mathrm{VE_I} = 1 - \phi$, where $\theta, \phi \in [0, 1]$ [1,10]. We assume at least one of the $\mathrm{VE_S}$, $\mathrm{VE_I} > 0$.

We assume that there are no infected individuals in the population at some time 0 other than the single infected person that is inserted. Then, we can model the beginning of the epidemic process as the following system of difference equations for y_{vi}, the expected number of secondary infections in population $i \in \{1, \ldots, m\}$, unvaccinated if $v = 0$ and vaccinated if $v = 1$, at generation g:

$$y_{0i}(g+1) = \sum_{j=1}^{m} R_{ij}(1 - f_j) y_{0j}(g) + R_{ij}\phi f_j y_{1j}(g),$$

$$y_{1i}(g+1) = \sum_{j=1}^{m} R_{ij}\theta(1 - f_j) y_{0j}(g) + R_{ij}\theta\phi f_j y_{1j}(g).$$

If we define the $2m \times 1$ column vector

$$\mathbf{y}(g) = [y_{01}(g), y_{11}(g), \ldots, y_{0m}(g), y_{1m}(g)]^{\mathrm{T}},$$

the linearized system of difference equations takes the form

$$\mathbf{y}(g+1) = M\mathbf{y}(g),$$

where M is the $2m \times 2m$ next generation matrix for vaccine allocation $\mathbf{f}$:

$$M = \begin{bmatrix} R_{11}(1-f_1) & R_{11}\phi f_1 & \cdots & R_{1m}(1-f_m) & R_{1m}\phi f_m \\ R_{11}\theta(1-f_1) & R_{11}\theta\phi f_1 & \cdots & R_{1m}\theta(1-f_m) & R_{1m}\theta\phi f_m \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ R_{m1}(1-f_1) & R_{m1}\phi f_1 & \cdots & R_{mm}(1-f_m) & R_{mm}\phi f_m \\ R_{m1}\theta(1-f_1) & R_{m1}\theta\phi f_1 & \cdots & R_{mm}\theta(1-f_m) & R_{mm}\theta\phi f_m \end{bmatrix}.$$

The reproductive number $R_{\mathbf{f}}$ is defined to be the spectral radius of M. If $R_{\mathbf{f}} > 1$, the epidemic grows. Note that $R_{\mathbf{f}}$ depends on the vaccine allocation, $\mathbf{f} = [f_1, \ldots, f_m]^{\mathrm{T}}$. We assume that R_0, the basic reproductive number, must be greater than 1 for an epidemic to occur. The notation is consistent in that $R_{\mathbf{f}} = R_0$ in the absence of vaccination when $\mathbf{f} = \mathbf{0}$, as will be shown later in this section. To determine the critical (i.e., minimal) vaccination fractions f_j of the populations needed to eliminate the disease, we set a threshold condition by requiring $R_{\mathbf{f}}$ to be 1.

Both M and R are non-negative matrices. We assume R, hence M, is indecomposable. That is, R admits no permutation of its indices that transforms it into a block triangular matrix. As noted in

[11, p. 53], 'this assumption eliminates the possibility that part of the community encounters a major outbreak whereas another subgroup in the population remains unaffected'. The Perron–Frobenius theorem [12,13] asserts that $R_{\mathbf{f}}$ is itself an eigenvalue of M, as opposed to merely being the modulus of an eigenvalue. This theorem is stated below along with associated relevant results. For matrices A and B of the same size, define $B \leqslant A$ if and only if each entry of $A - B$ is non-negative.

Theorem 2.1. *Let A be a non-negative indecomposable matrix and let $\rho(A)$ denote its spectral radius.*
(a) *$\rho(A) > 0$ and $\rho(A)$ is an eigenvalue of A.*
(b) *With $\rho(A)$ can be associated strictly positive left and right eigenvectors. This is the only eigenvalue for which this happens.*
(c) *If λ is any other eigenvalue of A, then $|\lambda| \leqslant \rho(A)$.*
(d) *$\rho(A)$ has algebraic multiplicity 1. That is, $\rho(A)$ is a simple eigenvalue.*
(e) *If $0 \leqslant B \leqslant A$ and β is an eigenvalue of B, then $|\beta| \leqslant \rho(A)$. If $|\beta| = \rho(A)$, then $B = A$, so that $\rho(A)$ increases when any element of A increases.*

Theorem 2.1(a)–(d) comprises the Perron–Frobenius theorem. The second statement of part (b) is not always included in textbook versions of the Perron–Frobenius theorem. However, it is a crucial result here and is proved in the Appendix A.

If A is a strictly positive matrix, Theorem 2.1(c) may be strengthened to read: *If λ is any other eigenvalue of A, then $|\lambda| < \rho(A)$.*

This result also holds if A is a non-negative primitive matrix, that is, one for which A^k is strictly positive for some natural number k. Primitive matrices are indecomposable, but the converse is not true [12]. In many cases, our R is primitive (often, though not always, it will be strictly positive, hence primitive). This means that eventually, after k generations say, the expected number of infections in every group from every other group will be positive.

The system of difference equations, hence the structure of M, takes into account $y_{1i}(g) = \theta y_{0i}(g)$. The linear operator represented by M is therefore defined on the m-dimensional subspace of $\mathbb{R}^{2m}$ spanned by $\mathbf{e}_i \otimes [1\ \theta]^{\mathrm{T}}$, $i = 1, \ldots, m$, where $\mathbf{e}_i$ denotes the standard $m \times 1$ basis vector of $\mathbb{R}^m$ with 1 in the ith entry and 0 elsewhere. The symbol $\otimes$ denotes the Kronecker product [12]. Let $\psi = 1 - \theta\phi$ so that $0 < \psi \leqslant 1$. Also define $d_i = 1 - \psi f_i$, $i = 1, \ldots, m$ so that $1 - \psi \leqslant d_i \leqslant 1$ with corresponding vector $\mathbf{d} = \mathbf{1} - \psi\mathbf{f}$, and diagonal matrices $F = \mathrm{diag}(f_1, \ldots, f_m)$, $D = I - \psi F$. Here $\mathbf{1}$ is the $m \times 1$ vector with all entries 1 and I is the $m \times m$ identity. The parameter ψ is a measure of the combined efficacy of $\mathrm{VE_S}$ and $\mathrm{VE_I}$ and has been denoted VE_R elsewhere for its effect on R_0 [14]. The significance of the quantity $\theta\phi$ has previously been noted by Halloran et al. [15], who refer to it as the *immunologically naive equivalent*. We have the following important characterisation of $R_{\mathbf{f}}$.

Theorem 2.2. *The reproductive number is given by $R_{\mathbf{f}} = \rho(RD)$. The threshold condition for eliminating the disease is therefore $\rho(RD) = 1$.*

If nobody is vaccinated, $\mathbf{f} = 0$ and $D = I$, in which case, $R_{\mathbf{f}} = R_0 > 1$ and there is no possibility of elimination. If $\rho(RD) = 1$ with associated right eigenvector $\mathbf{v}$, then:

$$RD\mathbf{v} = \mathbf{v} \Rightarrow R\mathbf{v} - \mathbf{v} = \psi RF\mathbf{v}.$$

Corollary 2.3. *When* $m = 1$, *so that* $R = R_0$, *set* $\mathbf{v} = 1$ *to obtain the threshold condition*:

$$R_0 - 1 = (1 - \theta\phi)R_0 f.$$

This is the result previously obtained by Longini et al. [10].

For a given efficacy parameter ψ, Theorem 2.2 ensures the threshold solutions are to be found among those of the equation $\det(RD - I) = 0$ which constrains the f_j to lie on a $(m-1)$-dimensional hypersurface of $\mathbb{R}^m$ and forces the existence of a (not necessarily largest) eigenvalue of 1. The hypersurface usually consists of more than one connected component. As will be shown in the next section, one of these components contains the *threshold* (*hyper*)*surface*, implicitly described by $R_{\mathbf{f}} = 1$. The rest of that component and all the other components satisfy $R_{\mathbf{f}} > 1$. For example, for fixed ψ when $m = 2$, f_1 and f_2 are constrained to lie on one branch of the hyperbola implicitly described by

$$\begin{aligned} &\det R(1 - \psi f_1)(1 - \psi f_2) + \psi(f_1 R_{11} + f_2 R_{22}) - \mathrm{tr}R + 1 = 0 \\ &\quad \iff \det R\psi^2 f_1 f_2 + (R_{11} - \det R)\psi f_1 + (R_{22} - \det R)\psi f_2 + \det(R - I) = 0. \end{aligned} \tag{2.1}$$

3. The threshold surface

In this section, we give bounds on ψ and $\mathbf{f}$, derive a method for generating threshold solutions, and show that given ψ, the threshold surface is a $(m-1)$-dimensional, compact, connected subset of a regular submanifold of $\mathbb{R}^m$. These properties all hinge on Perron–Frobenius theory. Loosely speaking, regular submanifolds do not 'fold back' on themselves in points of self-intersection. See Definition 1.11 of [16] for the precise meaning of this concept.

Define the polynomial $\Phi\colon \mathbb{R}^{m+1} \to \mathbb{R}$ by

$$\Phi(\mathbf{f}, \psi) = \det(RD - I) = \det(R - \psi RF - I),$$

where all quantities are as defined in Section 2. Efficacies, ψ, and vaccination levels, f_j, satisfying the threshold condition for growth of the epidemic are among the solutions implicitly described by the set $N \cap [0,1]^{m+1}$, where

$$N = \Phi^{-1}(0) = \{(\mathbf{f}, \psi) \in \mathbb{R}^{m+1} : \Phi(\mathbf{f}, \psi) = 0\}.$$

The restriction of $(\mathbf{f}, \psi)$ to $[0,1]^{m+1}$ introduces a relationship between R_0 and ψ. In the case of $m = 1$, simultaneously constraining $f, \psi \in [0,1]$ in Corollary 2.3 gives $\psi \in [1 - R_0^{-1}, 1]$, a result also found in [15]. As ψ is bounded above by 1, this also forces the critical vaccination fraction f to lie in the interval $[1 - R_0^{-1}, 1]$. Thus, for there to be any chance of eliminating the disease, both f and ψ can be no smaller than $1 - R_0^{-1}$ and $\theta\phi \leqslant R_0^{-1} < 1$.

The function Φ is invariant under the scaling transformation $(\mathbf{f}, \psi) \mapsto (\psi\mathbf{f}, 1)$, or, equivalently,

$$\Phi(\mathbf{f}, 1) = \Phi(\psi^{-1}\mathbf{f}, \psi).$$

It follows that the threshold surface for $\psi = 1$ (i.e., when either $\mathrm{VE_S} = 1$ or $\mathrm{VE_I} = 1$) may be scaled up to the surface for general $\psi < 1$. That is, threshold vaccination coverage for arbitrary efficacy $\psi < 1$ may be obtained by scaling up each $\psi = 1$ threshold f_j by a factor of ψ^{-1}, provided the resulting value is still less than 1.

As ψ decreases, scaling ensures that the corresponding threshold $\mathbf{f}$ that satisfies $\rho(RD) = 1$ is 'pushed outward' towards the vertex $(1, 1, \ldots, 1)$ of the unit hypercube $[0, 1]^m$. At this vertex, we have universal vaccine coverage in all groups, $\mathbf{f} = \mathbf{1} = (1, 1, \ldots, 1)^{\mathrm{T}}$, $D = (1 - \psi)I$ and $R_{\mathbf{f}} = (1 - \psi)\rho(R) = (1 - \psi)R_0$, where $R_0 > 1$ by assumption. The threshold condition $R_{\mathbf{f}} = 1$ yields $\psi = 1 - R_0^{-1}$, the minimum allowable value of ψ for epidemic elimination.

A similar argument and scale invariance show that for $f \neq 0$, $N = \Phi^{-1}(0)$ contains $(\mathbf{f}_0, \psi_0) = (f, \ldots, f, (1 - R_0^{-1})f^{-1})$ corresponding to $D = R_0^{-1}I$. Note that $R_{\mathbf{f}} = R_0^{-1}\rho(R) = 1$. The element $(\mathbf{f}_0, \psi_0)$ lies in $N \cap [0, 1]^{m+1}$ if and only if $f \in [1 - R_0^{-1}, 1]$ so that the corresponding $\psi_0 = (1 - R_0^{-1})f^{-1}$ lies in $[1 - R_0^{-1}, 1]$. In other words, for various ψ, the permissible vaccine coverage region for elimination contains the line segment joining the points $\mathbf{f} = (1 - R_0^{-1})\mathbf{1}$ (maximum $\psi = 1$) and $\mathbf{f} = \mathbf{1}$ ($\psi = 1 - R_0^{-1}$).

The bounds on ψ for elimination are therefore a straightforward generalization of the case when $m = 1$ discussed previously: $\psi \in [1 - R_0^{-1}, 1]$ or, equivalently, $\theta\phi \in [0, R_0^{-1}]$.

As an example, consider the influenza A example of Longini et al. [4] for which

$$R = \begin{bmatrix} 0.6 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.2 & 1.7 & 0.3 & 0.2 & 0.2 \\ 0.4 & 0.3 & 0.5 & 0.4 & 0.3 \\ 0.2 & 0.1 & 0.3 & 0.2 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \end{bmatrix}. \tag{3.1}$$

The $m = 5$ age groups are pre-school, school, young, middle-aged and old adults. For this matrix, $R_0 = 1.86$. For elimination by vaccination, we need $\psi \geqslant 1 - R_0^{-1} = 0.46$. A vaccine with $\mathrm{VE_S} = 0.9$ and $\mathrm{VE_I} = 0.8$ is efficacious enough since $\psi = 1 - (0.1)(0.2) = 0.98$. On the other hand, a vaccine with $\mathrm{VE_S} = 0.3$ and $\mathrm{VE_I} = 0.2$ has $\psi = 0.44$ and elimination is impossible.

Note that Theorem 2.1(e) ensures that, given ψ, for any threshold vaccine allocation $\mathbf{f}$, an increase in any one of the components f_j corresponds to a decrease in D so that $R_{\mathbf{f}} = \rho(RD) < 1$. This reflects the fact that if threshold elimination is achievable, more vaccine in any of the groups results in faster elimination. In terms of the threshold surfaces, this means that all possible elimination combinations $\mathbf{f}$, threshold and above, for all $\psi \in [1 - R_0^{-1}, 1]$, lie in the interior region of $[0, 1]^m$ bounded by the vertex $\mathbf{f} = \mathbf{1}$ and (see Proposition 3.2 below) the connected submanifold containing $\mathbf{f} = (1 - R_0^{-1})\mathbf{1}$, corresponding to threshold elimination for $\psi = 1$.

Bounds involving $\mathbf{f}$ follow from Theorem 2.1(e). Let $d_{\min}$ and $d_{\max}$ denote the smallest and largest components, respectively, of $\mathbf{d}$, and define $f_{\min}$ and $f_{\max}$ in the same fashion. There is a necessary, though not sufficient, condition for a solution to the determinantal equation $\det(RD - I) = 0$ to lie on the threshold surface.

Proposition 3.1. *Assume* $\mathbf{f} \in [0, 1]^m$ *satisfies* $\det(RD - I) = 0$. *If* $\mathbf{f}$ *lies on the threshold surface then* $1 - \psi \leqslant d_{\min} \leqslant R_0^{-1} \leqslant d_{\max} \leqslant 1$, *or, equivalently,* $0 \leqslant f_{\min} \leqslant \psi^{-1}(1 - R_0^{-1}) \leqslant f_{\max} \leqslant 1$. *Unless* $\mathbf{f} = \psi^{-1}(1 - R_0^{-1})$, *the inner inequalities are strict*: $f_{\min} < \psi^{-1}(1 - R_0^{-1}) < f_{\max}$.

As a result, if $f_{\min} > \psi^{-1}(1 - R_0^{-1})$ and $f_{\max} \leqslant 1$, the vaccine allocation is above critical level, i.e., $R_{\mathbf{f}} < 1$. Similarly, if $f_{\max} < \psi^{-1}(1 - R_0^{-1})$, then $R_{\mathbf{f}} > 1$, and elimination is not possible.

Now we specify an algorithm that generates exact threshold solutions. Theorem 2.1(b) ensures that any non-negative square matrix has exactly one eigenvector, up to scalar multiples, with

positive entries. Furthermore, this eigenvector corresponds to the spectral radius eigenvalue. This result applies to both left and right eigenvectors. Therefore, if $R_{\mathbf{f}} = \rho(RD) = 1$, there exists a positive vector $\mathbf{v} = [v_1, \ldots, v_m]^{\mathrm{T}}$ such that $\mathbf{v}^{\mathrm{T}}RD = \mathbf{v}^{\mathrm{T}}$. Equating components:

$$\sum_{i=1}^{m} v_i R_{ij} d_j = v_j \Rightarrow d_j = \frac{v_j}{\sum_{i=1}^{m} R_{ij} v_i} = \frac{v_j}{(\mathbf{v}^{\mathrm{T}}R)_j} \Rightarrow \psi f_j = 1 - \frac{v_j}{(\mathbf{v}^{\mathrm{T}}R)_j}, \tag{3.2}$$

where $(\mathbf{v}^{\mathrm{T}}R)_j$ denotes the jth component of $\mathbf{v}^{\mathrm{T}}R$. By computing the right hand side of (3.2), as $\mathbf{v}$ ranges over vectors with positive components and rejecting any $f_j \notin [0,1]$, we generate all threshold solutions. We can normalize the $\mathbf{v}$ so as to exclude scalar multiples which yield the same solution. One normalization is $\|\mathbf{v}\|_\infty = 1$, i.e., the largest entry of $\mathbf{v}$ is always 1. Another is $\|\mathbf{v}\|_1 = 1$, i.e., $\sum_{i=1}^{m} v_i = 1$, so that the $\mathbf{v}$ lie on the face of the unit simplex. In any case, whatever norm is chosen, the (interior) threshold solutions are to be found among

$$\left\{ f_j = \psi^{-1}\left[1 - \frac{v_j}{(\mathbf{v}^{\mathrm{T}}R)_j}\right] : \ \|\mathbf{v}\| = 1, \ v_i > 0 \ \forall i \right\}.$$

For example, take $\mathbf{u}$ to be the normalized left positive eigenvector corresponding to the spectral radius eigenvalue R_0 of R. As $\sum_{i=1}^{m} R_{ij}u_i = R_0 u_j$, the corresponding threshold solution is $\mathbf{f} = \psi^{-1}(1 - R_0^{-1})\mathbf{1}$.

That we are using left eigenvectors merely reflects the choice of RD for determining $R_{\mathbf{f}}$. However, $R_{\mathbf{f}} = \rho(DR)$ as well since AB and BA always have the same eigenvalues for any $m \times m$ matrices A and B. In this situation, right eigenvectors generate the solutions

$$\left\{ f_j = \psi^{-1}\left[1 - \frac{v_j}{(R\mathbf{v})_j}\right] : \ \|\mathbf{v}\| = 1, \ v_i > 0 \ \forall i \right\}. \tag{3.3}$$

Using this method, we can recover the threshold solution found by Britton [9]. Suppose the final attack rates p_i have been estimated in the case of an SIR system without vital dynamics. That is, p_i is the proportion of group i that ultimately gets infected. Then

$$R\mathbf{p} = -\ln(\mathbf{1} - \mathbf{p}), \tag{3.4}$$

where $\mathbf{p} = [p_1, \ldots, p_m]^{\mathrm{T}}$ and the logarithm is applied component-wise [4]. Taking $\mathbf{v} = \mathbf{p} > 0$ as a right eigenvector, this gives the threshold solution

$$f_j = \psi^{-1}\left(1 - \frac{p_j}{-\ln(1 - p_j)}\right), \quad j = 1, \ldots, m. \tag{3.5}$$

Britton [9] derives this result by producing a set of m positive quantities between whose minimum and maximum $R_{\mathbf{f}}$ is guaranteed to lie. Britton's R is the transpose of ours. By constraining all m values to be 1, he obtains $R_{\mathbf{f}} = 1$ yielding the threshold solution just given. This is a special case of a linear algebraic result found in both Rao [12, p. 472], and Varga [17, p. 52], which (transposed) states that for any non-negative $n \times n$ matrix A and positive vector $\mathbf{x} = [x_1, \ldots, x_n]^{\mathrm{T}}$:

$$\min_{1 \leqslant i \leqslant n} \frac{\sum_{i=1}^{n} a_{ij}x_i}{x_j} \leqslant \rho(A) \leqslant \max_{1 \leqslant i \leqslant n} \frac{\sum_{i=1}^{n} a_{ij}x_i}{x_j}.$$

Greenhalgh and Dietz [18] have previously made use of this result in obtaining bounds for reproductive numbers.

For an example of Britton's threshold solution, consider again the influenza A example of Longini et al. [4] with $\psi = 0.98$, and next generation matrix given by (3.1). Solution of (3.4) in MAPLE shows the final attack rate vector to be $\mathbf{p} = [0.32, 0.83, 0.58, 0.33, 0.20]^{\mathrm{T}}$. Substitution of this into (3.5) generates the threshold solution $\mathbf{f} = [0.17, 0.54, 0.34, 0.18, 0.11]^{\mathrm{T}}$.

As another example, the solutions generated from $\mathbf{1}^{\mathrm{T}}$ and $\mathbf{1}$ as left and right eigenvectors, respectively are $f_j = \psi^{-1}[1 - (\sum_{i=1}^{m} R_{ij})^{-1}]$ and $f_i = \psi^{-1}[1 - (\sum_{j=1}^{m} R_{ij})^{-1}]$. This requires all column and row sums of R to exceed 1 to be of any use (else the f can be negative).

As the spectral radius is a continuous function of $m \times m$ matrices [19], the boundary of the set of all unit norm vectors with positive components is also mapped to threshold solutions. That is, all threshold solutions are to be found among

$$\left\{ f_j = \psi^{-1}\left[1 - \frac{v_j}{(R\mathbf{v})_j}\right] : \quad \|\mathbf{v}\| = 1,\ v_i \geqslant 0\ \forall i \right\}. \tag{3.6}$$

Taking $\mathbf{v} = \mathbf{e}_j$ gives the solution $\mathbf{d} = R_{jj}^{-1}\mathbf{e}_j$. This d_j-axis intercept yields $\mathbf{f} \in [0,1]^m$ if and only if $\psi = 1$ and $R_{jj} \geqslant 1$. Although $\psi < 1$ in real applications, the above observation is useful for graphing the threshold surface. From these observations we also obtain:

Proposition 3.2. *Given $\psi \in [1 - R_0^{-1}, 1]$, the threshold surface $R_{\mathbf{f}} = 1$ is a compact, connected subset of $\mathbb{R}^m$.*

Returning to the determinantal equation $\det(RD - I) = 0$, of which the threshold points are solutions, we can establish regularity of the threshold surface. For a fixed $\psi \in [1 - R_0^{-1}, 1]$, define $\bar{\Phi} : \mathbb{R}^m \to \mathbb{R}$ by $\bar{\Phi}(\mathbf{f}) = \Phi(\mathbf{f}, \psi) = \det(RD - I)$. Regularity is proved by examining the gradient of $\bar{\Phi}$.

Theorem 3.3. *Let $\psi \in [1 - R_0^{-1}, 1]$. For each $j \in \{1, \ldots, m\}$, $\bar{\Phi}(\mathbf{f}) = d_j \det A_j - C_j$, where A_j is the $m \times m$ matrix obtained by replacing the jth column of $RD - I$ with that of R and C_j is the (j,j)-cofactor of $RD - I$. The gradient of $\bar{\Phi}$ as a function of $(f_1, \ldots, f_m)$ is given by*

$$\nabla\bar{\Phi} = -\psi[\det A_1, \ldots, \det A_m].$$

$\nabla\bar{\Phi}$ does not vanish on $\bar{N} = \bar{\Phi}^{-1}(0)$ so that the latter is a regular $(m-1)$-dimensional submanifold of $\mathbb{R}^m$. As a subset of the above submanifold, the threshold surface is also regular.

The m-dimensional submanifold $N = \Phi^{-1}(0)$ of $\mathbb{R}^{m+1}$ is also regular. Using the scale invariance mentioned earlier, it can be shown that $\partial\Phi/\partial\psi = -\sum_{j=1}^{n} f_j \det A_j$.

4. The case when $m = 2$

The situation when $m = 2$ is explicitly described by Eq. (2.1) and has also been investigated in [5] in the context of two culturally distinct population groups. We assume all $R_{ij} > 0$. Remaining indecomposable R may be similarly dealt with. The corresponding curves in $\mathbb{R}^2$ depend on the nature of the eigenvalues of R, namely R_0 and real λ, with $|\lambda| < R_0$. A case-by-case examination follows. Note that $\mathrm{tr}R = R_0 + \lambda$ and the determinant is $|R| = R_0\lambda$.

(a) If $|R| = 0$, R has eigenvalues 0 and $R_0 > 1$; $\mathrm{tr}R = R_0$. Simplification of (2.1) leads to

$$R_{11}\psi f_1 + R_{22}\psi f_2 = R_0 - 1 \iff f_1 = -\left(\frac{R_0}{R_{11}} - 1\right) f_2 + \frac{R_0 - 1}{\psi R_{11}}.$$

Since all entries of R are strictly positive, $R_{11} < R_0$. The submanifold is therefore a straight line of fixed negative slope with positive axis intercepts, decreasing with increasing ψ. Of interest is the portion of this line that lies inside the square $[0,1]^2$. The lower bound on ψ is found by setting $f_1 = f_2 = 1$, whence $\psi = 1 - R_0^{-1}$, as expected.

(b) If $|R| \neq 0$, the equation of the hyperbola is

$$\left(\psi f_1 - \frac{|R| - R_{22}}{|R|}\right)\left(\psi f_2 - \frac{|R| - R_{11}}{|R|}\right) = \frac{R_{12}R_{21}}{|R|^2}.$$

(i) If the other eigenvalue of R is $\lambda = 1$, this becomes

$$\left(\psi f_1 - \frac{R_0 - R_{22}}{R_0}\right)\left(\psi f_2 - \frac{R_0 - R_{11}}{R_0}\right) = \frac{(R_0 - R_{22})(R_0 - R_{11})}{R_0^2}.$$

We get a hyperbola with asymptotes $f_1 = \psi^{-1}(1 - R_{22}/R_0)$ and $f_2 = \psi^{-1}(1 - R_{11}/R_0)$, with lower branch passing through the origin. The threshold curve lies on the upper branch.

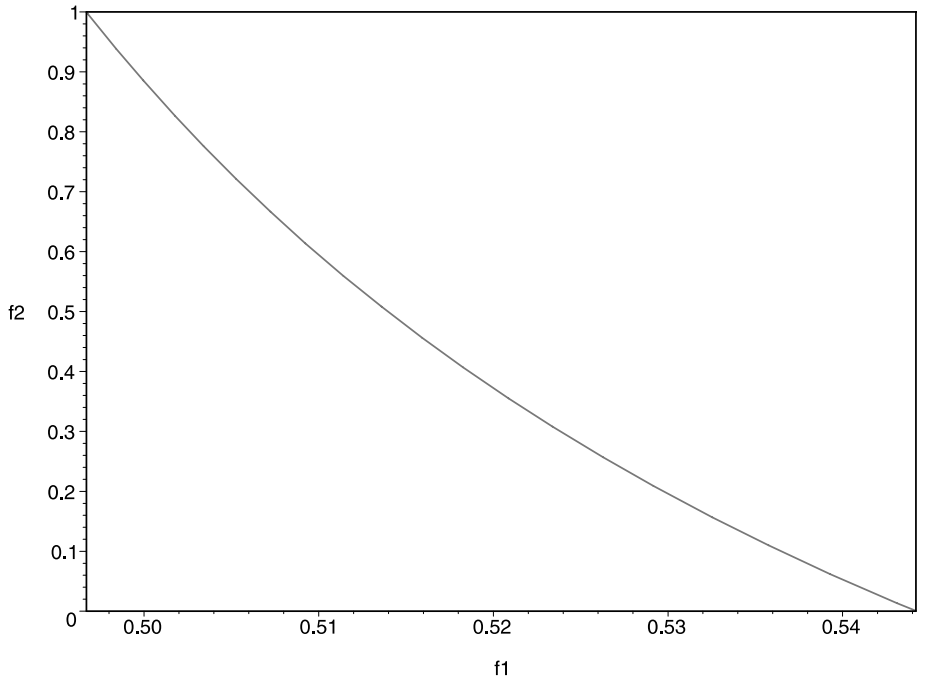

Fig. 1. An example of the threshold curve for a population consisting of two mixing groups with vaccine efficacy parameter $\psi = 0.9$. In this case, $R = \begin{bmatrix} 1.8 & 0.2 \\ 0.4 & 0.5 \end{bmatrix}$ and the threshold curve is on the upper branch of the hyperbola. The threshold solutions range from $(f_1, f_2) = (0.497, 1)$ to $(f_1, f_2) = (0.544, 0)$.

(ii) If the other (real) eigenvalue is $\lambda \neq 1$ with $|\lambda| < R_0$, we get another hyperbola. As $f = (1 - \lambda^{-1})\mathbf{1}$ is a solution to $\det(RD - I) = 0$, it follows that if $\lambda > 0$, the threshold curve lies on the upper branch, whereas if $\lambda < 0$, it lies on the lower branch. As examples, suppose

$$R = \begin{bmatrix} 1.8 & 0.2 \\ 0.4 & 0.5 \end{bmatrix},$$

for which $R_0 = 1.859$ and $\lambda = 0.441$. The equation of the hyperbola is

$$(\psi f_1 - 0.390)(\psi f_2 + 1.195) = 0.119,$$

for which ψf_1 varies between 0.444 and 0.490, corresponding to ψf_2 values of 1 and 0, respectively. The threshold curve lies on the upper branch as shown in Fig. 1 with $\psi = 0.9$. On the other hand, if

$$R = \begin{bmatrix} 1 & 4 \\ 3 & 2 \end{bmatrix},$$

then $R_0 = 5$ and $\lambda = -2$. Now the equation of the hyperbola is

$$(\psi f_1 - 1.2)(\psi f_2 - 1.1) = 0.12,$$

for which ψf_2 varies between 0.5 and 1, corresponding to ψf_1 values of 1 and 0, respectively. The threshold curve lies on the lower branch. See Fig. 2, where $\psi = 0.9$.

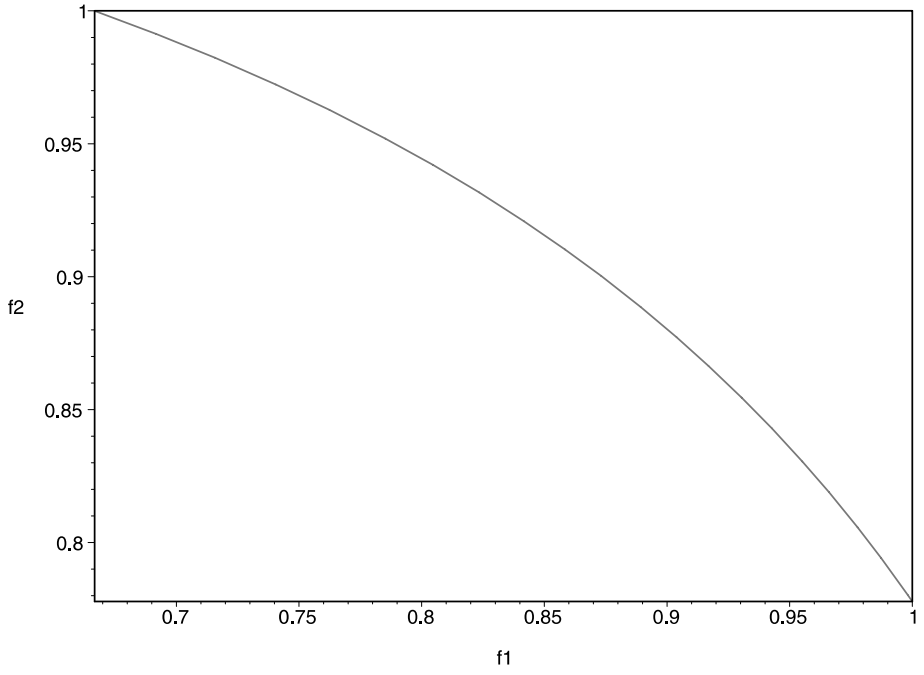

Fig. 2. Another example of the threshold curve for a population consisting of two mixing groups with vaccine efficacy parameter $\psi = 0.9$. This time, $R = \begin{bmatrix} 1 & 4 \\ 3 & 2 \end{bmatrix}$ and the threshold curve is on the lower branch of the hyperbola. The threshold solutions range from $(f_1, f_2) = (0.667, 1)$ to $(f_1, f_2) = (1, 0.778)$.

5. Conditions for elimination by vaccinating a single group

The determinantal description of the threshold surface is useful for finding threshold vaccination coverage restricted to a small number of groups. If we vaccinate the jth group only, $D = I - \psi F = \operatorname{diag}(1, \ldots, 1, 1 - \psi f_j, 1, \ldots, 1)$. Expand the determinant about the jth column as in the Proof of Theorem 3.3:

$$\det(RD - I) = (1 - \psi f_j) \det A_j - C_j = (1 - \psi f_j)(\det(R - I) + C_j) - C_j.$$

The last expression results from jth column expansion of $\det A_j$, with $d_k = 1$, $\forall k \neq j$. Setting to zero and solving

$$f_j = \psi^{-1} \frac{\det(R - I)}{\det(R - I) + C_j}. \tag{5.1}$$

As all the other $d_k = 1$, C_j is now the (j, j)-cofactor of $R - I$. According to Proposition 3.1, $R_{\mathbf{f}} = 1$ only if $\psi^{-1}(1 - R_0^{-1}) < f_j \leqslant 1$.

Consider the influenza A example of Longini et al. [4], with $R_0 = 1.86$ and R as in (3.1). If we vaccinate school-aged children with a vaccine having $\mathrm{VE_S} = 0.9$ and $\mathrm{VE_I} = 0.8$, then $\psi = 0.98$, $\det(R - I) = 0.0925$ and $C_2 = 0.015$. Eq. (5.1) gives $f_2 = 0.86/0.98 = 0.88$. Thus, 88% vaccination coverage of the school-aged group will achieve threshold. A quick check of the spectral radius shows that $R_{\mathbf{f}} = 1$ and our solution lies on the threshold surface and $1 - R_0^{-1} = 0.46 < \psi f_2 = 0.86$ as guaranteed by Proposition 3.1.

For $j \neq 2$, $C_j < 0$. As $\det(R - I) > 0$, Eq. (5.1) shows that threshold can only be achieved by single group vaccination for school-aged children ($\psi f_j > 1$ if $j \neq 2$).

6. Conditions for elimination by vaccinating two groups

It may be possible to achieve threshold elimination or better by vaccinating just two groups, particularly if the disease spreads more efficiently in these groups than in the rest of the population. If we wish to vaccinate groups i and j only, the threshold curve is a hyperbola, as with $m = 2$. The equation for the hyperbola may be obtained by taking successive determinant expansions about the ith and jth columns. Denote by E_{kk} the $m \times m$ matrix with 1 in the (k, k) position and zero elsewhere. Define $B = \det(R - I + E_{ii} + E_{jj})$ and let $B_{[i]}$, $B_{[j]}$ be the determinants of the $(m - 1) \times (m - 1)$ matrices obtained by deleting from $R - I + E_{ii} + E_{jj}$ the ith row and column and the jth row and column respectively. Furthermore, let $B_{[ij]}$ denote the determinant of the $(m - 2) \times (m - 2)$ matrix obtained by deleting the ith and jth columns and rows from $R - I$. If just $f_i, f_j > 0$, the threshold equation $\det(RD - I) = 0$ reduces to

$$(1 - \psi f_i)(1 - \psi f_j)B - (1 - \psi f_i)B_{[j]} - (1 - \psi f_j)B_{[i]} + B_{[ij]} = 0. \tag{6.1}$$

If $B = 0$, this is a straight line (suppressing the remaining $m - 2$ dimensions):

$$B_{[j]}\psi f_i + B_{[i]}\psi f_j = B_{[i]} + B_{[j]} - B_{[ij]}.$$

In accordance with Theorem 2.1(e), the slope of the line must be negative, that is, $B_{[i]}$ and $B_{[j]}$ must have the same sign.

If $B \neq 0$, we have a hyperbola:

$$\left(\psi f_i - \frac{B - B_{[i]}}{B}\right)\left(\psi f_j - \frac{B - B_{[j]}}{B}\right) = \frac{B_{[i]}B_{[j]} - BB_{[ij]}}{B^2}.$$

By Theorem 2.1(e), simultaneously increasing or decreasing f_i and f_j will increase or decrease $\rho(RD)$. Therefore, the right hand side of the hyperbola equation must be positive, so that an increase in f_i is accompanied by a decrease in f_j: $B_{[i]}B_{[j]} > BB_{[ij]}$. Note that this may not be the case, i.e., that the hyperbola is oriented to the left. This reflects the fact that there are no threshold solutions involving just the two groups i and j.

As an example, consider again the influenza A example of Longini et al. [4] with matrix R given in (3.1). Pre-school and school groups have the greatest numbers of secondary infections. A vaccination strategy might therefore be to vaccinate only these two groups with coverage f_1 for pre-school, f_2 for school and no vaccine for the rest. With $i = 1$ and $j = 2$, Eq. (6.1) becomes

$$-0.3755(1 - \psi f_1)(1 - \psi f_2) + 0.195(1 - \psi f_1) + 0.483(1 - \psi f_2) - 0.21 = 0. \tag{6.2}$$

The various coefficients are given by subdeterminants. For example,

$$B_{[2]} = \begin{vmatrix} 0.6 & 0.1 & 0.1 & 0.1 \\ 0.4 & 0.5-1 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.2-1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1-1 \end{vmatrix} = -0.195.$$

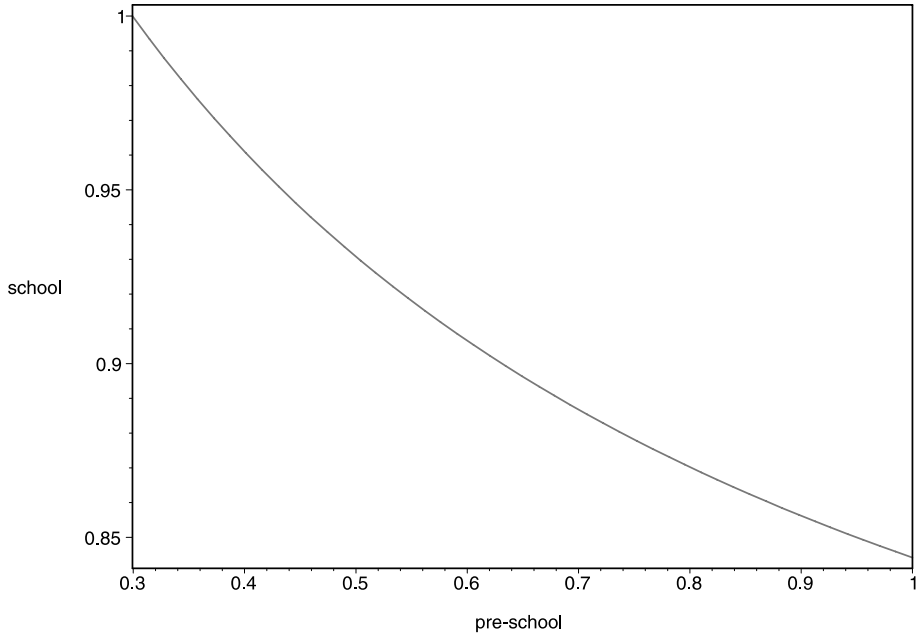

Fig. 3. Threshold vaccination levels for influenza A when only two of the five mixing groups are vaccinated, namely pre-school and school. The vaccine efficacy parameter is $\psi = 0.7$. In this case, f_1 represents pre-school vaccination coverage and f_2 school age coverage. The critical fractions range from $(f_1, f_2) = (0.299, 1)$ to $(f_1, f_2) = (1, 0.844)$ so that a high proportion of school age vaccination is required for elimination. As an example, 30% vaccination of pre-school and 100% vaccination of school age will achieve threshold.

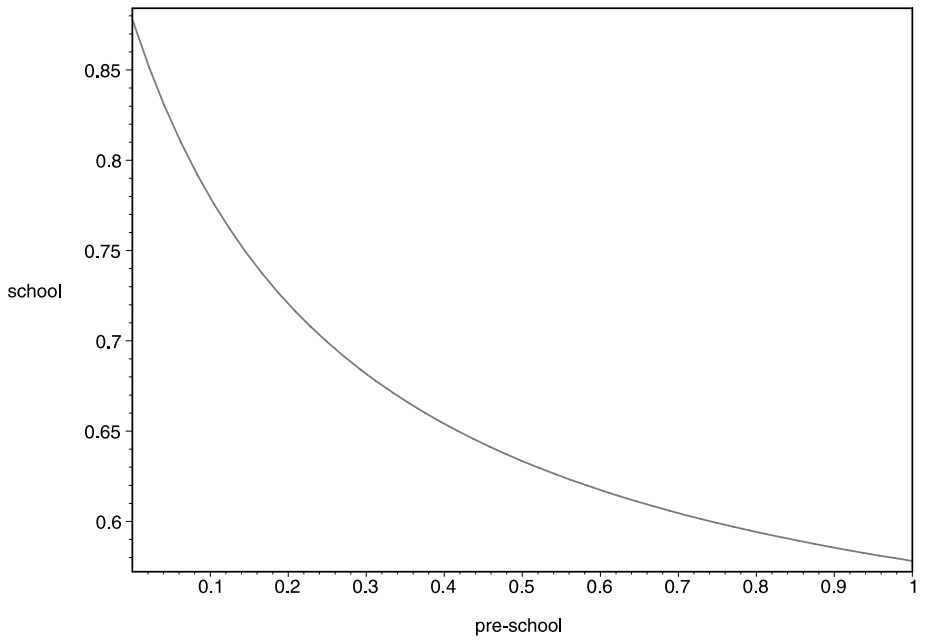

Fig. 4. Threshold vaccination levels for influenza A when only two of the five mixing groups are vaccinated, again pre-school and school. Now the vaccine efficacy parameter is $\psi = 0.98$. The critical fractions range from $(f_1, f_2) = (0, 0.878)$ to $(f_1, f_2) = (1, 0.578)$. For example, 25% vaccination of pre-school and 70% vaccination of school age will achieve threshold.

As $B_{[i]}B_{[j]} = -0.483 \times -0.195 = 0.094 > BB_{[ij]} = -0.3755 \times -0.21 = 0.079$, the hyperbola is correctly oriented.

For example, if $\mathrm{VE_S} = 0.7$ and $\mathrm{VE_I} = 0$, then $\psi = 1 - (0.3)(1) = 0.7$. 100% vaccination of the school-age population ($f_2 = 1$) leads to $f_1 = 0.30$ in Eq. (6.2) so that 30% vaccination of pre-schoolers is required. Calculation shows that $R_{\mathbf{f}} = 1$, and this allocation of vaccine does indeed lie on the threshold curve shown in Fig. 3.

Or suppose $\mathrm{VE_S} = 0.9$ and $\mathrm{VE_I} = 0.8$ so that $\psi = 0.98$. 70% vaccination of the school-age population gives $f_1 = 0.25$. Again $R_{\mathbf{f}}$ is found to be 1, so that 25% vaccination of pre-schoolers and 70% vaccination of school-age population will achieve threshold elimination when $\psi = 0.98$. The threshold curve is displayed in Fig. 4.

On the other hand, suppose we try vaccinating only middle-aged and old adults. A look at R suggests that this is probably not going to work as is now confirmed. In this case, $B = 0.0025$, $B_{[4]} = 7.8 \times 10^{-13}$, $B_{[5]} = 2.5 \times 10^{-50}$ and $B_{[45]} = 0.176$. Clearly, $B_{[4]}B_{[5]} < BB_{[45]}$ so that the resulting hyperbola is oriented the wrong way and there is no threshold solution, irrespective of the value of ψ.

7. Separable mixing

Assume that R has rank 1, so that $R_{ij} = a_i b_j$, $i, j = 1 \ldots, m$. This is known in the literature as *separable mixing* [3] and corresponds to group i being equally susceptible to all other groups and group j being equally infectious to all other groups. The special case when $a_i = kb_i$ for all i is

known as *proportionate mixing* [20]. This type of mixing has been extensively researched, for example, see Hethcote and van Ark [6] and Cairns [7]. As RD also has rank 1 and $R_{\mathbf{f}}$ is a simple eigenvalue, the sum of the eigenvalues of RD, counted by algebraic multiplicity, is $R_{\mathbf{f}}$. That is, $R_{\mathbf{f}} = \operatorname{tr} RD = \sum_{j=1}^{m} R_{jj}(1 - \psi f_j)$. The threshold condition of $R_{\mathbf{f}} = 1$ becomes

$$\psi \sum_{j=1}^{m} R_{jj} f_j = R_0 - 1. \tag{7.1}$$

Note that $\operatorname{tr} R = R_0$ for the same reasons as with RD. The threshold surface is a hyperplane with axis intercepts $f_j = (R_0 - 1)/\psi R_{jj} > 0$. The intercept $f_j \leqslant 1$ if and only if $\psi \geqslant (R_0 - 1)/R_{jj}$. In other words, for $1 - R_0^{-1} \leqslant \psi < (R_0 - 1)/R_{jj}$, threshold elimination is impossible by vaccinating only the jth group. From (7.1),

$$f_1 = R_{11}^{-1} \left\{ \psi^{-1}(R_0 - 1) - \sum_{j=2}^{m} R_{jj} f_j \right\}.$$

The f_j, $j \geqslant 2$ are chosen so that f_1 lies in $[0, 1]$.

8. A conjecture on convexity

Preliminary inspection of the surfaces generated when $m = 3$ suggests that, for some types of next-generation matrix R, the threshold surface corresponds to the graph of a convex or concave function. For general m and fixed ψ, the function is one of $m - 1$ variables obtained by solving the implicit expression $\bar{\Phi}(f_1, \ldots, f_m) = 0$ for one of the f_i. Theorem 2.1(e) ensures that this is possible for any f_i. A convex function $g : \mathbb{R}^k \to \mathbb{R}$ satisfies

$$g((1 - t)\mathbf{a} + t\mathbf{b}) \leqslant (1 - t)g(\mathbf{a}) + tg(\mathbf{b}),$$

for all $t \in [0, 1]$ and $\mathbf{a}, \mathbf{b} \in \mathbb{R}^k$. By a concave function $h : \mathbb{R}^k \to \mathbb{R}$ we mean that $-h$ is convex.

In view of the scale invariance observed in Section 3, if the threshold surface for given R is convex for a particular ψ_0, then it will be convex for all other values of ψ. Therefore, it is enough to show convexity when $\psi = 1$. Establishing convexity is equivalent to showing that $\rho((1 - t)\mathbf{f}_1 + t\mathbf{f}_2) \leqslant 1$, for any two critical vaccine solutions $\mathbf{f}_1$ and $\mathbf{f}_2$. This is a direct consequence of Theorem 2.1(e). Similarly, concavity holds if $\rho((1 - t)\mathbf{f}_1 + t\mathbf{f}_2) \geqslant 1$, for all critical $\mathbf{f}_1, \mathbf{f}_2$.

When $m = 2$, as seen in Section 4, the threshold curve is either a straight line or the upper branch of a hyperbola (both convex), or the lower branch of a hyperbola (concave). More generally, if the threshold surface is convex, we could generate points on the surface by the technique of Section 3 and approximate by the polyhedral surface with vertices consisting of these points. All points on the polyhedron would then satisfy $R_{\mathbf{f}} \leqslant 1$, giving a conservative approximation to the critical surface. We conjecture the following, also based on observation when $m = 3$.

Conjecture 8.1. *Assume R has only real eigenvalues. If all the eigenvalues are positive, then the threshold surface is convex. If all the eigenvalues except the spectral radius eigenvalue are negative, then the threshold surface is concave.*

As examples, see the accompanying MAPLE plots of the threshold surfaces for the following matrices. In all cases, ψ was taken to be 1. Although this is unrealistic, there is no loss in mathematical generality here owing to the scale invariance property discussed in Section 3.

(a) A 3×3 next generation matrix with three distinct positive eigenvalues

$$R_1 = \begin{bmatrix} 2.85 & 1.46 & 1.37 \\ 0.92 & 3.57 & 0.06 \\ 1.82 & 2.29 & 2.46 \end{bmatrix}.$$

The eigenvalues are $R_0 = 5.26$ and 2.39, 1.23. For any non-zero eigenvalue λ, $\mathbf{f} = (1 - \lambda^{-1})\mathbf{1}$ satisfies $\det(RD - I) = 0$. The threshold surface is therefore the upper branch as shown in Fig. 5, as $1 - \lambda^{-1} > 1$ is maximized when $\lambda = R_0$. Note that the surface is convex.

(b) A 3×3 next generation matrix with three distinct real eigenvalues, two of which are negative

$$R_2 = \begin{bmatrix} 5.16 & 2.74 & 4.27 \\ 6.54 & 2.32 & 5.82 \\ 5.28 & 2.73 & 2.47 \end{bmatrix}.$$

The eigenvalues are $R_0 = 12.16$ and -0.58, -1.63. The threshold surface is therefore the lower branch as shown in Fig. 6, for unless $\lambda = R_0$, $1 - \lambda^{-1} > 1$. Note that the surface is concave.

(c) A 3×3 next generation matrix with one real eigenvalue

$$R_3 = \begin{bmatrix} 1.78 & 3.69 & 1.62 \\ 2.46 & 2.95 & 3.74 \\ 3.17 & 0.71 & 3.67 \end{bmatrix}.$$

The eigenvalues are $R_0 = 7.87$ and 0.26 ± 1.24i. The threshold surface is shown in Fig. 7. It is neither convex nor concave; in fact, it contains a saddle-like point as is obvious from Fig. 8.

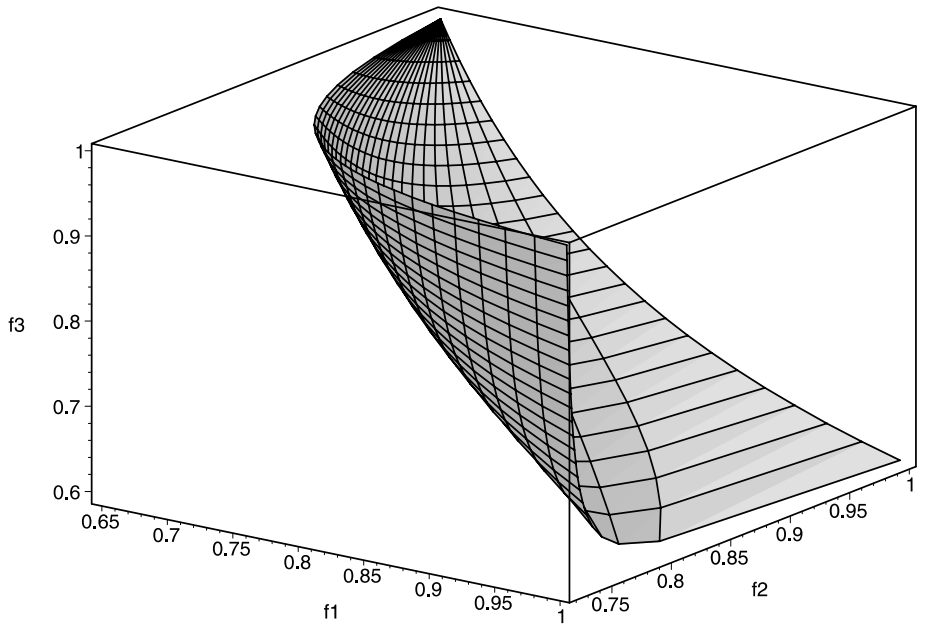

Fig. 5. The threshold surface for a population consisting of three mixing groups with vaccine efficacy parameter $\psi = 1$. The next generation matrix R_1 has three distinct positive eigenvalues and the surface corresponds to the graph of a convex function. The vertices of the surface are $(f_1, f_2, f_3) = (0.649, 1, 1)$, (1,0.720,1) and (1,1,0.593).

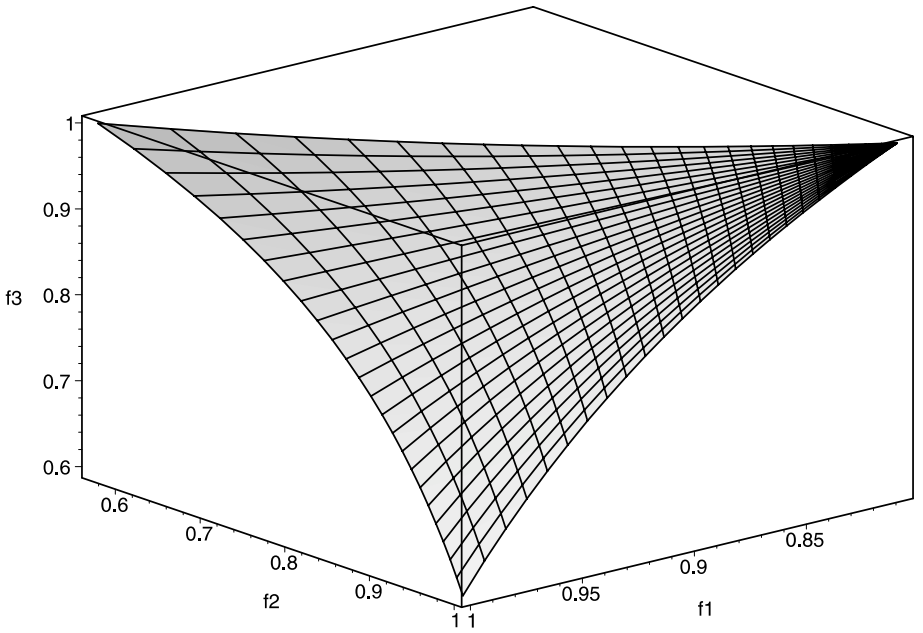

Fig. 6. The threshold surface for a population consisting of three mixing groups with vaccine efficacy parameter $\psi = 1$. The next generation matrix R_2 has three distinct real eigenvalues, two of which are negative, and the surface corresponds to the graph of a concave function. The vertices of the surface are $(f_1, f_2, f_3) = (0.806, 1, 1)$, (1,0.569,1) and (1,1,0.595).

9. An optimization application

Often one is interested in minimizing the total number of vaccinations administered while remaining at threshold level. Optimal strategies have previously been considered by Cairns [7], Agur et al. [5] in the context of two groups, Becker and Starczak [8] who consider household models and, for age structured populations, Castillo-Chavez and Feng [22], Hadeler and Müller [23] and Müller [24]. Mathematically, the goal is to minimize $\sum_{i=1}^{m} n_i f_i$ subject to the constraint $R_{\mathbf{f}} = 1$, where n_i is the population of subgroup i. This can be done using the method of Lagrange multipliers; the crucial gradient of the threshold surface is supplied by Theorem 3.3. The Lagrange multiplier system of equations is

$$n_i + k \det A_i = 0, \quad i = 1, \ldots, m$$
$$R_{\mathbf{f}} = 1.$$

For example, in the case of the Influenza A model, Longini et al. [4] give the subgroup sizes as $n_1 = 77$, $n_2 = 241$, $n_3 = 375$, $n_4 = 204$ and $n_5 = 103$ (the total population was 1000). For the matrix given in Eq. (3.1), with $\psi = 0.98$, solution of the Lagrange multiplier system, including the reduced systems on all possible lower dimensional boundaries of $[0,1]^5$, yields the optimal allocation $\mathbf{f} = (0.30, 0.68, 0, 0, 0)$. That is to say, the critical vaccine allocation which innoculates the least number of people is achieved by vaccinating 30% of pre-schoolers and 68% of school children. This corresponds to vaccinating a total of 187 people (out of the overall population of 1000).

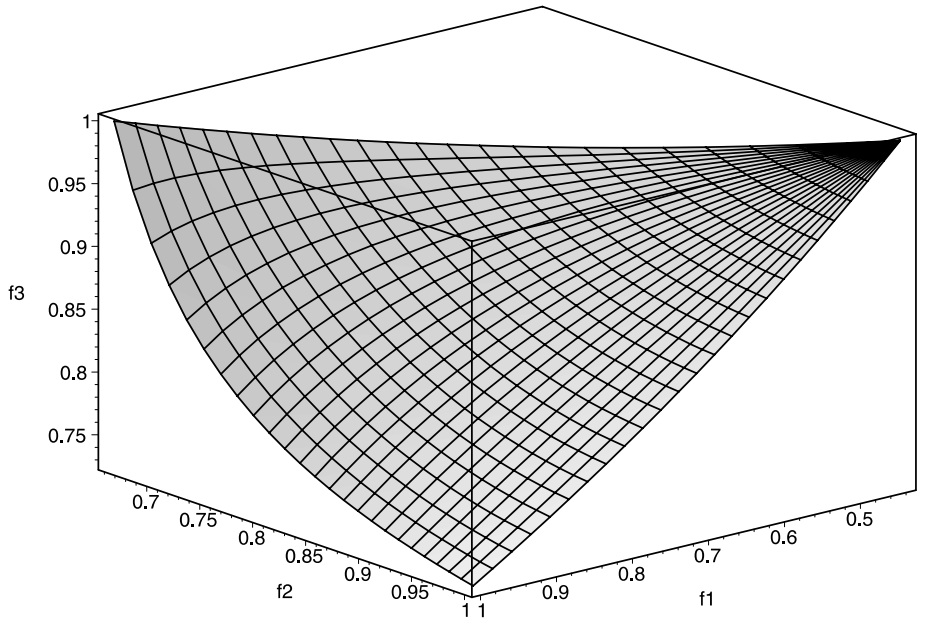

Fig. 7. The threshold surface for another population consisting of three mixing groups with vaccine efficacy parameter $\psi = 1$. The next generation matrix R_3 has just one real eigenvalue. The vertices of the surface are $(f_1, f_2, f_3) = (0.438, 1, 1)$, (1,0.661,1) and (1,1,0.728).

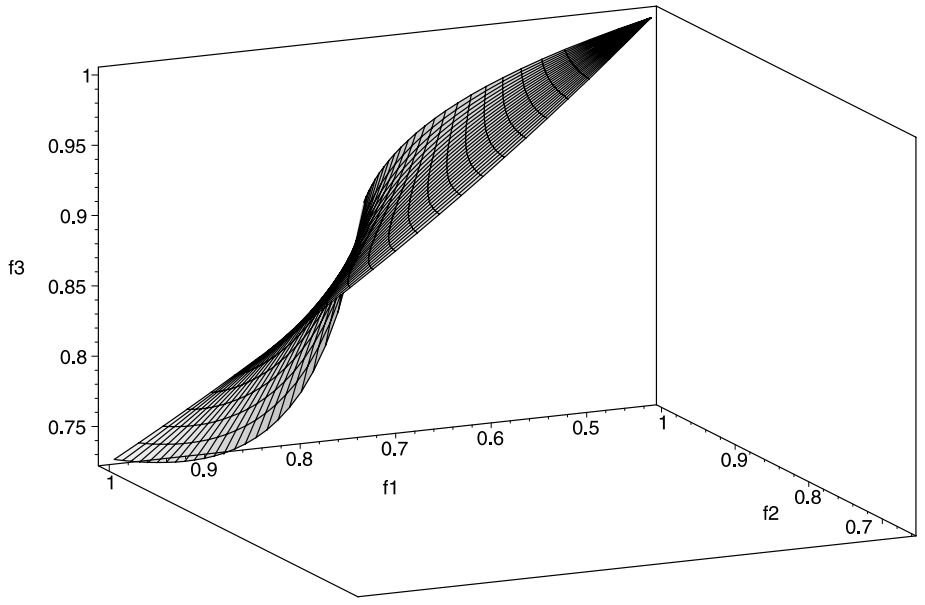

Fig. 8. The threshold surface in Fig. 7 from a different perspective. This view clearly shows the saddle nature of the surface.

10. Discussion

In this work, we have derived explicit conditions that govern when the reproduction number is 1. The threshold surface, the set of all vaccination allocations which lead to a reproduction number of 1, is a subset of solutions to a determinantal equation. The threshold surface is shown to be well-behaved, in that it is a subset of a compact, connected, regular manifold. Although the surface is described implicitly via a determinantal equation, another method was outlined for generating exact threshold solutions. In principle, this could be implemented in a numerical approach to generate a grid of points that approximate the surface. We have made a detailed examination of the case when the population is split into two subgroups only. In this case, it is relatively easy to write down an explicit equation for the threshold curve. For higher dimensions, we have given methods for achieving critical coverage, if possible, by singling out just one subgroup for vaccination or by allocating vaccine to two groups only. We also looked at the special case of separable mixing and conjectured convexity of the threshold surface in certain situations. Finally, we presented an optimization application.

We have given general formulations for the next generation matrix and vaccine effects. These results can be applied to a variety of epidemic and endemic situations, e.g., SIR, SEIR, SIS, including those with simple vital dynamics [21]. We give the explicit example of the SIR model without vital dynamics in Eqs. (3.4) and (3.5). However, our next generation matrix, M, may not apply directly to all models, especially those with complex age structure [22–25] or a long complex infectious period. In these cases, the next generation matrix needs to be derived from the linearization, in the neighborhood of the fixed point with no infected individuals, of the defining system of differential equations in the SI-plane. Conceptually, the elements R_{ij} of the next generation matrix are the products of three factors: (1) the contact rate between people in mixing group i and j; (2) the per-contact transmission probability between an infected, unvaccinated person in mixing group j and a susceptible, unvaccinated person in group i; and (3) the average duration of infectiousness for an infected in group j. This will be true regardless of the infectious disease being modeled. Vaccination reduces the magnitude of the elements of the next generation matrix as given by the matrix M for a variety of heterogeneous responses to vaccination. This includes all-or-none, leaky and mixtures of effects [1,26].

Once the elements of the next generation matrix have been determined, the theory outlined in this paper can be used to define the minimal vaccination coverage needed to ensure elimination of the infection. In some cases, the next generation matrix can be statistically estimated from field data [27]. In other cases, it can be roughly determined from field data via sensitivity analysis [4]. Vaccine efficacy can be estimated from vaccine trials [1,28]. In this paper, we have used influenza vaccination as a motivating example. Longini et al. [4] roughly determined the next generation matrix for pandemic Asian influenza (1957) for the five age groups shown in Section 3. The next pandemic strain of influenza could have similar behavior to this one. In addition, Longini et al. [28] estimated the efficacy of cold-adapted influenza vaccine in children to be $VE_S \approx 0.90$ and $VE_I \approx 0.80$. As we have shown in Section 6, Asian influenza could be eliminated by vaccinating 70% of school children and 25% of pre-school children, without vaccinating adults. Other vaccination fractions for children that fall on the curve in Fig. 4 could also result in the elimination of influenza epidemics. Mass vaccination of children with a highly efficacious influenza vaccine has

been argued to be effective in controlling influenza in the whole population based on both theoretical grounds [28] and observations [29,30].

Although we have assumed constant efficacy parameter ψ amongst all the population subgroups, it is a straightforward modification to allow different ψ_i, $i = 1, \ldots, m$ for each of the subgroups. As is clear from (3.6), this merely results in scale changes in each of the f_j and the threshold surface for constant ψ is expanded or contracted in each of the coordinate axis directions accordingly.

In this paper, we have derived exact methods for determining the critical vaccination fractions for a general next generation matrix. Exact solutions are derived for the cases when there are one or two mixing groups and when vaccination is carried out in one or two mixing groups in the presence of other mixing groups. For the special case of separable mixing, exact solutions are always available when they exist. In general, there is no easily obtainable exact solution for the case of vaccinating more than two mixing groups. The vaccine allocation for any particular group can be expressed as a function of the remaining allocations, namely, $f_j = \psi^{-1}(1 - C_j/\det A_j)$, whenever this is defined (see Theorem 3.3). However, it is still a hit and miss affair determining which of the remaining allocations yield $f_j \in [0, 1]$. A numerical algorithm could be devised to find exact solutions using the technique in Section 3. As this does not involve calculation of determinants, it would be computationally more efficient.

With respect to the optimization example presented in Section 9, we acknowledge that this is rather a simplistic case, as most considerations involve various weightings applied to the different subgroups and the objective function may well be non-linear [4]. We also note that the various boundary checks are quite extensive: in general, it is necessary to check

$$\sum_{k=0}^{m} \binom{m}{k} 2^{m-k} = 3^m$$

different boundary systems, each term in the sum corresponding to the number of k-dimensional boundaries, there being 2^{m-k} boundary choices (0 or 1 values for the boundary variables) for each specific choice of k variables. For the Influenza A example, this meant checking $3^5 = 243$ boundary systems. However, in principle, for a differentiable objective function, the method of Lagrange multipliers can be applied with software such as MAPLE and an algorithm implemented to take care of all the boundary checks. We are currently working on this and intend to address the optimization problem more generally in future work.

In conclusion, the methods in this paper provide tools to extend the quantitative science of infectious disease elimination through vaccination.

Acknowledgements

This work was partially supported by National Institute of Health grants R01-AI32042 and T32-AI07442-08. The authors are grateful to Dr. Pauline van den Driessche for her insights into Perron–Frobenius theory and to the two anonymous referees for their helpful suggestions.

Appendix A

Proof of second statement of Theorem 2.1(b). Let $\mathbf{u}^{\mathrm{T}}$ be the positive left eigenvector associated with $\rho(A)$ and let μ be an eigenvalue with positive right eigenvector $\mathbf{v}$. Then

$$\rho(A)\mathbf{u}^{\mathrm{T}}\mathbf{v} = (\mathbf{u}^{\mathrm{T}}A)\mathbf{v} = \mathbf{u}^{\mathrm{T}}(A\mathbf{v}) = \mu\mathbf{u}^{\mathrm{T}}\mathbf{v}.$$

Since $\mathbf{u}^{\mathrm{T}}\mathbf{v} > 0$, it follows that $\mu = \rho(A)$. □

This result also shows that the positive left eigenvector is orthogonal to all right eigenvectors of eigenvalues other than the spectral radius. The roles of left and right eigenvectors can be interchanged in the above arguments.

Proof of Theorem 2.2. Consider the action of M on the subspace basis $\{\mathbf{e}_i \otimes [1\ \theta]^{\mathrm{T}} : i = 1, \ldots, m\}$, where the $\mathbf{e}_i$ constitute the standard basis of $\mathbb{R}^m$. Denote the jth columns of M and R by $\mathbf{m}_j$ and $\mathbf{r}_j$, respectively. Recall that $\psi = 1 - \theta\phi$. Then

$$\begin{aligned} M\left(\mathbf{e}_i \otimes \begin{bmatrix} 1 \\ \theta \end{bmatrix}\right) = \mathbf{m}_{2i-1} + \theta\mathbf{m}_{2i} &= \begin{bmatrix} R_{1i}(1-f_i) \\ R_{1i}\theta(1-f_i) \\ \vdots \\ R_{mi}(1-f_i) \\ R_{mi}\theta(1-f_i) \end{bmatrix} + \theta \begin{bmatrix} R_{1i}\phi f_i \\ R_{1i}\theta\phi f_i \\ \vdots \\ R_{mi}\phi f_i \\ R_{mi}\theta\phi f_i \end{bmatrix} \\ &= (1-f_i)\mathbf{r}_i \otimes \begin{bmatrix} 1 \\ \theta \end{bmatrix} + \theta\phi f_i \mathbf{r}_i \otimes \begin{bmatrix} 1 \\ \theta \end{bmatrix} = (1-\psi f_i)\mathbf{r}_i \otimes \begin{bmatrix} 1 \\ \theta \end{bmatrix} = d_i\mathbf{r}_i \otimes \begin{bmatrix} 1 \\ \theta \end{bmatrix} \\ &= d_i \sum_{k=1}^{m} R_{ki}\mathbf{e}_k \otimes \begin{bmatrix} 1 \\ \theta \end{bmatrix}. \end{aligned}$$

Note that $d_i\mathbf{r}_i$ is the ith column of RD. The final expression above shows that M maps the subspace into itself. As the m-dimensional subspace is isomorphic to $\mathbb{R}^m$, the linear transformation may be represented on $\mathbb{R}^m$ by the matrix RD. It follows that $R_{\mathbf{f}} = \rho(M) = \rho(RD)$. □

Proof of Proposition 3.1. Let $R_{\mathbf{f}} = \rho(RD) = 1$. As $\mathbf{d} = R_0^{-1}\mathbf{1}$ corresponds to a threshold solution, it follows from Theorem 2.1(e) that any other $\mathbf{d}$ representing a threshold solution must have at least one component less than R_0^{-1} and one greater than R_0^{-1}. Hence, $d_{\min} \leqslant R_0^{-1} \leqslant d_{\max}$ (equality holding for $R_0^{-1}\mathbf{1}$). The bounds for $\mathbf{f}$ follow from $\mathbf{d} = \mathbf{1} - \psi\mathbf{f}$. □

Proof of Proposition 3.2. The threshold surface is the continuous image of the compact, connected set $\{\|\mathbf{v}\| = 1 : v_i \geqslant 0\}$ and is therefore a compact, connected subset of $\mathbb{R}^m$. Compactness is preserved on intersection with $[0,1]^m$. □

Proof of Theorem 3.3. We exploit the multi-linear property of the determinant function with respect to columns. Fix $\psi \in [1 - R_0^{-1}, 1]$ and recall that $d_j = 1 - \psi f_j$. For any $j \in \{1, \ldots, m\}$, expand about the jth column to obtain

$$\begin{aligned}\bar{\Phi}(\mathbf{f}) &= d_j \begin{vmatrix} d_1R_{11}-1 & \cdots & R_{1j} & \cdots & d_mR_{1m} \\ \vdots & & \vdots & & \vdots \\ d_1R_{j1} & \cdots & R_{jj} & \cdots & d_mR_{jm} \\ \vdots & & \vdots & & \vdots \\ d_1R_{m1} & \cdots & R_{mj} & \cdots & d_mR_{mm}-1 \end{vmatrix} \\ &+ \begin{vmatrix} d_1R_{11}-1 & \cdots & 0 & \cdots & d_mR_{1m} \\ \vdots & & \vdots & & \vdots \\ d_1R_{j1} & \cdots & -1 & \cdots & d_mR_{jm} \\ \vdots & & \vdots & & \vdots \\ d_1R_{m1} & \cdots & 0 & \cdots & d_mR_{mm}-1 \end{vmatrix} \\ &= d_j \det A_j - C_j,\end{aligned}$$

where A_j is the $m \times m$ matrix obtained by replacing the jth column of $RD - I$ with that of R and C_j is the (j,j)-cofactor of $RD - I$. Since D is diagonal, both $\det A_j$ and C_j are independent of d_j, hence f_j. Thus:

$$\frac{\partial \bar{\Phi}}{\partial f_j} = -\psi \det A_j, \quad \forall j = 1, \ldots, m.$$

For regularity, it suffices to show that the gradient $\nabla \bar{\Phi}$ has maximal rank, i.e., does not vanish on $\bar{N} = \bar{\Phi}^{-1}(0)$ [16]. If there are any points where the gradient is zero, they satisfy $\det A_j = 0$, $\forall j = 1, \ldots, m$. It follows that at any such critical point $\mathbf{f} \in \bar{N}$,

$$C_j = d_j \det A_j - \bar{\Phi}(\mathbf{f}) = 0, \quad \forall j = 1, \ldots, m.$$

For any square matrix B, with characteristic polynomial $p_B(z) = \det(B - zI)$, the coefficient of z, namely $p'_B(0)$, is equal to the (negative) sum of the principal cofactors. Hence:

$$\begin{aligned} p_{RD-I}(z) &= \det[(RD - I) - zI] = \det[RD - (z+1)I] = p_{RD}(z+1) \\ &\Rightarrow p'_{RD}(1) = p'_{RD-I}(0) = -\sum_{j=1}^{m} C_j = 0. \end{aligned}$$

This is a contradiction: $p'_{RD}(1)$ is non-zero because $\rho(RD) = 1$ is a *simple* eigenvalue via Perron–Frobenius. Hence, there are no critical points as claimed and the submanifold is regular. □

References

[1] M.E. Halloran, I.M. Longini Jr., C.J. Struchiner, Design and interpretation of vaccine field studies, in: A.S. Monto, S.B. Thacker (Eds.), Epidemiologic Reviews: vaccines, vol. 21, 1999, p. 73.
[2] K. Dietz, Discussion of estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data, Appl. Statist. 50 (2001) 283.
[3] 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.
[4] I.M. Longini Jr., E. Ackerman, L. Elveback, An optimization model for influenza A epidemics, Math. Biosci. 38 (1978) 141.

[5] Z. Agur, Y.L. Danon, R.M. Anderson, L. Cojocaru, R.M. May, Measles immunization strategies for an epidemiologically heterogeneous population: the Israeli case study, Proc. Roy. Soc. Lond. B 252 (1993) 81.
[6] H.W. Hethcote, J.W. van Ark, Epidemiological models for heterogeneous populations: proportionate mixing, parameter estimation, and immunization programs, Math. Biosci. 84 (1987) 85.
[7] A.J.G. Cairns, Epidemics in heterogeneous populations: aspects of optimal vaccination policies, IMA J. Math. Appl. Med. Biol. 6 (1989) 137.
[8] N.G. Becker, D.N. Starczak, Optimal vaccination strategies for a community of households, Math. Biosci. 139 (1997) 117.
[9] T. Britton, Epidemics in heterogeneous communities: estimation of R_0 and secure vaccination coverage, J. Roy. Statist. Soc. B 63 (2001) 705.
[10] I.M. Longini Jr., K. Sagetalian, W.N. Rida, M.E. Halloran, Optimal vaccine trial design when estimating vaccine efficacy for susceptibility and infectiousness from multiple populations, Statist. Med. 17 (1998) 1121.
[11] H. Andersson, T. Britton, Stochastic Epidemic Models and their Statistical Analysis, Lecture Notes in Statistics, 151, Springer, New York, 2000.
[12] C.R. Rao, M.B. Rao, Matrix Algebra and Its Applications to Statistics and Econometrics, World Scientific, Singapore, 1998.
[13] J.J. Hunter, Mathematical Techniques of Applied Probability, vol. 1, Academic Press, New York, 1983.
[14] A.L. Adams, D.C. Barth-Jones, S.E. Chick, J.S. Koopman, Simulations to evaluate HIV vaccine trial designs, Simulation 71 (4) (1998) 228.
[15] M.E. Halloran, L. Watelet, C.J. Struchiner, Epidemiologic effects of vaccines with complex direct effects in an age-structured population, Math. Biosci. 121 (1994) 193.
[16] P.J. Olver, Applications of Lie Groups to Differential Equations, Springer, New York, 1986.
[17] R.S. Varga, Matrix Iterative Analysis, Springer Series in Computational Mathematics, 27, second ed., Springer, New York, 2000.
[18] D. Greenhalgh, K. Dietz, Some bounds on estimates for reproductive ratios derived from the age-specific force of infection, Math. Biosci. 124 (1994) 9.
[19] R. Bhatia, Matrix Analysis, Springer, New York, 1997.
[20] A.D. Barbour, MacDonald's model and the transmission of bilharzia, Trans. Roy. Soc. Trop. Med. Hyg. 72 (1978) 6.
[21] H.W. Hethcote, Qualitative analyses of communicable disease models, Math. Biosci. 28 (1976) 335.
[22] C. Castillo-Chavez, Z. Feng, Global stability of an age-structure model for TB and its applications to optimal vaccination strategies, Math. Biosci. 151 (1998) 135.
[23] K.P. Hadeler, J. Müller, Vaccination in age structured populations II: optimal strategies, in: V. Isham, G. Medley (Eds.), Models for Infectious Human Diseases: their Structure and Relation to Data, Cambridge University, Cambridge, 1993, p. 102.
[24] J. Müller, Optimal vaccination patterns in age-structured populations, SIAM J. Appl. Math. 59 (1999) 222.
[25] D. Greenhalgh, Vaccination campaigns for common childhood diseases, Math. Biosci. 100 (1990) 201.
[26] I.M. Longini, M.E. Halloran, A frailty mixture model for estimating vaccine efficacy, Appl. Statist. 45 (1996) 165.
[27] C.P. Farrington, M.N. Kanaan, N.J. Gay, Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data—with discussion, Appl. Statist. 50 (2001) 251.
[28] I.M. Longini, M.E. Halloran, A. Nizam, M. Wolff, P.M. Mendelman, P. Fast, R.B. Belshe, Estimation of the efficacy of live, attenuated influenza vaccine from a two-year, multi-center vaccine trial: implications for influenza epidemic control, Vaccine 18 (2000) 1902.
[29] A.S. Monto, J.A. Davenport, J.A. Napier, T.F. Francis, Effect of vaccination of a school-age population upon the course of an A2 Hong Kong influenza epidemic, Bull. World Health Org. 41 (1969) 537.
[30] T.A. Reichart, N. Sugaya, D.S. Fedson, W.P. Glezen, L. Simonsen, M. Tashiro, The Japanese experience with vaccinating schoolchildren against influenza, N. Eng. J. Med. 344 (2001) 889.