Journal of Mathematical Biology manuscript No.
(will be inserted by the editor)

A simple SIS epidemic model with a backward bifurcation

P. van den Driessche*, James Watmough**

Department of Mathematics and Statistics, University of Victoria, Victoria, B.C. V8W 3P4,
e-mail: pvdd@math.uvic.ca,watmough@math.uvic.ca

April 26, 2000

Abstract It is shown that an SIS epidemic model with a non-constant contact rate may have multiple stable equilibria, a backward bifurcation and hysteresis. The consequences for disease control are discussed. The model is based on a Volterra integral equation and allows for a distributed infective period. The analysis includes both local and global stability of equilibria.

Key words SIS epidemic model – multiple equilibria – backward bifurcation – nonlinear Volterra integral equation – distributed delay

1 Introduction

Classical disease transmission models with constant contact rates typically have a unique stable equilibrium. The reproduction number of the disease, that is the number of new infective individuals produced by a single infective individual during its lifetime, serves as a threshold parameter. If the reproduction number is below one, then the disease dies out. Otherwise, the fraction of infective individuals in the population rises and the disease approaches an endemic level. In this paper, we show that if the contact rate is not constant, but rather a function of the fraction of infective individuals, then a constant population SIS epidemic model may have multiple stable equilibria, a backward bifurcation and hysteresis.

In an SIS model, the population is divided into two disjoint classes, individuals susceptible to infection and infective individuals. The dynamics of the disease are specified by two functions, the contact rate and the distribution of the infective period. Susceptible individuals become infective after contact with infective individuals. Infective individuals return to the susceptible class after an infective period. Models of this type not only form a basis for most other epidemiological models, but are directly applicable to gonnorhea and other sexually transmitted diseases or bacterial infections where exposed individuals typically become infective within 24 hours and do not gain immunity to the disease once the infection is passed (see e.g. [13, p. 18]).

* Research supported in part by an NSERC Research Grant and the University of Victoria Committee on faculty, research and travel.
** Research supported by an NSERC Postdoctoral Fellowship.
Correspondence to: James Watmough

The contact rate is the average number of adequate encounters per unit time between an infective individual and another individual in the population [10]. If this rate is constant, then the incidence of new infections is proportional to the product of the number of infective individuals and the number of susceptible individuals. The contact rate may depend on the fraction of infective individuals for several reasons, such as saturation, an increased likelihood of infection from multiple exposures, or behavioural changes in the population as the fraction of infective individuals rises. Another likely scenario is that the contact rate does not depend directly on the fraction of infective individuals in the population, but rather on the severity of infection in the infecting individual. This situation could be modelled using a non-constant contact rate provided a link can be made between the fraction of infective individuals in the population and the severity of infection in each individual.

The model of this paper is based on a Volterra integral equation, which includes ordinary differential equations (ODE) and delay differential equations as special cases. If the distribution of the infective period is exponential, then the integral equation reduces to an ODE; if the infective period is constant, then a differential delay model results. The analysis of the SIS model applies to a very general class of distributions. However, for more complex models it is often necessary to restrict attention to one or both of these special cases, even though an intermediate distribution is more realistic (see e.g. [1, p. 60]).

The main focus of this paper is the dependence of the number of equilibria and their stability on the reproduction number of the disease. The term 'reproduction number' is commonly reserved for the expected number of new infections produced by a single infective individual introduced into a *disease free* population. Hence, it is a measure of the invasive strength of the disease. Since the reproduction number is based on conditions at the disease free equilibrium, it does not necessarily indicate the ability of the disease to persist at endemic levels. We find that there are two critical values of the reproduction number, R_o^c and R_o^m, that relate to disease persistence. For reproduction numbers below R_o^c the disease free equilibrium is globally stable, and for reproduction numbers above R_o^m, there is a unique, globally stable endemic equilibrium. In the classical case, both critical values are equal to one and there is always a single globally stable equilibrium (see e.g. Figure 2(a)). If the contact rate is a function of the number of infective individuals, there may be multiple equilibria for reproduction numbers between these two values (see e.g. Figure 2(d)).

Most previous studies of multiple steady states in epidemic models have focused on multiple-group models. That is, models with either multiple susceptible classes [7, 15], multiple infective classes [4] or both [5, 8, 14]. Rather than specify a non-constant contact rate, these models specify the dependence of mixing between groups on the total population. The analysis of these models is more difficult than that of the simple model presented here. Each of these papers shows that backward bifurcations, which are indicative of multiple steady states, can occur for certain types of mixing behaviour. Previous analysis of single group epidemic models with a non-constant contact rate has focused on the existence of periodic solutions of SIRS and SEIRS models [9, 11, 16, 17]. Although multiple steady states were found in these papers, the question of what features of the contact rate lead to a backward bifurcation, multiple steady states and hysteresis was not explored.

The model is developed in Section 2. Local and global asymptotic stability of equilibrium solutions are explored in Section 3. Specific examples illustrating the main results of the paper are given in Section 4. A discussion of the consequences for disease control and of earlier studies of multiple steady states in epidemiological models is deferred to Section 5.

2 Model Formulation

Let $I(t)$ and $S(t) = 1 - I(t)$ be the fraction of the population in each of the two disjoint classes, infective and susceptible, at time $t \geq 0$. It is assumed that there is no disease related death and that the birth and death rate constants are both equal to $b > 0$. From this assumption, it follows that the size of the population remains constant.

It is assumed that all newborns are susceptible and that susceptible individuals become infective at a rate $\lambda(I(t))I(t)S(t)$. The contact rate, $\lambda(I)$, is assumed to satisfy the following conditions:

(A1) $\lambda(I) > 0$ on $(0, 1]$, $\lambda(0) \geq 0$, $\lambda(I)$ is continuous and the rate of infection $\lambda(I)I(1 - I)$ has a continuous derivative on $[0, 1]$.

Once infective, individuals either die or return to the susceptible class. Consider the individuals infected at time zero that survive to time t. Let $P(t)$ be the fraction of these individuals that remain infective through to time t. Since death occurs at the rate b, independent of infection, the fraction of all individuals infected at time zero that remain infective at least t time units before either dying or returning to the susceptible class is $P(t)e^{-bt}$. The integral $\tau = \int_0^\infty P(u)e^{-bu}\,du$ is the mean time an individual remains infective. It is assumed that

(A2) $P(t) \geq 0$ and non-increasing for $t \geq 0$, $P(0^+) = 1$ and $\tau \in (0, \infty)$.

Finally, let $I_o(t)$ represent the fraction of individuals that were infective at time zero and have remained infective through to time t. It is assumed that

(A3) $I_o(t) \geq 0$, non-increasing and differentiable with $I_o(0) \in [0, 1]$ and $\lim_{t\to\infty} I_o(t) = 0$.

With the above assumptions, the fraction of infective individuals at time $t \geq 0$ is given by the Volterra integral equation

$$I(t) = I_o(t) + \int_0^t \lambda(I(u))I(u)(1 - I(u))P(t - u)e^{-b(t-u)}\,du, \qquad t \geq 0, \tag{1}$$

where $S(t)$ has been replaced by $1-I(t)$. The integral sums the individuals that entered the infective class at time $u \geq 0$ and have remained infective through to time t.

Near the disease free equilibrium (i.e., I near zero, S near one), a single infective individual makes approximately $\lambda(0)$ adequate contacts per unit time. The expected infective period for this same individual is τ. Hence, the reproduction number of the disease is the product $\tau\lambda(0)$. It is apparent that the reproduction number makes a suitable parameter for study provided $\lambda(0)$ is non-zero. However, even in the special case $\lambda(0) = 0$, much of our analysis is the same. In order to simplify the presentation it is advantageous to make the following definitions. Let

$$R_o = \tau\lambda_o, \tag{2}$$

$$f(I) = \frac{1}{\lambda_o}\lambda(I)(1 - I), \qquad 0 \leq I \leq 1, \tag{3}$$

and

$$\tilde{P}(t) = \frac{1}{\tau}P(t)e^{-bt}, \qquad t \geq 0, \tag{4}$$

in which

$$\lambda_o = \begin{cases} \lambda(0) & \text{if } \lambda(0) > 0, \\ 1 & \text{if } \lambda(0) = 0. \end{cases} \tag{5}$$

Note that R_o is the reproduction number provided $\lambda(0)$ is non-zero. With these definitions, Equation (1) becomes

$$I(t) = I_o(t) + R_o \int_0^t I(u) f(I(u)) \tilde{P}(t-u)\, du. \tag{6}$$

By assumptions (A2), $\tilde{P}(t)$ is non-negative and non-increasing for all $t \geq 0$ and $\int_0^\infty \tilde{P}(u)\, du = 1$. Further, by Equation (5), $f(0)$ is either zero if $\lambda(0) = 0$, or one if $\lambda(0) > 0$.

The function $f(I)$ can be interpreted as the incidence of infection per infective individual, with the scale given by λ_o. To illustrate the rescaled functions and parameters, consider the three examples illustrated in Figure 1. If λ is constant, then $f(I)$ decreases monotonically from one to zero as shown in Figure 1(a). In Figure 1(b) and Figure 1(c) the contact rate is linear in I. For Figure 1(b), $\lambda(0) = f(0) = 0$.

3 Equilibria and asymptotic stability

Consider the model given by Equation (6) under assumptions (A1), (A2) and (A3). This section focuses on the asymptotic stability of equilibrium (constant) solutions. Before proceeding with that analysis, we present the following theorem, which states that the model is well posed, i.e., a unique solution $I(t)$ exists with both $I(t)$ and $S(t) = 1 - I(t)$ remaining non-negative for all time.

Theorem 1 *For the disease transmission model of Equation (6), with assumptions (A1), (A2) and (A3), there is a unique, differentiable solution for each $I_o(t)$. Further, if $0 \leq I_o(0) \leq 1$, then this solution remains in the interval $[0, 1]$ for all $t \geq 0$.*

The existence and uniqueness of solutions follow from the results in Miller [20, Ch. 2]. The bounds on the solution can be established by an extension of the technique used for constant $\lambda(I)$ in Hethcote and van den Driessche [12, Th. 2.1]. Note that for the results of Theorem 1 it is sufficient that $\lambda(I)$ be Lipschitz continuous. Further, the requirement for differentiability and continuity of $I_o(t)$ can be relaxed at the expense of the differentiability and continuity of the solution.

3.1 Number of equilibria

By Theorem 1, there is a unique solution to Equation (6) for each choice of $I_o(t)$. If $I(t) = \bar{I}$ is a constant solution to Equation (6) corresponding to $I_o(t) = I_o(t; \bar{I})$, then $I_o(t; \bar{I})$ must satisfy

$$\bar{I} = I_o(t; \bar{I}) + R_o \bar{I} f(\bar{I}) \int_0^t \tilde{P}(t-u)\, du. \tag{7}$$

In addition, $\lim_{t\to\infty} I_o(t; \bar{I}) = 0$ by assumptions (A3). Applying the limit $t \to \infty$ to Equation (7), $\bar{I}$ must satisfy

$$\bar{I} = R_o \bar{I} f(\bar{I}), \qquad 0 \leq \bar{I} \leq 1. \tag{8}$$

Conversely, Equation (7) defines a function $I_o(t; \bar{I})$ satisfying assumptions (A3) for each $\bar{I}$ satisfying Equation (8). Thus, $\bar{I}$ is an equilibrium solution of the model if and only if $\bar{I}$ is a solution of Equation (8). The equilibrium $\bar{I} = 0$ is referred to as the disease free equilibrium, and an equilibrium

$\bar{I} = I_e > 0$ is referred to as an endemic equilibrium. The fraction of infective individuals at an endemic equilibrium, I_e, is given implicitly as a function of R_o by the relation

$$R_o f(I_e) = 1. \tag{9}$$

The model may have multiple equilibria over a range of values of R_o. As can be seen from Figure 1, multiple endemic equilibria occur for a given value of R_o if and only if the graph of $f(I)$ has multiple intersections with a horizontal line of height $1/R_o$. To classify these points rigorously, let

$$X = \{0\} \cup \{I \in (0,1) \mid I \text{ is a critical point of } f(I)\}, \tag{10}$$

and define

$$R_o^c = \frac{1}{\max_{I \in X} f(I)}, \tag{11}$$

$$R_o^m = \frac{1}{\min_{I \in X} f(I)}. \tag{12}$$

Thus, $1/R_o^c$ is the global maximum of $f(I)$, and $1/R_o^m$ is the smallest local extremum of $f(I)$ other than the right endpoint. These definitions are illustrated in Figure 1. Obviously, $R_o^c \leq R_o^m$. By assumptions (A1) and Equation (3), $f(I)$ is continuous and non-negative on $[0,1]$, positive on $(0,1)$ and has the global minimum $f(1) = 0$. Thus $0 < R_o^c < \infty$. Recall that $f(0) \in \{0,1\}$. If $f(0) = 1$ then $1 \leq R_o^m < \infty$, but if $f(0) = 0$ then R_o^m is undefined $(+\infty)$ as shown in Figure 1(b). The properties of $f(I)$ noted above are also sufficient to prove the following theorem.

Theorem 2 *The disease transmission model of Equation (6), with assumptions (A1), (A2) and (A3)*

(i) always has the disease free equilibrium,
(ii) has no endemic equilibrium if $R_o < R_o^c$,
(iii) has at least one endemic equilibrium if $R_o > R_o^c$, and
(iv) has exactly one endemic equilibrium if $R_o > R_o^m$.

R_o^c and R_o^m (if it is defined) correspond to the minimum and maximum values of R_o where the number and stability of equilibrium solutions of the model change.

3.2 Local asymptotic stability

By the result of Miller [19, Th. 4], the equilibrium $\bar{I}$ is locally asymptotically stable if the characteristic equation,

$$1 = R_{\bar{I}} \int_0^\infty e^{-zu} \tilde{P}(u)\, du, \tag{13}$$

where

$$R_{\bar{I}} = R_o \left(\bar{I} f(\bar{I})\right)' = R_o \left(f(\bar{I}) + \bar{I} f'(\bar{I})\right), \tag{14}$$

has no root with non-negative real part. The coefficient $R_{\bar{I}}$ arises from the linearization of Equation (6) about the equilibrium solution $\bar{I}$. Note that by assumptions (A1), $R_{\bar{I}}$ is bounded.

If Equation (13) has a root with positive real part then the equilibrium $\bar{I}$ is unstable. Since all perturbations from the disease free equilibrium are positive, instability of the disease free equilibrium means that all perturbations from this equilibrium initially grow exponentially [6]. However, for an endemic equilibrium, I_e, instability means only that there are arbitrarily small perturbations of $I_o(t)$ from $I_o(t; I_e)$ for which $I(t)$ is not asymptotic to I_e [2]. Indeed, an unstable endemic equilibrium may in fact be stable to some perturbations (e.g. [2, p. 42]).

Cooke and Yorke [3] examined solutions to Equation (13) under the additional assumption that $\tilde{P}(u)$ has compact support. Theorem 3, below, states that their results also hold for the less restrictive assumptions (A2). The proof of the theorem relies on two lemmas. Lemma 1 is used to show that Equation (13) has no purely imaginary roots. Lemma 2 then shows that all roots with non-negative real parts are in fact real and goes on to examine the real roots.

Lemma 1 *If $g(u)$ is non-negative, bounded and non-increasing for $u > 0$, and $0 < \int_0^\infty g(u)\,du < \infty$, then*

$$\int_0^\infty g(u)\sin(yu)\,du = 0 \tag{15}$$

if and only if either $y = 0$ or $g(u)$ is constant on every subinterval $2\pi n/y < u < 2\pi(n+1)/y$ where n is a non-negative integer.

Proof If $y = 0$ then Equation (15) is true. It remains to show that if $y \neq 0$ then Equation (15) holds if and only if $g(u)$ is constant on each subinterval $2\pi n/y < u < 2\pi(n+1)/y$.

Since $\sin(-yu) = -\sin(yu)$, it is sufficient to assume that $y > 0$. The case $y < 0$ follows similarly. The integral in Equation (15) is the sum of the infinite sequence $a_n = \int_{2\pi n/y}^{2\pi(n+1)/y} g(u)\sin(yu)\,du$, which is convergent since $g(u)$ is non-negative and integrable. From the properties of sine and $g(u)$ it follows that

$$a_n = \int_{2\pi n/y}^{2\pi(n+1/2)/y} (g(u) - g(u+\pi/y))\sin(yu)\,du \geq 0. \tag{16}$$

Hence, Equation (15) holds if and only if $a_n = 0$ for each n. By Equation (16), $a_n = 0$ if and only if $g(u+\pi/y) = g(u)$ in the interval $2\pi n/y < u < 2\pi(n+1/2)/y$. Since $g(u)$ is non-increasing, this is equivalent to requiring that $g(u)$ be constant on the larger subintervals $2\pi n/y < u < 2\pi(n+1)/y$, which completes the proof. □

Lemma 2 *If $P(u)$ satisfies assumptions (A2) and $\tilde{P}(u)$ is defined by Equation (4), then for Equation (13), any non-real root has negative real part and there is at most one real root. Further, if it exists, this real root is unique, simple and positive if $R_{\bar{I}} > 1$, negative if $R_{\bar{I}} < 1$, and zero if $R_{\bar{I}} = 1$.*

Proof Let $z = x + iy$ be a root of Equation (13) with x and y both real. Then x and y must be roots of the two equations

$$0 = R_{\bar{I}} \int_0^\infty e^{-xu}\tilde{P}(u)\sin(yu)\,du, \tag{17}$$

$$1 = R_{\bar{I}} \int_0^\infty e^{-xu}\tilde{P}(u)\cos(yu)\,du. \tag{18}$$

First, consider a root with non-negative real part (i.e., $x \geq 0$). If $R_{\bar{I}} = 0$ then Equation (18) has no root. If $R_{\bar{I}} \neq 0$, then by Lemma 1, Equation (17) has a root if and only if $y = 0$ or $e^{-xu}\tilde{P}(u)$ is

constant on every subinterval $2\pi n/y < u < 2\pi(n+1)/y$. In the latter case, Equation (18) has no root, since the integral is zero. Thus, $x \geq 0$ implies that $y = 0$.

Now, consider a real root, x, (i.e., $y = 0$) of Equation (13), which must satisfy

$$1 = R_{\bar{I}} \int_0^\infty e^{-xu} \tilde{P}(u)\, du. \tag{19}$$

If $x \geq -b$, where $b > 0$ is the birth rate, then by Equation (4) and assumptions (A2), the integral in Equation (19) is positive, bounded, and decreasing in x. Hence, the root of Equation (19), if it exists, is unique and simple. Since $\tilde{P}(u)$ is positive and $\int_0^\infty \tilde{P}(u)\, du = 1$, there is no real root for $R_{\bar{I}} \leq 0$ and the real root is positive if $R_{\bar{I}} > 1$, negative if $0 < R_{\bar{I}} < 1$ and zero if $R_{\bar{I}} = 1$. □

As a direct consequence of Lemma 2, an equilibrium, $\bar{I}$ is locally stable if $R_{\bar{I}} < 1$ and unstable if $R_{\bar{I}} > 1$. The details of local stability are given by the following theorem.

Theorem 3 *For the model of Equation (6), with assumptions (A1), (A2) and (A3),*

(i) if $\lambda(0) > 0$, then the disease free equilibrium is locally asymptotically stable if $R_o < 1$, but unstable if $R_o > 1$,
(ii) if $\lambda(0) = 0$, then the disease free equilibrium is locally asymptotically stable for all R_o,
(iii) an endemic equilibrium I_e is locally asymptotically stable if $f'(I_e) < 0$, but unstable if $f'(I_e) > 0$.

Proof The value of $R_{\bar{I}}$ can be computed from Equation (14) with $f(I)$ given by Equation (3). If $\bar{I} = 0$ and $\lambda(0) > 0$, then $R_{\bar{I}} = R_o$. Hence, the disease free equilibrium is locally stable if $R_o < 1$ and unstable for $R_o > 1$, proving *(i)*. If $\bar{I} = 0$ and $\lambda(0) = 0$, then $R_{\bar{I}} = 0$. Although Miller [19] does not deal explicitly with this case, the techniques used therein can be applied to prove statement *(ii)*. For an endemic equilibrium $\bar{I} = I_e > 0$, $R_{I_e} = 1 + R_o I_e f'(I_e)$ and $\mathrm{sgn}(R_{I_e} - 1) = \mathrm{sgn}(f'(I_e))$. Thus, the endemic equilibrium is locally stable if $f'(I_e) < 0$ and unstable if $f'(I_e) > 0$, proving *(iii)*. □

An important corollary of Theorems 2 and 3 is that the number and local stability of endemic equilibria as functions of R_o can be determined solely from $f(I)$ and $f'(I)$, respectively. Since $f(I)$ is continuous, the graph of the points (R_o, I_e) satisfying Equation (9) is a continuous curve composed of branches where $f(I)$ is either increasing (unstable endemic equilibria) or decreasing (locally stable endemic equilibria). These equilibria can be plotted either as in Figure 1, which shows plots of $f(I)$, or as bifurcation diagrams as shown in Figure 2.

The stability of the disease free equilibrium is determined by the value of $R_o f(0) = \tau\lambda(0)$. There are only two cases to consider, either $\lambda(0) = 0$, indicating the disease free equilibrium is locally stable for all values of R_o, or $\lambda(0) > 0$, indicating a bifurcation at $R_o = 1$. This bifurcation is in fact an exchange of stability between the disease free equilibrium and a branch of endemic equilibria intersecting the R_o-axis. If $f'(0) < 0$ then a branch of locally stable endemic equilibria defined for $R_o > 1$ intersects the axis at $R_o = 1$ (see Figure 2(a)). This situation is referred to as a forward bifurcation. If, $f'(0) > 0$ then a branch of unstable endemic equilibria intersects the axis at $R_o = 1$, with the equilibria defined for $R_o < 1$ (see Figure 2(c)). This is a backward bifurcation. If $f'(0) = 0$, then the nature of the bifurcation can be deduced from the first non-zero derivative of $f(I)$. In terms of the contact rate $\lambda(I)$, the direction of this bifurcation is given by the following theorem.

Theorem 4 *For the model of Equation (6), with $\lambda(0) > 0$ and assumptions (A1), (A2) and (A3), the bifurcation at $R_o = 1$, $\bar{I} = 0$ is in the forward direction if $\lambda'(0) < \lambda(0)$ and in the backward direction if $\lambda'(0) > \lambda(0)$. Further, if the model has a backward bifurcation, then $R_o^c < 1$ and there are multiple stable equilibria for $R_o^c < R_o < 1$.*

Proof The criteria for the direction of the bifurcation follow directly from the definition of $f(I)$. If there is a backward bifurcation at $R_o = 1$, $\bar{I} = 0$, then there is a branch of unstable endemic equilibria defined for $R_o < 1$. Since these equilibria are roots of Equation (9), the result that $R_o^c < 1$ follows from Equation (11). Further, since $f(I)$ is continuous and $f(1) = 0$ there is a second branch of equilibria in the range $R_o^c < R_o < 1$ with $f'(\bar{I}) < 0$, which are therefore locally stable. By Theorem 3, the disease free equilibrium is also locally stable for $R_o < 1$. Hence, there are at least two locally stable equilibria for $R_o^c < R_o < 1$. □

In terms of the contact rate, the endemic equilibria are the points satisfying $\lambda(I) = L(I) = \tau^{-1}(1 - I)^{-1}$. This follows from Equation (9) and the definitions of Equations (3) and (5). By Theorem 3, an endemic equilibrium is stable if $L(I)$ is increasing faster than $\lambda(I)$ at the equilibrium. Since $L(I)$ is unbounded as I approaches 1, and $\lambda(I)$ is bounded (by assumptions (A1)), a sufficient condition for the existence of at least one stable endemic equilibria is that $\lambda(I) > L(I)$ over some range of I. Since $\tau\lambda(0) < 1$ implies a stable disease free equilibrium, the two inequalities together give a criterion for the existence of multiple stable equilibria.

3.3 Global asymptotic stability

Theorem 5 *For the model of Equation (6), with assumptions (A1), (A2) and (A3), the disease free equilibrium is globally asymptotically stable for $R_o < R_o^c$.*

Proof By Brauer's corollary [2, p. 35] to a theorem of Londen [18, Th. 1], if $I(t)$ is a bounded solution of Equation (6) then $\lim_{t\to\infty} I(t)$ exists and is equal to a root, $\bar{I}$, of Equation (8). By Theorem 1, all solutions to Equation (6) are bounded and so must be asymptotic to equilibrium solutions. However, if $R_o < R_o^c$ then Equation (8) has the unique equilibrium $\bar{I} = 0$. Hence, all solutions must be asymptotic to zero and the disease free equilibrium is globally asymptotically stable. □

Theorem 6 *If, in addition to assumptions (A1), (A2) and (A3), $\lambda(0) > 0$, and $R_o > R_o^m$, then the unique endemic equilibrium, I_e, guaranteed by Theorem 2, is globally stable in the sense that all positive solutions of Equation (6) are asymptotic to I_e.*

Proof The proof again uses the corollary of Brauer referred to in Theorem 5. If $R_o > R_o^m$, then Equation (8) has two roots, $\bar{I} = 0$ and $\bar{I} = I_e > 0$. Since the disease free equilibrium is unstable for $R_o > R_o^m \geq 1$, any solution of Equation (6) with $I_o(0) > 0$ must be asymptotic to the equilibrium solution I_e. □

Theorems 5 and 6 have some immediate consequences for disease control. By Theorem 5, the disease free equilibrium is globally asymptotically stable for $R_o < R_o^c$. Therefore, the disease will eventually die out of the population if R_o can be lowered below R_o^c. By Theorem 6, the disease remains endemic if $R_o > R_o^m$. Hence, lowering R_o or raising R_o^m are the only possible means of disease control. R_o can be lowered by decreasing the disease free equilibrium contact rate, $\lambda(0)$, decreasing the mean infective period τ, or increasing the birth/death rate b. If $R_o^c < R_o < R_o^m$, then the above control methods still apply. In addition, it is possible to control the disease by removing

infective individuals from the population. This can be determined from the model by looking for perturbations from the initial conditions $I_o(t; I_e)$ for which the solution is asymptotic to the disease free equilibrium. This last method is not applicable if $R_o > R_o^m$, since the endemic equilibrium is globally stable. Examples of these control methods applied to specific contact rates are the subject of the next section.

4 Examples

Example 4.1

Consider the contact rate

$$\lambda(I) = \kappa I^{p-1}(1-I)^{q-1}, \qquad p > 1, q \geq 1. \tag{20}$$

A contact rate of this form has been studied for an ODE model with the additional classes of latent and removed individuals [16,17]. Hethcote et al. [9] have also used this contact rate with $q = 1$ in an SIRS model with a delay in the transition from the removed class to the susceptible class. In each of these models, an equation qualitatively similar to Equation (8) determines the number of equilibrium solutions.

Since $\lambda(0) = 0$, the disease free equilibrium is always locally stable and R_o^m is undefined. The endemic equilibria are given by Equation (9) with $f(I) = \kappa I^{p-1}(1-I)^q$, $I \in [0,1]$. The function f is concave down, has a single global maximum at $I_c = (p-1)/(p+q-1) \in (0,1)$ and is zero at the endpoints (Figure 1(b) illustrates the case $\kappa = 1$, $p = 2$, and $q = 1$). The bifurcation diagram is shown in Figure 2(b) with $R_o^c = 1/f(I_c)$.

The model has multiple stable equilibria for $R_o > R_o^c$. Possible control strategies for the disease are to raise R_o^c, most probably by lowering κ, to remove infective individuals from the population, and to lower R_o by lowering the mean infective period τ or increasing the birth/death rate b.

Example 4.2

Consider the contact rate

$$\lambda(I) = \beta\left(1 + \nu I^{p-1}\right), \qquad \beta > 0, \nu > 0, p \geq 1. \tag{21}$$

This corresponds to a rate of infection of $\beta\left(IS + \nu I^p S\right)$. For low I, the bilinear term dominates, but if $\nu > 1$, then for larger I the higher order term dominates.

Case 1 The case $p = 1$ corresponds to a constant contact number, which is applicable to a mass action incidence with only bilinear interactions. Since $f(I)$ decreases monotonically from one to zero as I is increased from zero to one (see Figure 1(a)), $R_o^c = R_o^m = 1$ and there is a forward bifurcation at $R_o = 1$. The bifurcation diagram is shown in Figure 2(a). If $R_o < 1$, then the disease free equilibrium is the sole equilibrium and is globally asymptotically stable. If $R_o > 1$, then by Theorem 6, there is a unique globally stable endemic equilibrium and the disease free equilibrium is unstable. Note that there is always a single globally stable equilibrium and that the level of infective individuals at this equilibrium is a continuous function of R_o. The only option for disease control is to lower R_o below one.

Case 2 The case $p = 2$ corresponds to the inclusion of an increased rate of infection due to two exposures over a short time period. New infective individuals arise from double exposures at a rate $\beta\nu I^2 S$, while single contacts still lead to infection at the rate βSI.

The endemic equilibria satisfy Equation (9) with $f(I) = (1+\nu I)(1-I)$. This function is always concave down and has a global maximum value of $(1+\nu)^2/4\nu$ at $I_c = (\nu-1)/2\nu$ (see Figure 1(c)). If $\nu \leq 1$, $I_c \leq 0$ and $f(I)$ is decreasing over the interval $[0,1]$. This is qualitatively the same as case 1.

If $\nu > 1$, then $0 < I_c < 1$ and $f'(0) > 0$, indicating a backward bifurcation at $R_o = 1$, $\bar{I} = 0$. There are two endemic equilibria for $R_o^c < R_o < 1$, where $R_o^c = 4\nu/(1+\nu)^2$. Figure 2(c) shows these equilibria as functions of the parameter R_o. The upper branch of endemic equilibria is always locally asymptotically stable, but the lower branch is unstable. By Theorem 3, the disease free equilibrium is globally stable for $R_o < R_o^c$, locally stable for $R_o < 1$ and unstable for $R_o > 1$. When R_o is in the range $R_o^c < R_o < 1$ both the disease free equilibrium and one endemic equilibrium are locally stable. If $R_o > R_o^m = 1$ then the endemic equilibrium is globally stable.

As ν increases, R_o^c decreases and the range over which there are multiple stable equilibria increases. This information is conveyed by the graph of R_o^c versus ν for $\nu > 1$. Figure 3(a) shows this curve along with the locus of forward and backward bifurcations. The curves divide the R_o-ν parameter plane into regions where there are (I) no endemic equilibria, (II) a single globally stable endemic equilibria and (III) multiple stable equilibria. Note that the locus of saddle node bifurcations meets the locus of forward and backward bifurcations at $R_o = 1$, $\nu = 1$. The different regions of the figure correspond to the different control options available. If a population is in region II, R_o must be decreased. If a population is in region III, either R_o or ν can be decreased to return to region I. Alternately, since the disease free equilibrium is locally stable, isolation of infective individuals may be effective.

Case 3 The cases $1 < p < 2$ and $p > 2$ correspond to a more general nonlinear contact rate. The nonlinearity may be due to crowding, multiple pathways to infection, or a combination of several effects with p determined from experimental data. For the general contact rate, the endemic equilibria are given by Equation (9) with $f(I) = (1-I)(1+\nu I^{p-1})$. If $1 < p < 2$ then $f(I)$ has a single maximum in the interval $0 \leq I \leq 1$ and $f(0) = 1$. Although the derivative at zero is undefined, the endemic equilibria exist for $R_o < 1$ similar to the case $p = 2$ above. Thus the bifurcation diagram is qualitatively the same as Figure 2(c).

If $p > 2$, then $f'(0) = -1$, indicating a forward bifurcation at $R_o = 1$ for all ν. If $0 < \nu < \nu^*$, where

$$\nu^* = \left(\frac{p}{p-2}\right)^{p-2}, \tag{22}$$

then $f(I)$ is strictly decreasing and the bifurcation diagram is similar to that of the case $p = 1$ shown in Figure 2(a). However, if $\nu > \nu^*$, then $f'(I)$ has two zeros. Hence, $R_o^c < R_o^m$, and there are multiple stable equilibria for $R_o^c < R_o < R_o^m$ as shown in Figure 2(d). This example shows that although a backward bifurcation is sufficient, it is not necessary for hysteresis and multiple stable equilibria.

The range of R_o over which the model has multiple stable equilibria depends on both p and ν. Figure 3(b) shows the region of multiple stable equilibria in the R_o-ν parameter plane. The two saddle nodes coalesce at a cusp bifurcation for $\nu = \nu^*$. The options available for disease control are similar to the previous case with $\nu > 1$. However, for populations in the upper portion of region III, both stable equilibria are endemic. Hence, isolation of infective individuals will at best shift the population to a lower endemic level. A disease free population is only possible if $R_o < 1$.

Example 4.3

One supporting argument for a nonlinear contact rate is that the number of effective contacts may saturate at high infective levels due to crowding of infective individuals. Cooke and Yorke [3] suggest a contact rate similar to Equation (21) but with $\nu < 0$. The analysis for the case $\nu < 1$ also holds for $\nu \leq 0$ and so the equilibria are qualitatively similar to Figure 2(a).

Liu et al. [17] propose a saturated contact rate $\lambda(I) = \kappa I^{p-1}(1-I)^{q-1}/(1+mI^{p-1})$, $m > 0$, $\kappa > 0$, and $p, q \geq 1$ for an SIRS model without delay. Hethcote and van den Driessche [11] use a similar saturation, but with $q = 1$ for an SEIRS model without delay and an SIRS model with delay. To see the effect of saturation on the equilibria, consider the function $f_s(I) = f(I)/g(I)$ for some non-negative, increasing, differentiable function $g(I)$. The local stability of the endemic equilibrium is determined by the sign of $f_s'(I)$. Since, $f(I)$, $g'(I)$ and $g(I)$ are all positive, $f_s'(I) < 0$ for $f'(I) < 0$. The effect of saturation is to reduce the length of the unstable branches of endemic equilibria. This reduction obviously cannot introduce new bifurcations to the model and may in fact reduce the number of saddle node bifurcations appearing. Thus, saturation reduces the range of R_o over which there are multiple equilibria.

5 Discussion

A simple SIS epidemic model with a non-constant contact rate may have multiple stable equilibria. In particular, both the disease free equilibrium and an endemic equilibrium may be locally stable for the same set of parameter values. The disease free equilibrium is stable if $\tau\lambda(0) < 1$ and there is a stable endemic equilibrium if $\tau\lambda(I) > (1-I)^{-1}$ for some values of $I \in (0,1)$. If both these conditions are met for a particular set of parameter values, then there are multiple stable equilibria. Alternately, the rate of increase of the logarithm of the contact rate as the fraction of infective individuals is increased from zero can be evaluated when $\tau\lambda(0) = 1$. If this quantity is larger than one, then there is a backward bifurcation at $R_o = 1$, $I = 0$ (Theorem 4), and there are multiple stable equilibria for reproduction numbers in the range $R_o^c < R_o < 1$.

For all contact rates satisfying assumptions (A1), there is a critical value, R_o^c, of the reproduction number, R_o, below which the disease free equilibrium is globally stable. This means that the disease can be controlled by lowering R_o or raising R_o^c. There may also be a second critical value, R_o^m, above which there is a unique, globally stable endemic equilibrium. The sole option for disease control in this case is to lower the reproduction number below R_o^m. For reproduction numbers between these two critical values, there may be multiple stable equilibria.

One mechanism giving rise to multiple stable equilibria is an increase in the infection rate for susceptible individuals contacting two infective individuals over a short time. As shown in Figure 2(c), there is both a locally stable endemic and a locally stable disease free equilibria for reproduction numbers between R_o^c and $R_o^m = 1$. An additional option of disease control for reproduction numbers in this range is to temporally isolate infective individuals or implement a large scale treatment program [4]. Since the disease free equilibrium is locally stable, the disease will not return once it is eradicated.

The model of Equation (6) along with assumptions (A2) permits a very large class of delays. For our simple SIS model, the analysis was done for a general delay. However, for more detailed models, it is often necessary to restrict attention to special cases. The usual special forms of $P(t)$ are exponential and constant. If the distribution of the infective period is exponential, then $P(t) = e^{-\gamma t}$ and $I_o(t) = I_o P(t)$. Here, $\omega = 1/\gamma$ is the mean infective period barring death and $R_o = \lambda_o/(b+\gamma)$. Differentiating Equation (6) with respect to time yields the ODE

$$I'(t) = R_o I(t) f(I(t)) - (\gamma + b) I(t), \qquad \text{with } I(0) = I_o. \tag{23}$$

If the infective period is constant, then by assumptions (A2), $P(u) = 1$ for $0 \leq u \leq \omega$ and $P(u) = 0$ otherwise. If we also assume that $I_o(t) = 0$ for $t \geq \omega$, then differentiating Equation (6) yields the differential delay equation

$$I'(t) = \lambda_o I(t) f(I(t)) - \lambda_o I(t-\omega) f(I(t-\omega)) e^{-b\omega} - bI(t), \qquad t \geq \omega, \tag{24}$$

with $I(t)$ given by Equation (6) for $0 \leq \omega < t$. In this case the mean infective period, taking account of birth and death, is $(1 - e^{-b\omega})/b$ and $R_o = \lambda_o(1 - e^{-b\omega})/b$.

Previous studies of backward bifurcations in epidemic models focused on multiple susceptible and/or infective classes [4,5,7,8,14,15]. The analyses of those papers was necessarily more complex, and the results were restricted to specific examples. The rate of infection depends on the distribution of susceptible individuals among the various groups and is influenced by the assumptions about mixing between groups. For certain assumptions, a backward bifurcation may arise (see the summary in [5]). However, the more general question of the existence of multiple stable equilibria remains open. Although multi-group models are undoubtably more realistic, by focusing on an SIS model, our results may be stated in terms of general properties of the contact rate.

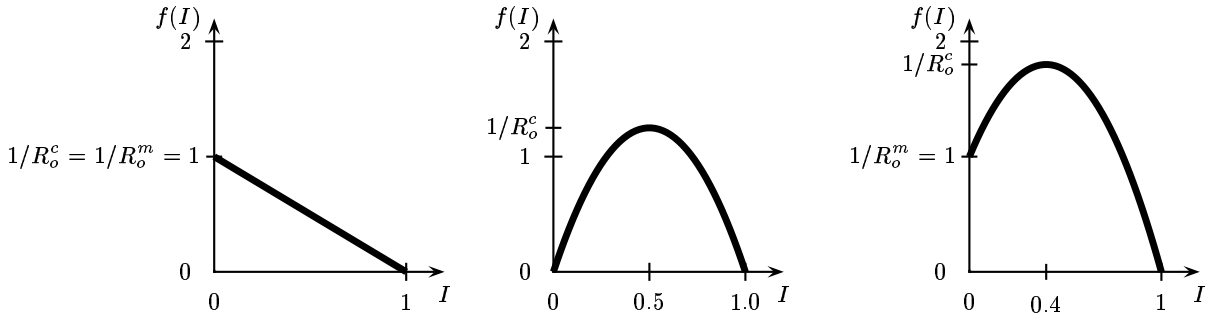
Previous modelling with a non-constant contact rate focused on the existence of periodic solutions rather than on the existence of multiple stable equilibria [9,11,16,17]. Although the model of Equation (6) does not have periodic solutions (see the proof of Theorem 5), an SIS model with a variable population size and constant infective period can have periodic solutions over a small parameter range [12]. By focusing on a constant population SIS model, the results of this paper hold for a general distribution of infective periods.

Acknowledgements The authors thank Christopher Kribs-Zaleta for providing them with a preprint of [15].

References

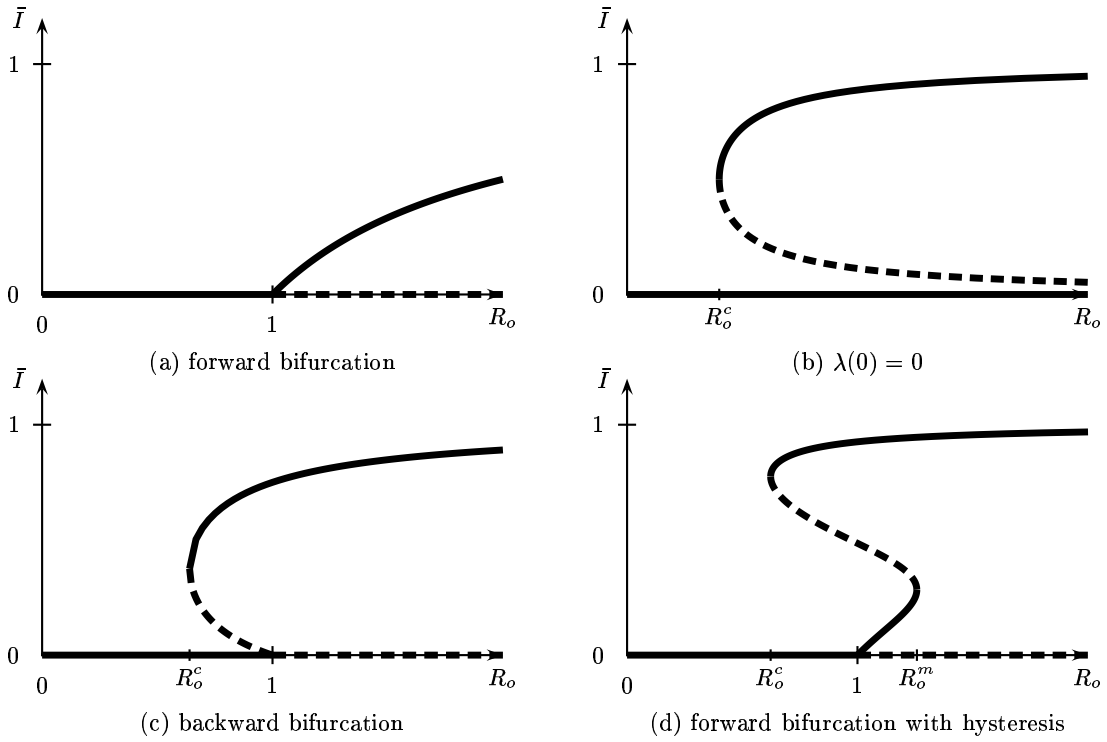
1. Roy M. Anderson and Robert M. May. *Infectious Diseases of Humans.* Oxford University Press, Oxford, 1991.
2. Fred Brauer. Perturbations of the nonlinear renewal equation. *Advances in Mathematics*, 22(1):32–51, 1976.
3. Kenneth L. Cooke and James A. Yorke. Some equations modelling growth processes and gonorrhea epidemics. *Mathematical Biosciences*, 16:75–101, 1973.
4. Jonathan Dushoff. Incorporating immunological ideas in epidemiological models. *Journal of Theoretical Biology*, 180:181–187, 1996.
5. Jonathan Dushoff, Wenzhang Huang, and Carlos Castillo-Chavez. Backwards bifurcations and catastrophe in simple models of fatal diseases. *Journal of Mathematical Biology*, 36:227–248, 1998.
6. Willy Feller. On the integral equation of renewal theory. *The Annals of Mathematical Statistics*, 12:243–267, 1941.
7. K. P. Hadeler and Carlos Castillo-Chavez. A core group model for disease transmission. *Mathematical Biosciences*, 128:41–55, 1995.
8. K. P. Hadeler and P. van den Driessche. Backward bifurcation in epidemic control. *Mathematical Biosciences*, 146:15–35, 1997.
9. Herbert W. Hethcote, Mark A. Lewis, and P. van den Driessche. An epidemiological model with a delay and a nonlinear incidence rate. *Journal of Mathematical Biology*, 27:49–64, 1989.
10. Herbert W. Hethcote and James W. van Ark. Epidemiological models for heterogeneous populations: proportionate mixing, parameter estimation, and immunization programs. *Mathematical Biosciences*, 84:85–118, 1987.
11. Herbert W. Hethcote and P. van den Driessche. Some epidemiological models with nonlinear incidence. *Journal of Mathematical Biology*, 29:271–287, 1991.
12. Herbert W. Hethcote and P. van den Driessche. An SIS epidemic model with variable population size and a delay. *Journal of Mathematical Biology*, 34:177–194, 1995.
13. Herbert W. Hethcote and James A. Yorke. *Gonorrhea Transmission Dynamics and Control*, volume 56 of *Lecture Notes in Biomathematics.* Springer-Verlag, Berlin, 1984.

14. Wenzhang Huang, Kenneth L. Cooke, and Carlos Castillo-Chavez. Stability and bifurcation for a multiple-group model for the dynamics of HIV/AIDS transmission. *SIAM Journal of Applied Math*, 52(3):835–854, 1992.
15. Christopher M. Kribs-Zaleta and Jorge X. Velasco-Hernández. A simple vaccination model with multiple endemic states. *Mathematical Bioscience*, 164(2):183–201, 2000.
16. Wei-min Liu, Herbert W. Hethcote, and Simon A. Levin. Dynamical behavior of epidemiological models with nonlinear incidence rates. *Journal of Mathematical Biology*, 25(4):359–380, 1987.
17. Wei-min Liu, Simon A. Levin, and Yoh Iwasa. Influence of nonlinear incidence rates upon the behavior of SIRS epidemiological models. *Journal of Mathematical Biology*, 23(2):187–204, 1986.
18. Stig-Olof Londen. On the asymptotic behaviour of the bounded solutions of a non-linear Volterra equation. *SIAM Journal on Mathematical Analysis*, 5(6):849–875, 1974.
19. Richard K. Miller. On the linearization of Volterra integral equations. *Journal of Mathematical Analysis and Applications*, 23:198–208, 1968.
20. Richard K. Miller. *Nonlinear Volterra Integral Equations*. Mathematics Lecture Note Series. W. A. Benjamin, Inc., Philippines, 1971.



(a) $f(I) = 1 - I$, $X = \{0\}$ (b) $f(I) = I(1 - I)$, $X = \{0, 0.5\}$ (c) $f(I) = (1 + 5I)(1 - I)$, $X = \{0, 0.4\}$

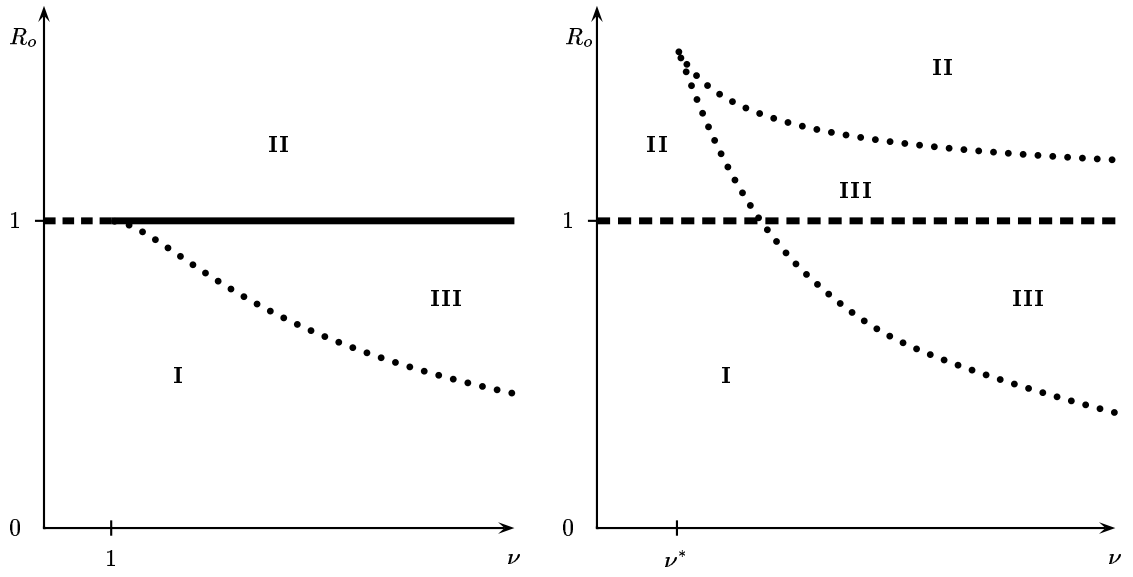
Fig. 1 Example plots of the function $f(I)$. (a) Contact rate of Equation (21) with $\nu = 0$ (constant contact rate). (b) Contact rate of Equation (20) with $\kappa = 1$, $p = 2$, $q = 1$. (c) Contact rate of Equation (21) with $p = 2$ and $\nu = 5$. The values for X given below each figure refer to the set of critical points defined by Equation (10).



(a) forward bifurcation (b) $\lambda(0) = 0$

(c) backward bifurcation (d) forward bifurcation with hysteresis

Fig. 2 Bifurcation diagrams for Equation (6). The dashed lines indicate unstable equilibria and continuous lines indicate locally stable equilibria. (a) Contact rate of Equation (21) with $\nu = 0$ (constant contact rate). (b) Contact rate of Equation (20) with $\kappa = 1$, $p = 2$, $q = 1$. (c) Contact rate of Equation (21) with $p = 2$ and $\nu = 5$. (d) Contact rate of Equation (21) with $p = 5$, $\nu = 17 > \nu^*$.



(a) $p = 2$ (b) $p > 1$, $p \neq 2$

Fig. 3 Bifurcation sets in the (R_o, ν) parameter plane for Equation (6) with contact rates of Equation (21) and (a) $p = 2$, (b) $p = 5$. The horizontal line represents the emergence of the endemic equilibrium at a reproduction number of one. The line is dashed to indicate a forward bifurcation and solid to indicate a backward bifurcation. The dotted curves indicate the saddle node bifurcations. In region I, the disease free equilibrium is globally stable. In region II, the disease free equilibrium is unstable and there is a single globally stable endemic equilibrium. In region III, there are multiple stable equilibria. In (b), the two saddle nodes appear at a cusp bifurcation and decrease to $R_o = 1$ and $R_o = 0$ as $\nu \to \infty$. The diagram is qualitatively similar for all $p > 2$. For $p = 5$, $\nu^* = 125/9$.