Dynamics of an Epidemic Model with Quadratic Treatment

John C. Eckalbar and Walter L. Eckalbar

BSS, California State University, Chico and
School of Life Sciences, Arizona State University

This paper introduces a novel treatment function into an SIR model with bi-linear infection force. Treatment is assumed to increase at a decreasing rate as the sub-population of infected rises. But at some finite number of infected individuals, society's ability to treat the infected reaches a peak and then begins to fall, due perhaps to diminishing supplies or efficiency of health care resources. The system is found to have as many as four equilibria, with possible bi-stability, backward bifurcations, and limit cycles. Particular attention is paid to the effect of variations in the key treatment parameter, $r$. It is found that when $r$ is either low or high, small changes in $r$ do not affect the equilibrium outcome.

DRAFT January 2010

1. Introduction

Epidemiological models have recently begun to explore the role of treatment functions within their dynamic equations. The central idea is that as the number of infectious individuals, $I$, increases, society's resources are mobilized to counter the potential spread of the infection. Hence, a treatment function, $T(I)$ is proposed, where $T(I)$ works to cure infectious individuals and, therefore, to reduce the value of the time derivative $dI/dt$.

Using an SIR model, Wang and Ruan (2004) have examined the case where $T(I)$ equals 0 for $I = 0$ and equals a constant value $r$ for all $I > 0$. In this case treatment is fully "on" at the first appearance of an infectious disease. The effect of the magnitude of parameter $r$ on the behavior of the system is explored. The authors find the model can yield homoclinic orbits even though incidence rates are assumed to be bi linear. And further, from a policy perspective, it is found that under certain parameter assumptions it may not be necessary for policy-makers to set $r$ so high as to eliminate endemic equilibria--instead, such equilibria may be unstable, and society can approach a disease free equilibrium without incurring the expense of maintaining a high value for $r$.

W. Wang (2006) incorporates the following piecewise linear treatment function into an SIR model: $T(I) = \min(rI, rI_0)$, where $I_0$ is the infective level at which the health care system reaches capacity, and $r$ is a positive constant. Hence, treatment rises linearly with $I$ until capacity is reached, after which treatment is constant. Bi linear incidence is assumed, and by construction, the dynamics of $I$ and $S$ are independent of $R$ and $N$, where $S$ and $R$ are the numbers of susceptible and recovered individuals, and $N = S + I + R$ is the total population. The model is found to have bistable endemic equilibria when $I_0$ is low. The author also finds backward bifurcations.

Hu, S. Liu, and H. Wang (2008), X. Li, W. Li, and Ghosh (2009), and Zhang and X. Liu (2009) employ the same treatment function as W. Wang (2004), but apply it in different settings. Zang and X. Liu use an SIS model. X. Li, W. Li, and Ghosh deal with an SIR model with saturated incidence rate. While Hu, S. Liu, and H. Wang model all three dimensions of an SIR system. Backward bifurcations and bistable endemic equilibria are found by all under appropriate parameter values. And all three are able to discuss the effect of capacity on the nature of the equilibria.
Zhang and Liu (2008) use the saturated treatment function: \( T(I) = rl/(1 + \alpha I) \) for \( r > 0, \alpha > 0 \). The authors argue that the saturated function has the advantage of giving near-linear treatment response when \( I \) is low, the function approaches a capacity limit as \( I \) gets large, and it has the convenience of being continuous for \( I \geq 0 \). Using an SIR model with saturated incidence rate, the authors find sufficient conditions for global stability of endemic and disease free equilibria, and they discuss the equilibrium effects of the magnitudes of the parameters in the treatment function. As with the other papers cited, it is found that the basic reproduction number does not govern the stability of the disease-free equilibrium.

In the present paper, we propose an alternative treatment function. We then embed the treatment function in an SIR model with bilinear infection force and explore the existence and stability of the various equilibria. The treatment function is investigated in the next section. Section 3 specifies the model, and Section 4 examines the conditions for for the existence of real positive equilibria. Section 5 looks at local stability conditions, and Section 6 explores global stability issues. Section 7 explores global issues. Section 7 displays a wide range of numerical simulations. We close with a discussion of the role or the treatment function.

2. The Treatment Function

It is no doubt true that all but the poorest societies will mobilize their health sectors to fight a perceived infection. A fully realistic treatment function would be hopelessly complex. Consider even the barest outline of how one would start to build such a function. Treatment would be a function of the availability and use of many thousands of types of physical objects--syringes, medications, gloves, masks, imaging equipment, etc. All of these items would be available in finite quantities at the outbreak of an infectious threat and would only be augment able with the passage of time. Furthermore, inanimate objects obviously don't provide effective treatment, only objects in the hands of trained health care workers can do that. Indeed, the objects and the healthcare workers are highly complementary inputs in the economic sense, meaning that neither objects alone, not workers alone can yield effective results.

As an epidemic develops, it may be that there is enough excess capacity in the system to shift resources in such a way as to increase effective treatment in a near-linear fashion to match the level of infectious cases. But as \( I \) increases, shortages of equipment or personnel will begin to develop, and \( T(I) \) will grow at a decreasing rate. At some value for \( I \), \( T(I) \) will attain a maximum and begin to decline with further increases in \( I \). Perhaps individuals begin hoarding critical supplies. Or perhaps the number of effective healthcare workers drops as some fall victim to the infection. This was certainly the case in the influenza pandemic of 1918. Or treatment centers could fill up, leaving a higher portion of the infected untreated and out in society where they could spread the disease more easily.

In the interest of capturing this possibility, we propose a treatment function \( T(I) \) with the property that as \( I \) rises from zero, \( T(I) \) rises to a maximum value at \( I = I_m \) and then falls toward zero, so \( T'(I) > 0 \) for \( I < I_m \) and \( T'(I) < 0 \) for \( I > I_m \). Further, in order to enforce the diminishing marginal increment to effective treatment as \( I \) increases, we propose that \( T''(I) < 0 \).

Consider a second-order Taylor approximation to such a treatment function in the form:

\[
T(I) = rl - gl^2, \quad r, g > 0.
\]

Care must be taken with this function in that the second term can cause \( T(I) < 0 \), i.e., "treatment" could do more harm than good. Though this is not impossible, a priori, we feel it would be better to leave that possibility aside. To that end, define our treatment function as follows:

\[
T(I) = \max[rl - gl^2, 0].
\]
Though we will not formally introduce any restrictions on $g$ other than that it is positive, we should think of $g$ as being "small," perhaps on the order of $r/N$, where $N$ is the population size. If $g$ were precisely $r/N$, we would have: \( T(I) = r \left( I - \frac{l}{N} \right) \), which yields $T(I) = 0$ for $I = 0$ and for $I = N$, and gives the maximum value for $T$ when $I = N/2$, at which point $T = rN/4$. Hence, $r$ could be seen as the parameter which "tunes" the societal effort to fight the infection. For interpretative purposes and for bifurcation analysis, we will in fact focus on $r$, thinking of it as a metric for the societal effort to fight the infection.

### 3. The Model

Our model is as follows:

\[
\begin{align*}
\frac{dS}{dt} &= A - dS - \lambda IS \\
\frac{dI}{dt} &= \lambda IS - (d + \gamma + \epsilon)I - \max\{rI - gI^2, 0\} \\
\frac{dR}{dt} &= \gamma I + \max\{rI - gI^2, 0\} - dR \\
\frac{dN}{dt} &= A - dN - \epsilon I,
\end{align*}
\]

where $A$ is the recruitment rate; $d$ is the natural death rate; $\lambda$ is the infection force; $\gamma$ is the recovery rate for the infectious class; $\epsilon$ is the disease-related death rate; and $r$ and $g$ are as previously discussed. All parameters are assumed to be positive. If we define $\Omega = \{ (S, I, R, N) \mid S \geq 0, I \geq 0, R \geq 0, N \geq 0 \}$, it is obvious from inspection of the above system that if the initial values $S(0), I(0), R(0),$ and $N(0)$ are in $\Omega$ then $S(t), I(t), R(t),$ and $N(t)$ will be in $\Omega$ at all $t > 0$.

As written, the system has four dynamic variables, $S, I, R,$ and $N,$ but it is decomposable in the sense that, $dS/dt$ and $dI/dt$ depend only upon only upon $S$ and $I$. This is due primarily to assumptions on the role of parameter $A$. Specifically, parameter $A$ is a constant inflow of newborns into the population, and all newborns are assumed to arrive into the susceptible class.

We will refer to a point $(S, I, R, N)$ as an "equilibrium" if $dS/dt = dI/dt = 0,$ even if $dR/dt$ and $dN/dt$ are non-zero. It is obvious from inspection that if $S$ and $I$ are fixed at an equilibrium, then as $t \to \infty$ $N \to (A - \epsilon I)/d$ and $R \to (\gamma I + \max\{rI - gI^2, 0\})/d$, which might be called a "full equilibrium." Consequently, we will concentrate on the system:

\[
\begin{align*}
\frac{dS}{dt} &= A - dS - \lambda IS \\
\frac{dI}{dt} &= \lambda IS - (d + \gamma + \epsilon)I - \max\{rI - gI^2, 0\}
\end{align*}
\]

(System 3.1) can be decomposed into two sub-systems according to whether $\max\{rI - gI^2, 0\} = 0$ or $rI - gI^2$. To this end, we define two domains: $D_{31} = \{(S, I) \mid rI - gI^2 \geq 0, \ S, I \geq 0 \}$ and $D_{32} = \{(S, I) \mid rI - gI^2 < 0, \ S, I \geq 0 \}$, and we explore the two associated sub-systems:

\[
\begin{align*}
\frac{dS}{dt} &= A - dS - \lambda IS \\
\frac{dI}{dt} &= \lambda IS - (d + \gamma + \epsilon)I - (rI - gI^2)
\end{align*}
\]

and
The Epidemic model paper F January 2010.nb

\[
\begin{align*}
\frac{dS}{dt} &= A - dS - \lambda IS \\
\frac{dI}{dt} &= \lambda IS - (d + \gamma + \epsilon) I \\
(S, I) &\in D_{32}
\end{align*}
\]

**Comment 1.** Note that (3.0) gives rise to a continuously differentiable vector field in \( \Omega \). This is obvious from the fact that (3.1) is continuous in \( D_{31} \), (3.2) is continuous in \( D_{32} \), and the two subsystems have the same direction and magnitude of flow on the border between them, i.e., where \( I = r/g \).

Dynamic systems of this sort are often called "hybrid" systems. (See Branicky (1997).)

### 4. Equilibria

System (3.0) admits four possible equilibrium vectors, \( E_0, E_1, E_2, \) and \( E_3 \), where each vector \( E_i, i = 0, 1, 2, 3 \) is given by the equilibrium values of \( (S_i, I_i) \). The possible equilibria are:

\[
E_0 = (S_0, I_0) = \left( \frac{\lambda}{d}, 0 \right)
\]

\[
E_1 = (S_1, I_1) = \left( \frac{d g + d \lambda + r \lambda + \gamma \lambda + \epsilon \lambda + \sqrt{-A g \lambda^2 + ((r+\gamma+\epsilon) \lambda + d (g+\lambda))^2}}{2 \lambda}, -\frac{d g + d \lambda + r \lambda + \gamma \lambda + \epsilon \lambda - \sqrt{-A g \lambda^2 + ((r+\gamma+\epsilon) \lambda + d (g+\lambda))^2}}{2 \lambda} \right)
\]

\[
E_2 = (S_2, I_2) = \left( \frac{d g + d \lambda + r \lambda + \gamma \lambda + \epsilon \lambda - \sqrt{-A g \lambda^2 + ((r+\gamma+\epsilon) \lambda + d (g+\lambda))^2}}{2 \lambda}, -\frac{d g + d \lambda + r \lambda + \gamma \lambda + \epsilon \lambda + \sqrt{-A g \lambda^2 + ((r+\gamma+\epsilon) \lambda + d (g+\lambda))^2}}{2 \lambda} \right)
\]

\[
E_3 = (S_3, I_3) = \left( \frac{d + \gamma + \epsilon}{\lambda}, \frac{A \lambda - d^2 - d (\gamma + \epsilon)}{(d + \gamma + \epsilon) \lambda} \right)
\]

\( E_0 \) is the disease-free equilibrium, and \( E_1, E_2, \) and \( E_3 \) are the possible endemic equilibria. Since the equations above can yield solution values outside their stipulated domains, we must be careful to consider \( E_0, E_1, \) and \( E_2 \) only if they exist in \( D_{31} \) and to consider \( E_3 \) only if it exists in \( D_{32} \). A real and weakly positive \( E_0 \) will always exist in \( D_{31} \), but the number of real and positive endemic equilibria depends upon the parameters chosen. Sometimes all three \( E_1, E_2, \) and \( E_3 \) will exist in real and positive forms, and sometimes either none, one, or two of them will exist. Let us explore the possibilities.

As an aid to intuition, we construct isoclines, iso-S and iso-I, for (3.0). These are defined as iso-S = \{ \( (S, I) \mid dS/dt = 0 \) \} and iso-I = \{ \( (S, I) \mid dI/dt = 0 \) \}. Equilibria are found at points where iso-S meets iso-I, and at non-equilibrium points, all paths through iso-S cross vertically, and all paths through iso-I cross horizontally. Figure 1 depicts these curves for a case where the assumed parameters yield four equilibria. Note that iso-I is in three parts--the vertical part running through \( D_{32} \), the negatively sloped part passing through \( E_2 \) and \( E_3 \), and the horizontal part coextensive with the S-axis.
Figure 1. Iso-S and iso-I curves showing four equilibria with: \[^{[A, \lambda, d, \gamma, \epsilon, r, g]} = \{460, .00125, .2, .312, .002, 1.35, .0025\}. \] \(I = r/g\) divides the domain into \(D_{32}\) and \(D_{31}\).

Figure 1 makes it easy to get a general idea of how parameter changes affect equilibria. For example, starting with the parameter values used in the figure, we see that if \(r\) falls, the down-sloping leg of iso-I falls and the upper border of \(D_{31}\) drops. This will not change \(E_3\), but it may simultaneously eliminate \(E_1\) and \(E_2\). On the other hand, an increase in \(r\) will raise both the down-sloping leg of iso-I and the border between \(D_{31}\) and \(D_{32}\). This can obviously eliminate \(E_3\). (The \(r\) value which eliminates \(E_3\) will be noted \(r_2\).) Also, when \(r\) rises, \(E_2\) rises with it. When \(r\) reaches \(r_2\), \(E_3 = E_2\), and further increases in \(r\) annihilate both equilibria simultaneously.

Since our focus here is on the role of parameter \(r\), we will leave aside further manipulations of Figure 1 and proceed to a bifurcation diagram showing equilibrium \(I\) values as a function of \(r\). We will explore the conditions under which real and positive values will exist for each of the \(E_i\). Again, care must then be take to verify that each \(E_i\) exists in its appropriate domain.

Figure 2 may help clarify the following discussion. The heavy lines in the figure denote equilibrium values of \(I_1\) when these values are in their required domains. The dashed ray from the origin divides the space into \(E_{32}\) above and \(E_{31}\) below. Various critical \(r\) values, to be discussed shortly, are also depicted.
Figure 2. Bifurcation diagram using \( \{A, \lambda, d, \gamma, \epsilon, g\} = \{460, .00125, .2, .312, .002, .0025\} \).

The primary existence result is given below. Recall throughout that \( S = A/(d + \lambda I) \) at any equilibrium, so if we verify that \( I \) is real and positive, it follows that \( S \) is also.

**Theorem 4.1.** Assuming \( A, \lambda, d, \gamma, \epsilon, r, g > 0 \), real and positive values for the \( E_i \) will exist in their appropriate domains under the following conditions:

1. \( E_0 \) is real and weakly positive, and \( E_0 \) will lie in \( D_{31} \) at all parameter values.
2. Equilibrium \( E_1 \) will be real and positive, if and only if
   
   \[
   (2a) \quad \frac{\sqrt{A}}{\sqrt{g}} - \frac{d}{\lambda} > 0 \quad \text{and} \quad \frac{-d g - d \lambda + \sqrt{A g} - \gamma \lambda - \epsilon \lambda}{\lambda} = r_0 \quad \text{and} \quad \frac{-d^2 d \gamma - d \epsilon + A \lambda}{d} = r_1.
   \]

   (Note: \( r_1 - r_0 = \left(d \sqrt{g} - \sqrt{A \lambda}\right)^2/d \lambda, \) so \( r_1 \geq r_0 \).)

3. Assuming (2a) and (2b) hold, real and positive \( E_1 \) will exist in \( D_{31} \) if any of the following hold:
   
   \( i \) \( r_0 > 0 \) and at \( r_0, E_1 \in D_{31} \) (as in Figure 2), or
(ii) \( r_0 > 0 \) and at \( r_0, E_1 \in D_{32}, \) but \( r \) is restricted to \( r > r_2 = \frac{g(-d^2 - d \gamma - d \epsilon + A \lambda)}{(d+\gamma+\epsilon)\lambda}, \) or

(iii) \( r_0 < 0 \) but \( I_1 \) is positive at \( r = 0 \) (as in Figure 3) and \( r \) is restricted to \( r > r_2. \)

(3) Equilibria \( E_2 \) will be real, positive, and in \( D_{31} \) if and only if \( r > r_0 \) and

(3a) \( r_0 \geq 0, \) and

(i) if \( I_2 \) at \( r_0 \) is positive and less than \( r_0/g. \) And \( r \in [r_0, r_2 = \frac{g(-d^2 - d \gamma - d \epsilon + A \lambda)}{(d+\gamma+\epsilon)\lambda}], \) or

(ii) if \( I_2 \) at \( r_0 \) is negative and \( r > r_1. \) Or

(3b) \( r_0 < 0, \) \( I_2 \) at \( r_0 \) is negative and \( r > r_1. \)

(4) Real and positive values for \( E_3 \) will only obtain when \( r_2 > 0 \) and \( r < r_2. \)

**Proof.**

(1) Since \( E_0 = (A/d, 0), \) it is obviously real and weakly positive, and since \( I_0 = 0 < r/g, E_0 \in D_{31} \) for all acceptable parameter values.

(2) Simple inspection of the formula for the equilibrium \( E_1 \) reveals that \( E_1 \) will be real if and only if

\[-4 A g \lambda^2 + ((r + \gamma + \epsilon) \lambda + d (g + \lambda))^2 \geq 0, \]

which implies that \( r \geq -\frac{-d \times 8 - d \lambda + 2 \sqrt{A g \lambda - \gamma \lambda - \epsilon \lambda}}{\lambda} = r_0. \) And \( E_1 \) will be positive if \( I_1 \) is positive, which requires two conditions:

(2a) Since \( I_1 \) declines with increases in \( r, I_1 \) must be positive at \( r_0, \) where \( I_1 \) has its highest value. (We know that \( dI_1/dr < 0, \) since \( dI_1/dr = \frac{J}{28 \left( 1 - \frac{(r+y+\epsilon)\lambda + d (g + \lambda))^2}{\sqrt{-4 A g \lambda^2 + ((r+y+\epsilon)\lambda + d (g + \lambda))^2}} \right) \). And since the value of \( I_1 \) at \( r_0 \) is

\[ \frac{\sqrt{A}}{\sqrt{g}} - \frac{d}{\lambda}, \]

to have a positive \( E_1 \)-type equilibrium, \( \frac{\sqrt{A}}{\sqrt{g}} - \frac{d}{\lambda} \) must be positive. And

(2b) \( r \) must not be too large. Direct solution of \( I_1 = 0 \) for \( r, \) reveals that \( I_1 \) is positive if

\[ r < \frac{-d^2 - d \gamma - d \epsilon + A \lambda}{d} = r_1. \]

(2c) Assuming throughout that (2a) and (2b) hold,

(i) If \( r_0 > 0 \) and at \( r_0, E_1 \in D_{31}, \) then since \( dI_1/dr < 0, I_1 \) will remain in \( D_{31} \) until an increase in \( r \) forces \( I_1 = 0, \) which will happen when \( r = r_1. \)

(ii) If \( r_0 > 0 \) and at \( r_0, E_1 \in D_{32}, \) then \( I_1 \) will cross the border from \( D_{32} \) into \( D_{31} \) when \( r \) increases to \( r_2, \) which is easily checked by solving \( I_1 = r/g \) for \( r. \)

(iii) If \( r_0 < 0 \) but \( I_1 \) is positive at \( r = 0, \) then \( I_1 \) drops in to \( D_{31} \) at \( r = r_2. \) Note that \( I_1 \) is positive at \( r = 0 \) if and only if

\[ d \lambda + \gamma \lambda + \epsilon \lambda - \sqrt{-4 A g \lambda^2 + ((\gamma + \epsilon) \lambda + d (g + \lambda))^2} > d g. \]

Figure 2 is offered as an illustration.
Figure 3. Bifurcation diagram using \( \{A, \lambda, d, \gamma, \epsilon, g\} = \{3300, .002, 1.26, 2.5, .001, .0015\} \)

(3) Having a real-valued \( E_2 \) requires
\[
-4Ag \lambda^2 + ((r + \gamma + \epsilon) \lambda + d(g + \lambda))^2 \geq 0,
\]
which is equivalent to \( r \geq r_0 \).

And the value of \( I_2 \) at \( r_0 \) will be \( \frac{\sqrt{A}}{\sqrt{g}} - \frac{d}{\lambda} \). Checking for positive \( I_2 \), we find several cases to consider:

(3a) If \( r_0 \geq 0 \), and

(i) \( I_2 \) at \( r_0 \) is above the line given by \( I = r/g \), i.e., \( I_2 \) is outside its proper domain \( D_{31} \) when \( r = r_0 \), then \( I_2 \) will also be outside \( D_{31} \) at all \( r > r_0 \). (This follows from the fact that)
\[
\frac{dI_2}{dr} = \frac{1}{2g} \left( I + \frac{(r+\gamma+\epsilon) \lambda + d(g + \lambda)}{\sqrt{-4Ag \lambda^2 + ((r+\gamma+\epsilon) \lambda + d(g + \lambda))^2}} \right) > 0,
\]
which declines monotonically to \( I/g \) as \( r \to \infty \). Thus, \( I_2 \) never falls from \( E_{32} \) across the border into \( E_{31} \).) In this case, an \( E_2 \) -type equilibrium will not exist in \( D_{31} \). (See the Appendix Proof of 3a(i).) But if \( I_2 \) at \( r_0 \) is positive and below \( I = r/g \), then \( E_2 \) will be in \( D_{31} \) at \( r = r_0 \). And as \( r \) rises, \( I_2 \) will remain in \( D_{31} \) until it crosses \( I = r/g \) and leaves the proper domain when \( r = r_2 = \frac{g(-d^2 - d \gamma - d \epsilon + A \lambda)}{(d+\gamma+\epsilon) \lambda} \). See Figure 2.
(ii) $I_2$ at $r0$ is below zero, then $E_2$ is initially outside its proper domain $D_{31}$, but at some $r > r0$, $I_2$ will become positive and $E_2$ will enter its proper domain. Solving $I_2 = 0$, we find that value of $r$ is the same $rI$ described earlier. Hence, $rI$ gives the value at which $I_1$ drops down out of $D_{31}$, or, if applicable, where $I_2$ rises into $D_{31}$ from below. See Figures 1 and 4. Obviously, for any given parameter set, only one of the above might happen.

(3b) If $r0 < 0$, then a real and positive $E_2$ requires that the $I_2$ line crosses the $I$ axis below zero, i.e., $I_2(r = 0) < 0$. So, using the equilibrium value for $I_2$, we must in this case have

$$d \lambda + \gamma \lambda + \epsilon \lambda + \sqrt{-4 A g \lambda^2 + ((\gamma + \epsilon) \lambda + d (g + \lambda))^2} < d g.$$ 

At $r0$, $I_2$ will be below $E_{31}$, but it will rise into $E_{31}$ when $r = rI$. (This is revealed by direct solution of $I_2 = 0$ for $r$.)

(4) Finally, the worst case equilibrium is $E_3 = (S_3, I_3) = \left( \frac{d + \gamma + \epsilon}{\lambda}, \frac{A \lambda - d^2 - d (\gamma + \epsilon) \lambda}{(d + \gamma + \epsilon) \lambda} \right)$, which is obviously real. Positive values for $I_3$ will only obtain when $r2 > 0$, since $I_3 = r2/g$. It follows that $I_3$ will only be in $D_{32}$ when $r < r2$, since being in $D_{32}$ requires $I_3 > r/g$. So $E_3$ drops out of existence altogether when $r$ rises above $r2$. □

Note that 2c(iii) and 3b indicate that when $r0 < 0$ we can have a real and positive $E_1$ or a real and positive $E_2$, but we cannot have both, and we will have neither if

$$d \lambda + \gamma \lambda + \epsilon \lambda + \sqrt{-4 A g \lambda^2 + ((\gamma + \epsilon) \lambda + d (g + \lambda))^2} > d g$$

$$d \lambda + \gamma \lambda + \epsilon \lambda - \sqrt{-4 A g \lambda^2 + ((\gamma + \epsilon) \lambda + d (g + \lambda))^2}.$$

Figure 4. $\{A, \lambda, d, \gamma, \epsilon, g, r\} = \{1500, .002, 1, .5, .001, .008, 2.15\}$
Also notice that both $E_2$ and $E_3$ drop out of existence when $r$ rises above $r_2$. What is happening is that each solution is crossing out of its domain as $r$ crosses from below to above $r_2$.

5. Local Stability

The stability examination begins with the Jacobian matrix for sub-system 3.1:

$$J_{31}(S, I) = \begin{pmatrix} -d - I \lambda & -S \lambda \\ I \lambda & -d + 2g I - r - \gamma - \epsilon + S \lambda \end{pmatrix}$$

Local stability of an equilibrium point in $D_{31}$, together with whether the point is a node, saddle, focus, etc., depend upon the eigenvalues $J_{31}$ evaluated at the equilibrium. When the eigenvalues prove to be too cumbersome for interpretation, stability information can also be derived from the trace and determinant of $J_{31}$, evaluated at the point of equilibrium. These will be written $\text{Tr}(J_{31}(S, I))$ and $\text{Det}(J_{31}(S, I))$, respectively. Local stability requires $\text{Tr}(J_{31}(S, I)) < 0$ and $\text{Det}(J_{31}(S, I)) > 0$. In general,

$$\text{Tr}(J_{31}) = -2d + 2gI - r - \gamma - \epsilon - I\lambda + S\lambda$$

and

$$\text{Det}(J_{31}) = d^2 + d\gamma + d\epsilon - 2gI^2\lambda - dS\lambda + I(-2dgr + d\lambda + r\lambda + \gamma\lambda + \epsilon\lambda).$$

We investigate the equilibria in turn:

**Theorem 5.1.** The disease-free equilibrium, $E_0 = (S, I) = (A/\lambda, 0)$, is a stable node if and only if $r > r_1 = (A\lambda)/d - d - \gamma - \epsilon$, and it is a saddle otherwise.

**Proof:** The eigenvalues of $J_{31}(E_0)$ are $-d$ and $-d - r - \gamma - \epsilon + \frac{A\lambda}{d}$, which are both clearly real and negative iff $r > r_1 = (A\lambda)/d - d - \gamma - \epsilon$. And if $r < r_1$, the eigenvalues are of opposite sign, and $E_0$ is a saddle. □

If the basic reproduction number, $R_0$, is defined such that $E_0$ is locally asymptotically stable when $R_0 < 1$, then $R_0 = \frac{A\lambda}{d(d+\gamma+r+\epsilon)}$. As others (Zhang and Liu, 2009) have found, $R_0$ is not particularly helpful in this class of models, and we will seldom refer to it here.

**Theorem 5.2.** If $r1 > r3 > r0$, $E_1$ will be locally asymptotically stable when $r \in (r3, r1]$ and unstable when $r \in [r0, r3)$.

**Proof.** Evaluating the trace and determinant at $E_1$, we have

$$\text{Det}(J_{31}(E_1)) = d^2 + d\gamma + d\epsilon - 2gI_1^2\lambda - dS_1\lambda + I_1(-2dgr + d\lambda + r\lambda + \gamma\lambda + \epsilon\lambda) = \sqrt{-4Ag\lambda^2 + ((r + \gamma + \epsilon)\lambda + d(g + \lambda))^2} \frac{2g\lambda}{d(g + \lambda)} I_1 \sqrt{-4Ag\lambda^2 + ((r + \gamma + \epsilon)\lambda + d(g + \lambda))^2}. $$

Since $\sqrt{-4Ag\lambda^2 + ((r + \gamma + \epsilon)\lambda + d(g + \lambda))^2}$ is real and positive when $r > r0$, $\text{Det}(J_{31}(E_1)) > 0$ when $r > r0$ and $I_1 > 0$. Equivalently, the determinant is real and positive when $r1 > r > r0$. It follows that $E_1$ will never be a saddle point.

Analysis of the trace at $E_1$ is more complex. In general, the trace is given by:
\[
\text{Tr}(J_{31}(E_1)) = -\frac{1}{2g\lambda} \left[ d g^2 - g r \lambda - g \gamma \lambda - g \epsilon \lambda + d \lambda^2 + r \lambda^2 + \gamma \lambda^2 + \epsilon \lambda^2 + (g - \lambda) \sqrt{-4 A g \lambda^2 + ((r + \gamma + \epsilon) \lambda + d (g + \lambda))^2} \right]
\]

and, assuming a positive \( E_1 \) exists, then at the right-hand edge of the domain of \( E_1 \), i.e., where \( r = r_1 \), direct substitution yields \( \text{Tr}(J_{31}(E_1(r))) = -d < 0 \). It can be shown that the derivative of \( \text{Tr}(J_{31}(E_1(r))) \) with respect to \( r \) is

\[
\frac{(\lambda - g) \lambda \left[ d g + d + \lambda + r \lambda + \gamma \lambda + \lambda + \sqrt{-4 A g \lambda^2 + ((r + \gamma + \epsilon) \lambda + d (g + \lambda))^2} \right]}{2 g \lambda - \lambda (I_1 + 2 d g)} = \frac{(\lambda - g) \lambda (I_1 + 2 d g)}{\sqrt{-4 A g \lambda^2 + ((r + \gamma + \epsilon) \lambda + d (g + \lambda))^2}}.
\]

Thus, the sign of the slope of the trace with respect to \( r \) is the same as the sign of \( \lambda - g \). So if \( \lambda - g > 0 \), \( \text{Tr}(J_{31}(E_1(r))) < -d < 0 \) for all \( r < r_1 \), and all \( E_1(r) \) will be locally stable for \( r_0 < r < r_1 \). But if \( \lambda - g < 0 \), then \( \text{Tr}(J_1(E_1(r))) \) is above - \( d \) for \( r \) below \( r_1 \). At some \( r < r_1 \), the trace may become positive. We will define this point as \( r_3 \), where

\[
r_3 = \frac{-d g (\gamma + \epsilon) (g - \lambda) + d^2 g \lambda + A (g - \lambda)^2 \lambda}{d (g - \lambda)}.
\]

If \( r_3 < r_0 \), there is no effect on stability of \( E_1 \), since \( E_1 \) is not defined for \( r < r_0 \). But if \( r_1 > r_3 > r_0 \), \( E_1 \) will be locally stable when \( r \in (r_3, r_1) \) and unstable when \( r \in [r_0, r_3) \).  

**Theorem 5.3.** If a real and positive \( E_2 \) exists, then \( E_2 \) is a saddle.

**Proof.** By direct substitution of \( E_2 \) into \( J_{31} \), we get

\[
\text{Det}(J_{31}(E_2)) = -\sqrt{-4 A g \lambda^2 + ((r + \gamma + \epsilon) \lambda + d (g + \lambda))^2} \left(-d g + d \lambda + \lambda + \sqrt{-4 A g \lambda^2 + ((r + \gamma + \epsilon) \lambda + d (g + \lambda))^2} \right)
\]

\[
= \frac{-I_2 \sqrt{-4 A g \lambda^2 + ((r + \gamma + \epsilon) \lambda + d (g + \lambda))^2}}{2 g \lambda}.
\]

\( I_2 \) is real if and only if \(-4 A g \lambda^2 + ((r + \gamma + \epsilon) \lambda + d (g + \lambda))^2 \geq 0 \), in which case \( \text{Det}(J_{31}(E_2)) \) is also real. And \( \text{Det}(J_{31}(E_2)) < 0 \) whenever \( I_2 > 0 \). Hence, \( E_2 \) is a saddle whenever \( E_2 \) is real and positive.  

Furthermore, when \( \frac{\sqrt{-A g}}{\sqrt{g}} - \frac{d}{\lambda} > 0 \), \( \text{Det}(J_{31}(E_2)) < 0 \) if \( r > r_0 \) and when \( \frac{\sqrt{-A g}}{\sqrt{g}} - \frac{d}{\lambda} < 0 \), \( \text{Det}(J_{31}(E_2)) < 0 \) if \( r > r_1 \). See Figures 3 and 4.

**Theorem 5.4.** If a real and positive \( E_3 \) equilibrium exists in \( D_{32} \), then that equilibrium is locally asymptotically stable at all possible positive parameter values.

**Proof.** In this case, the Jacobian can be written:

\[
J_{32}(S_3, I_3) = \begin{pmatrix}
-d - I_3 \lambda & -S_3 \lambda \\
I_3 \lambda & -d - \gamma - \epsilon + S_3 \lambda
\end{pmatrix}
\]

So we have,

\[
\text{Tr}[J_{32}(E_3)] = -d - \lambda I_3 < 0 \quad \text{and} \quad \text{Det}[J_{32}(E_3)] = \lambda^2 I_3 S_3 > 0.
\]

Hence, if a real and positive \( E_3 \) exists, it is locally asymptotically stable.  

**6. Global Issues**

We define \( \Sigma(t) = S(t) + I(t) \) and assert that \( \Sigma(t) \) is falling everywhere outside the triangle \( \Phi = \{(S, I)| S \geq 0, I \geq 0, S + I \leq A/d\} \). This will establish that the global flow into \( \Phi \). See Figure 5.
Theorem 6.1. If $\Sigma(t) = A1/d$, where $A1 \geq A$, then $d\Sigma(t)/dt \leq 0$ under (3.0).

Proof: Under (3.0)

$$\frac{dS}{dt} + \frac{dI}{dt} = A - dS - \lambda IS + \lambda IS - (d + \gamma + \epsilon)I - \max[(rI - gI^2), 0].$$

Canceling terms and using $S = A1/d - I$, we can write

$$\frac{d\Sigma(t)}{dt} = \frac{dS}{dt} + \frac{dI}{dt} = A - A1 - (\gamma + \lambda)I - \max[(rI - gI^2), 0] \leq 0,$$

since $A \leq A1, I \geq 0$, and $\max[., .] \geq 0$. $\square$

Figure 5.

Several comments on the implications of Theorem 6.1: First, any initial point $(S(0), I(0))$ in $\Omega\Phi$ has a positive semi-path that leads into $\Phi$. Second, since $d\Sigma(t)/dt < 0$ outside $\Phi$, no equilibrium point can exist in $\Omega\Phi$, or equivalently, all real and positive equilibria to (3.0) will be in $\Phi$. Third, all points (other than $E_0$) on the border of $\Phi$ defined by $S + I = A/d$ flow into $\Phi$. Fourth, since the flow of (3.0) is left to right on the vertical axis, and left to right along the $S$-axis for $S < A/d$, $\Phi$ is positively invariant.

Theorem 6.2. The disease-free equilibrium $E_0$ is globally asymptotically stable if $r > \max(r_1, r_2)$. 
**Proof.** By Theorem 6.1, all points in $\Omega\Phi$ flow into $\Phi$, so we can limit our consideration to the positive semi-paths originating on the border of $\Phi$. Initial points on the S-axis flow horizontally into $E_0$. Points on the I-axis flow into the interior of $\Phi$, as do all points other than $E_0$ on the border $\{(S, I)| I + S = A/d\}$. Since $\Phi$ is positively invariant, any positive semi path entering $\Phi$ will approach either an equilibrium or a limit cycle. (See Hirsch and Smale, 1974, p. 251.) By direct implication of Theorem 4.1, $E_0$ is a unique equilibrium when $r > \max(r_1, r_2)$. Further, closed paths around $E_0$ are impossible due to the fact that trajectories cannot cross $I = 0$. It follows that all positive semi-paths approach $E_0$ under the assumptions of the theorem. □

**Theorem 6.3** There are no limit cycles surrounding $E_3$ if $A g < 4 d^2$:

**Proof.** Since all positive semi-paths originating in $\Omega$ pass into $\Phi$ and all equilibria reside in $\Phi$, we can limit our attention to points in $\Phi$. Recall also that the vector field for (3.0) is $C^1$. Form the Dulac function $D = I(SI)$ and define

$$Y(S, I) = \frac{\partial D(dS/dt)}{\partial S} + \frac{\partial D(dI/dt)}{\partial I}.$$  

We show that $Y(S, I) < 0$ in $\Phi$ under the conditions of the theorem. There are three cases.

(i) In the domain $D_{32}$, we have $Y(S, I) = -\frac{A}{IS^2} < 0$.

(ii) In $D_{31}$ we have $Y(S, I) = -\frac{A + gIS}{IS^2}$, which seems to have the potential of becoming positive if $gIS$ is "large." But note that $I \leq (A/d) - S$ if $(I, S) \in \Phi$. So the maximum possible value of $IS$ in $D_{31}$ will occur when $I = S = \frac{A}{2d}$, provided $I = \frac{A}{2d} \leq r/g$. Assume so for the moment. Then substituting this into $Y(S, I)$, we get:

$$Y(S, I) = \frac{-A + gIS}{IS^2} = \frac{-A + g \frac{x^2}{4d}}{IS^2},$$

which is negative if $A g < 4 d^2$.

(iii) But if $(S, I)$ is outside $D_{31}$ when $I = S = \frac{A}{2d}$ (due to, e.g., a low value for $r$), then $gIS$ is constrained below the value attained in (ii) above, and therefore, we must also have $Y(S, I) < 0$. □

**Theorem 6.4.** When $E_3$ is the only endemic equilibrium and $A g < 4 d^2$, then $E_3$ attracts all points in $\Omega$ other than $E_0$ and its stable manifolds.

**Proof.** As we have just seen in Theorem 6.3, there will be no limit cycle around $E_3$ in this case, and then the method of proof under Theorem 6.2 establishes the present theorem. □

We now employ Poincaré index theory to explore possible combinations of equilibria. See Figure 6.2. (Moghadas is helpful on this topic.)
Theorem 6.5. If $E_0$ is a saddle point, then there are either 1 or 3 additional equilibria in $\Phi$; and if $E_0$ is a stable node, there will be either 0 or 2 additional equilibria in $\Phi$.

**Sketch of Proof.** If we considered the angle of the vector field as we moved counterclockwise along the line segments from $V$ to $W$ to $X$ to $Y$ to $Z$ in Figure 6, we would find a net rotation of $180^\circ$ counterclockwise. Now consider the simple closed curve $G$ (Figure 6) passing through points $Z$ and $V$ and enclosing a small neighborhood of $E_0$ but no other equilibrium point. Let $G$ be the union of the upper portion, $G^+$, where $I > 0$, and the lower portion, $G^-$. Define $H$ to be the piecewise continuous closed loop $H = \{(S, I)| (S, I) \in G^- \cup VW \cup WX \cup XY \cup YZ\}$, where, for example, VW denotes the set of points on the line segment from $V$ to $W$. Theorem 5 establishes that $E_0$ is either a stable node or a saddle point.

(i) If $E_0$ is a saddle point, then the index of the $G$ is $-1$, and if we checked the net change in the angle of the vector field as we moved counterclockwise from $Z$ to $V$ along $G^-$, we would find that it rotates clockwise by $180^\circ$. Hence, if $E_0$ is a saddle point, the net change in the angle of the vector field as we move counterclockwise around $H$ is $0^\circ$, and hence the index the $H$ would be zero. Since the index of the curve $H$ is the sum of the indices of the enclosed equilibrium points, and since a saddle point is presently assumed at $E_0$, then there must be at least one other point with index of +1 (i.e., a spiral, node, or center) which is also inside $H$. There may also be three equilibrium points other than $E_0$, and the sum of the indices of those three points must be +1. We could for example have a node, a spiral, and a saddle in addition to $E_0$, when $E_0$ itself is a saddle.

(ii) The only other possible case, given Theorem 5.1, is that $E_0$ is a stable node with index +1. In this case the rotation of the angle of the vector field from $Z$ to $V$ along $G^-$, is $180^\circ$ counterclockwise. Thus the index of the closed curve $H$ will be +1. Given this, there may be no other equilibria than $E_0$, or there might be two other equilibria with indices of -1 and +1, that is a saddle plus a node, spiral, or center. □

The next theorem pulls together the results of the last two sections and, by combining local stability results with Poincare index information, we are able to derive further results.

**Theorem 6.6.** A non-negative disease-free equilibrium, $E_0$, will always exist for (3.0) and will always be either a stable node or a saddle point.
1. If $E_0$ is a stable node, there are either no other or two other equilibria in $\Phi$ whose indices sum to 0. If there are two equilibria, they will be $E_2$ and $E_3$, which are a saddle point and a stable node or focus, respectively.

2. If $E_0$ is a saddle point, then there are either 1 or 3 other equilibria.

1a. The lone extra equilibrium must be either $E_1$ or $E_3$. In the absence of a limit cycle surrounding $E_1$, either either $E_1$ or $E_3$, whichever exists, will be stable.

2b. If all three other equilibria exist, $E_1$ must have index 1.

**Proof.** The proof that $E_0$ can only be a stable node or saddle point is found in Theorem 6.1.

1. Theorem 6.5 establishes that when $E_0$ is a stable node there can only be 0 or 2 other equilibria in $\Phi$. We now show that if there are 2, they will be $E_2$ and $E_3$. This follows from the fact that stability of $E_0$ requires $r > r_1$, but $E_1$ is not in $\Phi$ if $r > r_1$. (See 2. in Theorem 4.1) The index of the closed curve $H$ in this case is 1. And since the index of $E_0$ is also 1, then the two other equilibria, must have indices of +1 and -1. The index of $E_3$ is +1, so $E_2$ must in this case be a saddle. (See Theorem 5.4.)

2. If $E_0$ is a saddle, the index of $H$ is 0, so the sum of the indices of the other equilibria in $\Phi$ must be +1.

1a. There may be just one other equilibrium point in $\Phi$, in which case it could not be $E_2$, since $E_2$ is a saddle with index -1. Hence, it will be either $E_1$ or $E_3$. Suppose for a moment the other equilibrium is $E_1$. Then either $E_1$ is stable or it is surrounded by a limit cycle, since the up-going unstable manifold of $E_0$ must approach either a limit cycle or a fixed point due to the Poincare-Bendixson Theorem. If the other equilibrium is $E_3$, there will be no limit cycle due to Theorem 6.3.

2b. If there are a total of four equilibria in $\Phi$, two will be saddles and two will be foci or nodes, though again $E_1$ may be surrounded by a limit cycle. □

Numerical simulations also suggest the possible presence of an unstable homoclinic orbit surrounding $E_1$, with the path leaving and returning to $E_2$.

One final remark: If government behavior rendered $r$ a function of $I$, a large amplitude cycle could set up in the following way. Refer to Figure 4 and suppose a new virus is introduced when $r = 0$. The system then moves to $I_3$. If $r$ is then steadily increased to combat the infection, $I$ will remain at $I_3$ until $r$ passes $r_2$, at which time the equilibrium shifts to $I_0$. To conserve resources, $r$ might then be reduced with no initial adverse effect on $I$. But if $r$ is pushed below $r_1$, the equilibrium will shift back to $I_3$. Other scenarios can be visualized if the parameters yield bifurcation diagrams like Figure 2 or Figure 3.

7. Numerical Simulations

With seven parameters and multiple critical values for $r$, there are a great many dynamic possibilities for system (3.0). In the interest of brevity, we will assume the parameters of Figure 2 and explore the evolution of equilibria as $r$ sweeps upward from zero.
Our first case covers a situation of low societal effort to combat infection, *i.e.*, the value of $r$ is between 0 and $r0$. (Note also that $r0 < r3 < r2 < r1$.) See Figure 7. There are two equilibria: the disease-free equilibrium $E_0$ and the worst case $E_3$. $E_0$ is a saddle, and all positive semi-paths leading from points in $\Omega$ other than $E_0$ and its stable manifolds, converge to $E_3$. There is a heteroclinic path from $E_0$ to $E_3$. Notice that since $E_3 = (S_3, I_3) = \left( \frac{d + y + \epsilon}{\lambda}, \frac{A \lambda - d^2 - d(y + \epsilon)}{(d + y + \epsilon) \lambda} \right)$, the ultimate outcome for all initial points other than $E_0$ and its stable manifolds is invariant to small changes in $r$ as long as $r \in [0, r0]$.

![Figure 7](Epidemic model paper F January 2010.nb)

**Figure 7.** Positive trajectories from all initial points other than $E_0$ and its stable manifolds lead to $E_3$. \{A, \lambda, d, y, \epsilon, g, r\} = \{460, .00125, .2, .312, .002, .0025, 1\}. $r < r0$. $R_0 = 1.90.$

When $r$ initially rises past $r0$, equilibria $E_1$ and $E_2$ simultaneously appear. With the present assumption on the parameters, $r0 > 0$ and $\sqrt{A} - d > 0$, so there will now be four equilibria in $\Phi$. Since our current assumptions also yield $r3 > r0$, $E_1$ will be locally unstable. Figure 8 depicts a representative case. $E_0$ and $E_2$ are saddles, and $E_3$ is locally stable. Note that both unstable manifolds of $E_2$ have heteroclinic connections (shown as bold lines) with $E_3$, with the two paths forming an invariant set surrounding $E_1$. Positive simi-paths originating within this set at points other than $E_1$ and those on the stable manifold of $E_2$ flow to $E_3$, while the $\alpha$-limit set for the contained stable manifold of $E_2$ is $E_1$. $E_3$ is approached by all positive semi-paths originating in $\Omega$ other than $E_0$, $E_1$, $E_2$, and the stable manifolds of $E_0$ and $E_2$. 
Figure 8. The dominant flow is to $E_3$. \{A, \lambda, d, \gamma, \epsilon, g, r\} = \{460, .00125, .2, .312, .002, .0025, 1.3\}. $r_0 < r < r_3$. $R_0 = 1.58$.

As $r$ increases further to values where $r_3 < r < r_2$, $E_1$ becomes locally stable. The result is bi-stability of $E_1$ and $E_3$, with the domain of attraction for $E_1$ given by the interior of an unstable limit cycle surrounding $E_1$. The limit cycle is the $\alpha$-limit set of the right-side stable manifold of $E_2$. See Figure 9. All initial points in $\Omega$ other than $E_0$, $E_1$, $E_2$, the points on the stable manifolds of $E_0$ and $E_2$, and points on the interior of the limit cycle approach $E_3$ as $t \to \infty$. 

Figure 9. Bi-stable case with unstable limit cycle. \( \{A, \lambda, d, \gamma, \epsilon, g, r\} = \{460, .00125, .2, .312, .002, .0025, 1.35\} \). \( r_1 > r_2 > r > r_3 > r_0 \). \( R_0 = 1.54 \).

Compare Figure 9 and Figure 11 noticing that in Figure 9 (with \( r = 1.35 \)) the lower unstable manifold of \( E_2 \) below and to the right of the lower stable manifold of \( E_2 \), while in Figure 11 (where \( r = 1.52 \)) the lower unstable manifold of \( E_2 \) passes above and inside the stable manifold of \( E_2 \). This suggests that there will be an \( r \) between 1.35 and 1.52 which yields an homoclinic loop through \( E_2 \). This appears to happen with \( r \) at approximately 1.3805, as shown in Figure 10. Points inside the loop spiral in to \( E_1 \).
Figure 10. Bi-stable case with unstable homoclinic cycle surrounding stable $E_1$. \( \{A, \lambda, d, \gamma, \epsilon, g, r\} = \{460, 0.00125, 2, 0.312, 0.002, 1.3805\} \). $R_0 = 1.52$. 
**Figure 11.** Bi-stability. \( \{ A, \lambda, d, \gamma, \epsilon, g, r \} = \{460, 0.00125, 0.2, 0.312, 0.002, 0.0025, 1.52\} \). \( R_0 = 1.41 \).

Figure 11 shows a further evolution as \( r \) rises to 1.52. \( E_1 \) is still a stable focus, but the limit cycle surrounding it has vanished. The two depicted paths flowing down and right from the neighborhood of \( (S, I) = (50, 340) \) divide as they head in the direction of \( E_2 \), one turning right to \( E_3 \) and the other making a sharp left and spiraling into \( E_1 \). Between those two paths must lie the stable manifold of \( E_2 \), which originates outside \( \Omega \). Otherwise, the characteristics are much like those of Figure 9.

A further increase in \( r \) to a value between \( r_2 \) and \( r_1 \) will cause \( E_2 \) and \( E_3 \) to vanish simultaneously. With the sole exception of \( E_0 \) and its stable manifolds, all initial points in \( \Omega \) have positive semi-paths that approach \( E_1 \). See Figure 12. A further increase in \( r \) beyond \( r_l \) will push \( E_1 \) out of \( \Omega \) leaving the disease-free equilibrium globally stable.

**Figure 12.** Except for the point \( E_0 \) and its stable manifolds, all points starting in \( \Omega \) flow to \( E_1 \). \( \{ A, \lambda, d, \gamma, \epsilon, g, r \} = \{460, 0.00125, 0.2, 0.312, 0.002, 0.0025, 1.9\} \), with \( r_0 < r_3 < r_2 < r < r_l \). \( R_0 = 1.19 \).
In the interest of preserving simplicity, we have held all parameters but \( r \) constant through the preceding exercise. One final example shows that yet other equilibria configurations are possible. Figure 13 shows that under some alternative parameter values, all equilibria but \( E_1 \) will remain in \( \Omega \) with \( E_0 \) and \( E_3 \) both locally stable. In this case the domain of attraction for the disease-free equilibrium is all points in \( \Omega \) below the stable manifold of \( E_2 \), while all points above the stable manifold converge to \( E_3 \). In this case it is clear that small displacements from \( E_0 \) will lead back to \( E_0 \), while large displacements from \( E_0 \) will lead to \( E_3 \). The displacements from \( E_0 \) might be thought of as the mutation of a new virus in a population initially at \( E_0 \), or perhaps the immigration of an infected population into the current population.

Figure 13. Bi-stability. Positive \( E_1 \) does not exist. \( \{A, \lambda, d, \gamma, \epsilon, g, r\} = \{1500, .002, 1, .5, .001, .008, 2.15\} \). 
\( r2 > r > r3 > r1 > r0 \). \( R_0 = .82 \).

7. Discussion
Notice the \( r \) term, which we have taken to be the indicator of the extent of the social mobilization for treatment, has two distinct sorts of effects. First, changes in \( r \) affect the existence and/or the values of three of the four equilibria--\( E_1 \), \( E_2 \), and \( E_3 \). (See Figures 2, 3, and 4.) And second, changes in \( r \) affect both the stability properties and the domains of attraction (if appropriate) of all equilibria. Note further that changes in \( r \) do not yield continuous effects upon the ultimate outcome of the spread of an infection. For example, when \( E_3 \) is the only stable equilibrium due to very low values of \( r \), small increases in \( r \) have no effect upon either the location or the stability of \( E_3 \). Also, once \( r \) passes a threshold value \( r = \text{max}(r1, r2) \), the disease-free equilibrium is globally stable, and additional increases in \( r \) have no long-term advantage. In intermediate cases, we see a smooth decrease in \( I_1 \) as \( r \) is increased. Models of this sort point to the social value of determining realistic parameter ranges. Also, having \( r > r2 \) ensures that the worst case equilibrium, \( E_3 \), will vanish.

The critical role of the assumptions on and the use of parameter \( A \) were mentioned in Section 3. The model requires that all newborns come into the susceptible class and that the number of births be independent of all other variables. Though these assumptions are common in the literature, we should be conscious of the limitations their use imposes on the domain of applicability of the model. (The assumption is certainly widespread. See, for example, Wang 2006, Xiao and Zhou 2006, Ruan and Wang 2003, Wang and Ruan 2004, Wang 2006, Zhang and Liu 2008, Li, Li, Ghosh 2009, Mareira and YuQuan 1997, Hu, Liu, and Wang 2008 who all use this assumption. See Hethcote 2000 for a more general discussion.) One problem with the above is that it means that the birthrate rises when the population falls and vice versa. It would make more sense to replace \( A \) with something like \( bN \), so that the inflow rate varies with the size of the population. The price we would pay for this added realism would be high, however, since it would preclude 2-dimensional analysis in (\( S, I \))-space. (Hu, Liu, and Wang 2008 have an interesting 3-D SIR model with treatment, though the \( A \)-assumption is still used. And numerous authors use the "\( bN \)" construct in models without treatment. See Wang 2006, for example.) This limitation is less critical if the model is intended for short run use, since population change would then be less significant. But more importantly, the use of \( A \) requires that all newborns arrive exclusively into the susceptible class. This may not be appropriate for an important subset of diseases, such as HIV, where newborns can arrive into the infected class. Also, for some diseases, natural immunity or a high level of disease resistance can be inherited. In this case newborns may arrive in the \( R \) sub-population. The examples of sickle cell trait and malaria fit this case.

Another area for further research involves the assumed impact of the treatment function. In the present construct, an increase in treatment due to a rise in \( r \) works by diminishing \( dI/dt \) compared with the no-treatment case, i.e., increases in \( r \) (weakly) diminish the instantaneous rate of spread of the infection. This can reduce the number of people who become infected, which in turn will change the number who die from the disease. But as currently modeled, both here and elsewhere, treatment does not directly impact the death rate among those who do become infected. Perhaps it would be worthwhile to model a system where the disease-related death rate is a function of society's level of treatment.

**Appendix**

Proof of 3a(i). We suppose that \( I_2(\tau 0) - \frac{r0}{g} > 0 \), and show that \( I_2(\tau 0 + f) - \frac{r0 + f}{g} > 0 \) is then positive for all \( f > 0 \).
$$I_2(r) - \frac{r_0}{g} = \frac{2\lambda d-2\lambda \sqrt{A} \sqrt{g} + 2\lambda \gamma + 2\lambda \epsilon}{2g\lambda}, \quad \text{and} \quad I_2(r) + f - \frac{r_0 + f}{g} = \frac{2\lambda f - f^2 \lambda + 4\sqrt{A} f^2 \sqrt{g} \lambda^2 - f \lambda}{2g\lambda}.$$ 

Since $$\frac{\sqrt{f^2 \lambda^2 + 4\sqrt{A} f^2 \sqrt{g} \lambda^2 - f \lambda}}{2g\lambda} > 0$$ and $$I_2(r) > 0$$, their sum, which is $$I_2(r + f) - \frac{r_0 + f}{g}$$, must be greater than zero. Hence, if $$I_2$$ is outside its domain at $$r_0$$, it will be outside its domain at all $$r > r_0$$.

References


