Abstract. We study the interplay between effects of disease burden on the host population and the effects of population growth on the disease incidence, in an epidemic model of SIR type with demography and disease-caused death. We revisit the classical problem of periodicity in incidences of certain autonomous diseases. Under the assumption that the host population has a small intrinsic growth rate, using singular perturbation techniques and the phenomenon of the delay of stability loss due to turning points, we prove that large-amplitude relaxation oscillation cycles exist for an open set of model parameters. Simulations are provided to support our theoretical results. Our results offer new insight into the classical periodicity problem in epidemiology. Our approach relies on analysis far away from the endemic equilibrium and contrasts sharply with the method of Hopf bifurcations.

Key words. epidemic models, periodicity in disease incidence, interepidemic period, turning point, delay of stability loss, relaxation oscillation cycles

AMS subject classifications. 34C26, 92D25

DOI. 10.1137/15M1038785

1. Introduction. Investigation of oscillations in disease incidence is of fundamental importance in mathematical epidemiology. Empirical data of disease incidence has shown clearly identifiable cyclic patterns in many common diseases, including diseases for which environmental influences do not appear to play an important role, such as measles, pertussis, chicken pox, and mumps [2, 20]. Mechanisms for this type of “autonomous oscillation” have been extensively studied in the mathematical epidemiology literature. These include, together with papers that introduced them, time delays in the transmission process [14, 20], varying total population size with density dependent demography and transmission [1, 22], nonlinear incidence forms [16], discrete age-structures with a nonsymmetric contact matrix among age groups [8], and seasonality in the transmission process in both deterministic and stochastic models [3, 4, 12, 20]. The mathematical approach for these earlier works was bifurcation analysis (e.g., Hopf bifurcation theory), which analyzes model behaviors in a neighborhood of an endemic equilibrium. In the case of Hopf bifurcation, a certain degree of complexity needs to be introduced into the transmission process to produce instability of the endemic equilibrium, and the bifurcation may occur in parameter regimes...
that are not biologically realistic. For more complete reviews of related work, we refer
the reader to [2, 13].

In the present paper, we apply a singular perturbation approach to this investi-
gation. Our goal is to reveal a simple and biologically sound mechanism that can
produce large-amplitude oscillations in disease incidence. Our basic assumption is
that the host population has a small intrinsic growth rate \( \varepsilon > 0 \), the difference be-
tween the natural birth rate and the natural death rate. This slow-growth assumption
is not biologically unrealistic. Demographic data has shown that annual population
growth rates in many industrialized countries have been only slightly above zero, in
the range of 0.01–0.001 per year, for a long period of time [28]. The slow-growth
assumption may also apply to animal populations on livestock farms, where, for eco-
nomic reasons, population may be kept near its carrying capacity, where the growth is
close to zero. Using the intrinsic growth rate \( \varepsilon \) as a perturbation parameter, we show
that a standard SIR epidemic model can be reformulated as a singularly perturbed
problem. Applying techniques from geometric singular perturbation and global cen-
ter manifold theory, we prove that for an open and biologically realistic parameter
regime, stable periodic oscillations exist in rather simple SIR models. Furthermore,
our analysis demonstrates that the periodic solution has a large amplitude of order \( O(1) \).
This overcomes a common drawback of Hopf bifurcation analysis where the
bifurcating periodic solutions are of small amplitude.

Relaxation oscillations demonstrate distinguished and robust cyclic patterns that
consist a gradual (slow) change in the state variables over a long period of time
followed by a sudden (fast) change. A distinction between relaxation oscillations and
harmonic oscillations was first made by van der Pol [29]. Relaxation oscillation cycles
have been used to explain fast-slow dynamics frequently observed in electrical circuits,
mechanics, and many other physical and natural systems. In the present paper, for
a simple epidemic model with a slowly growing host population, we show that the
periodic solutions are of relaxation oscillation type. An important characteristic of
the model under the slow-growth assumption is the existence of a turning point. This
is a point on the slow manifold with the population size at the critical community size
to support an epidemic [2]. In the presence of turning points, the dynamics under
general perturbations are extremely rich and complicated (see [23, 24, 25, 27]). For
the specific model problem at hand, the disease-free subspace is invariant for \( \varepsilon \geq 0 \)
since the disease will not develop if it is not present at the initial time. Under the
invariance of the disease-free subspace, the turning point yields a critical phenomenon
called delay of stability loss, in which a solution starts with a fast motion to approach
a vicinity of the slow manifold, moves slowly along the slow manifold, passes through
the turning point, and continues the slow motion along the slow manifold, then, up to
some point, moves away from the slow manifold in a fast motion (see, e.g., [7, 17, 18,
23, 24, 25, 27]). In our model, the slow manifold is in the disease-free region, and the
time period a solution spends in the vicinity of the slow manifold corresponds to the
interepidemic period (IEP) with low disease incidence: the period between epidemics
(fast dynamics) away from the slow manifold. The fast-slow oscillations characterize
the global dynamics of the model and capture the qualitative nature of the oscillatory
behaviors in empirical disease data; see Figure 1 for a set of data on reported cases
of rubella in Canada during the period 1925–1960. Our analysis of the simple SIR
models has demonstrated that the existence of turning points and the associated delay
of stability loss due to the slow growth of the population offers a simple and robust
mechanism for sustained oscillations of disease incidence.
Mathematically, our singular perturbation analysis is done for a three-dimensional system, and the presence of turning points leads to a significant challenge. At a turning point, two eigenvalues are zero. This results in the loss of normal hyperbolicity of the one-dimensional slow manifold, and the standard geometric singular perturbation theory of Fenichel [9, 10] no longer applies. Another difficulty we encounter in the analysis is having to deal with the nonlinear dynamics in a large neighborhood of the slow manifold. Such a difficulty does not seem to appear in the analysis of many other biological models, e.g., in the analysis of relaxation oscillation of a predator-prey model [19].

The primary objective of our paper is to establish the mathematical framework and carry out detailed mathematical analysis for the singular perturbation approach to the study of epidemic models. We have chosen a simple SIR model to keep the mathematical technicality to its minimum, and the analysis is applicable to more complex models. In a subsequent paper, we will investigate relaxation oscillations in a SEIR model and give a more in-depth discussion of biological implications of the mathematical results. The singular perturbation approach and associated asymptotic analysis have been successfully applied to the analysis of relaxation oscillation phenomena in many mechanical, physical, chemical, and biological systems. We hope that our study will lead to more applications of singular perturbation analysis to the study of disease transmission processes.

2. The model and statements of main results.

2.1. The model problem. Consider the spread of an infectious disease in a host population of size $N$. Partition the population into susceptible, infectious, and recovered classes, and denote the sizes by $S$, $I$, and $R$, respectively, so that $N = S + I + R$.

In the absence of the disease, we assume that $N$ satisfies

$$N' = \varepsilon g(N),$$

where constant $\varepsilon > 0$ is assumed to be small. A typical example of $g(N)$ is the quadratic form $N(1 - N/N^*)$, such that $N$ has the logistic growth with carrying capacity $N^*$ and intrinsic growth rate $\varepsilon$. It is natural to require the following.

(A1) The function $g(N)$ satisfies

$$g''(N) < 0, \quad g(0) = g(N^*) = 0 \text{ for some } N^* > 0.$$
As a consequence, we have the following properties.

**Lemma 1.** Assume (A1). Then, $N^*$ is unique, and $g(N) > 0$ for $N \in (0, N^*)$ and $g(N) < 0$ for $N > N^*$.

We further assume that the per capita natural death rate is a constant $d > 0$, and newborns $b(N)$ has a density dependent form $b(N) = dN + \varepsilon g(N)$. For simplicity, we assume that all newborns are susceptible to the disease. We consider the type of diseases that spread through direct contact of hosts, and incidence is given by $h(S, N)I$, where $h(S, N)$ is a smooth function. We will assume the following basic properties on $h(S, N)$:

(A2) The function $h(S, N)$ is increasing in $S$ and $h(0, N) = 0$.

A specific form of $h(S, N)$ that is commonly used is

$$h(S, N) = \frac{\beta(N) S^q}{K + S}, \quad q \geq 1, \ K \geq 0.$$  

This incidence form $h(S, N)I$ includes the bilinear incidence $\beta SI$ (with $K = 0$, $q = 2$), nonlinear incidence $\beta S^{q-1}I$ (with $K = 0$, $q > 2$), standard incidence $\lambda SI$ (with $\beta(N) = \lambda/N$, $K = 0$, $q = 2$), and saturation incidence $\frac{\beta SI}{K + S}$.

The transmission process is demonstrated in the following diagram:

![Transmission process diagram]

The parameter $\gamma$ denotes the recovery rate, and $p$ denotes vaccination rate for a simple vaccination strategy. We assume that the infectious individuals suffer a disease-caused death $\alpha I$ with a constant rate $\alpha$. It is assumed that disease confers permanent immunity, and all parameters are assumed to be positive. The transfer diagram leads to the following system of differential equations:

$$S' = b(N) - h(S, N)I - (d + p)S,$$

$$I' = h(S, N)I - (d + \gamma + \alpha)I,$$

$$R' = pS + \gamma I - dR.$$  

(1)

As a consequence, the total population size $N$ satisfies

$$N' = \varepsilon g(N) - \alpha I.$$  

(2)

It follows that for $\varepsilon > 0$ and $\alpha > 0$, $N$ varies with time, and model (1) is a three-dimensional system.

Using $b(N) = dN + \varepsilon g(N)$ and replacing the $R$ equation by (2), we rewrite the model (1) as the following equivalent system:

$$S' = dN + \varepsilon g(N) - h(S, N)I - (d + p)S,$$

$$I' = h(S, N)I - \alpha I,$$

$$N' = \varepsilon g(N) - \alpha I.$$  

(3)
where \( a = d + \alpha + \gamma \). We study system (3) for \( \epsilon \geq 0 \) in the feasible region

\[
\mathcal{D} = \{(S, I, N) \in \mathbb{R}^3 : S \geq 0, \ I \geq 0, N \geq 0 \text{ and } S + I \leq N \leq N^*\}.
\]

From Lemma 1 and (2) we know that \( N' < 0 \) if \( N > N^* \). If follows that the region \( \mathcal{D} \) is positively invariant with respect to system (3) and globally attracts all nonnegative solutions of (3).

Global dynamics of model (3) for the case \( \epsilon = 0 \) were studied in [11]. It was shown that the essential dynamics consist of a local, stable, two-dimensional invariant manifold and, on the invariant manifold, a line of equilibria exists and all other solutions are heteroclinic orbits each connecting a pair of equilibria. This is a highly unstable structure and small perturbations can dramatically change the nature of the global dynamic. We will study the global dynamics of (3) for the case \( \epsilon > 0 \) and show that, under certain conditions, there exists a stable relaxation periodic cycle for small \( \epsilon \).

In the rest of this section, we describe the structure of the equilibria and their stability and state our main result on relaxation oscillations.

2.2. Structure of equilibria and statement of the main result. For \( \epsilon \geq 0 \), \((0, 0, 0)\) and \((S^*, 0, N^*)\), with \( N^* \) defined in (A1) and \( S^* = dN^*/(d+p) \), are equilibria of system (3).

**Proposition 2.** There are no other equilibria for \( \epsilon > 0 \) if and only if

\[
h(S(N), N) < a = d + \gamma + \alpha
\]

for all \( N \), where

\[
S(N) = \frac{d}{d + p} N - \epsilon \frac{a - \alpha}{\alpha(d + p)} g(N).
\]

Furthermore, if \( h(S, N) < a \) for all \( S \) and \( N \), then the equilibrium \((S^*, 0, N^*)\) attracts all solutions except \((0, 0, 0)\). The global dynamics are trivial.

**Proof.** The first statement can be checked directly. Assume that \( h(S, N) < a \) for all \( S \) and \( N \). Then, for any initial condition other than \((0, 0, 0)\), the solution \((S(t), I(t), N(t))\) satisfies that \( I(t) \to 0 \) as \( t \to +\infty \) and, on the plane \( \{I = 0\} \), \((S(t), N(t)) \to (S^*, N^*)\) as \( t \to \infty \).

In this work, we will focus on the cases where nontrivial dynamics are possible. In view of the statements in Proposition 2, we assume the following:

(A3) The function \( h(dN/(d + p), N) \) is nondecreasing for \( N \in (0, N^*) \). There is a unique \( N_0 \in (0, N^*) \) such that \( h(S_0, N_0) = a \), where \( S_0 = \frac{d}{d + p} N_0 \). Furthermore,

\[
\frac{d}{d + p} h_S(S_0, N_0) + h_N(S_0, N_0) > 0.
\]

Assumption (A3) is biologically intuitive since the force of infection \( h(S, N) \) should increase as the population size \( N \) increases. We note that \((S_0, 0, N_0)\) in (A3) has both dynamical and biological significance. In the case when \( h(S, N) = \beta S \), the equation \( h(S_0, N_0) = a \) becomes \( \beta S_0 = d + \gamma + \alpha \) and thus \( S_0 = (d + \gamma + \alpha)/\beta \). In the classical SIR model with no demography \((b = d = 0)\) and no disease-caused death \((\alpha = 0)\), we have \( S_0 = \gamma/\beta \), which is known as the critical size of susceptible population to sustain an epidemic [2, 12]. The dynamical significance of point \((S_0, 0, N_0)\) is that it is a turning point, whose existence is the foundation of the relaxation oscillation phenomenon.

**Lemma 3.** Assume that (A3) holds. For \( \epsilon > 0 \) small, there is a unique equilibrium \( E_\epsilon = (S_\epsilon, I_\epsilon, N_\epsilon) \) with \( S_\epsilon, I_\epsilon, N_\epsilon > 0 \), and \( E_\epsilon \to (S_0, 0, N_0) \) as \( \epsilon \to 0 \).
Proof. In addition to \((0,0,0)\) and \((S^*,0,N^*)\), other equilibria of system (3) are determined by

\[
h(S,N) = a, \quad I = \frac{\varepsilon}{\alpha}g(N), \quad S = \frac{d}{d+p} N - \varepsilon \frac{a-\alpha}{\alpha(d+p)} g(N).
\]

The \(N\) coordinates are roots of

\[
f(N;\varepsilon) := h\left(\frac{d}{d+p} N - \varepsilon \frac{a-\alpha}{\alpha(d+p)} g(N), N\right) - a = 0.
\]

It follows from assumption (A3) that

\[
f(N_0;0) = 0, \quad f_N(N_0;0) = \frac{d}{d+p} h_S(S_0,N_0) + h_N(S_0,N_0) > 0.
\]

An application of the implicit function theorem gives that for \(\varepsilon > 0\) small, there is \(N_{\varepsilon}\) such that \(f(N_{\varepsilon};\varepsilon) = 0 \) and \(N_{\varepsilon} \rightarrow N_0\) as \(\varepsilon \rightarrow 0\). Note that the corresponding \(I\)-coordinate is \(I_{\varepsilon} = \frac{\varepsilon}{\alpha}g(N_{\varepsilon}) > 0\) for \(\varepsilon > 0\) small.

Stability of equilibria of system (3) is described in the next result, whose proof is given in Appendix I. Denote

\[
\Delta_0 = \left(\frac{a}{\alpha} - \frac{d}{d+p}\right) h_S(S_0,N_0)g(N_0) - (d+p)g_N(N_0).
\]

**Theorem 4.** Assume that (A1), (A2), and (A3) hold. Then, for \(\varepsilon > 0\) small,

(i) the equilibria \((0,0,0)\) and \((S^*,0,N^*)\) are saddles each with two negative eigenvalues and one positive eigenvalue;

(ii) the equilibrium \(E_{\varepsilon}\) always has a real negative eigenvalue and a pair of complex conjugate eigenvalues. If \(\Delta_0 > 0\), then the complex eigenvalues have a negative real part and \(E_{\varepsilon}\) is locally stable; if \(\Delta_0 < 0\), then the complex eigenvalues have a positive real part and \(E_{\varepsilon}\) is a saddle.

A rough statement of our main result is given in the following. A more technical statement (Theorem 10) of this result and its proof will be given in section 4.

**Theorem 5.** Assume that (A1), (A2), and (A3) hold. Then, for system (3) with \(\varepsilon > 0\) small, one of the following holds:

(i) the equilibrium \(E_{\varepsilon}\) is a sink and it attracts all orbits except equilibria \((0,0,0)\) and \((S^*,0,N^*)\);

(ii) there exists an invariant annulus-like or disk-like two-dimensional region that attracts all but equilibria orbits and contains at least one stable periodic orbit.

We note that for fixed \(\varepsilon > 0\) small, as \(\Delta_0\) varies from positive to negative, in view of statement (ii) in Theorem 4, it is possible that a periodic solution can be created through a supercritical Hopf bifurcation of \(E_{\varepsilon}\). This has been extensively studied for many biological models in the literature. We will not pursue this direction. Instead, we will investigate the existence of a relaxation oscillation using a global approach. More precisely, we will treat \(\varepsilon\) as a parameter, first understand the limiting global behaviors when \(\varepsilon = 0\), and then examine how a relaxation oscillation is created for \(\varepsilon > 0\), far from the endemic equilibrium \(E_{\varepsilon}\). In particular, the example in section 4.3 shows that a stable relaxation oscillation may exist even if \(E_{\varepsilon}\) is stable.
3. Global dynamics of system (3) for \( \epsilon = 0 \). In this section, we give a complete description of the dynamics for the limiting system (3) at \( \epsilon = 0 \). The result extends the work in [11] for a semilocal description of the dynamics. We recall that system (3) for \( \epsilon = 0 \) is

\[
\begin{align*}
S' &= dN - h(S,N)I - (d + p)S, \\
I' &= (h(S,N) - a)I, \\
N' &= -\alpha I
\end{align*}
\]

with feasible region \( \mathcal{D} = \{(S,I,N) \in \mathbb{R}^3 : S \geq 0, I \geq 0, N \geq 0, S + I \leq N \leq N^*\} \), which is positively invariant for (5).

It can be verified that the disease-free plane \( \{I = 0\} \) and the half-line

\[
Z_0 := \left\{ S = \frac{dN}{d+p}, I = 0, \quad N \geq 0 \right\}
\]

are both invariant under system (5). In particular, \( Z_0 \) consists of equilibria of (5).

3.1. A complete characterization of dynamics of (5). On the invariant plane \( \{I = 0\} \), all solutions \( (S(t), I(t), N(t)) \) satisfy that

\[
I(t) \equiv 0, \quad N(t) \equiv N(0), \quad \text{and} \quad S(t) \rightarrow \frac{d}{d+p}N(0) \quad \text{as} \quad t \rightarrow \infty.
\]

The set \( Z_0 \) of equilibria attracts all solutions within \( \{I = 0\} \).

The linearization at each point \((dN/(d+p), 0, N) \in Z_0\) is

\[
\begin{pmatrix}
-(d+p) & -h(dN/(d+p), N) & d \\
0 & h(dN/(d+p), N) - a & 0 \\
0 & -\alpha & 0
\end{pmatrix}
\]

with eigenvalues \( \lambda_1 = 0, \lambda_2 = -(d+p) < 0, \) and \( \lambda_3 = h(dN/(d+p), N) - a \). The eigenvectors associated with \( \lambda_1 \) and \( \lambda_2 \) span the plane \( \{I = 0\} \) and that associated with \( \lambda_3 \) is transversal to the plane \( \{I = 0\} \). The eigenvalue \( \lambda_3 = h(dN/(d+p), N) - a \) changes sign across the point \((S_0, 0, N_0) \in Z_0\), where \( S_0 \) and \( N_0 \) are defined in (A3).

The complete dynamics for the case \( \epsilon = 0 \) are described in the following result and depicted in Figure 2. The proof is given in Appendix I.

**Theorem 6.** Assume that (A2) and (A3) are satisfied. Then the following statements hold:

(i) Every solution of system (5) is bounded for \( t \geq 0 \) and the set \( Z_0 \) is the global attractor.

(ii) The unstable manifold of each equilibrium \((dN/(d+p), 0, N) \in Z_0\) with \( N > N_0 \) is a heteroclinic orbit to an equilibrium \((S, 0, N) \in Z_0\) with \( 0 < \bar{N} < N_0 \). The relationship \( \bar{N}_1 < \bar{N}_2 < N_0 \) if \( N_1 > N_2 > N_0 \) holds. Furthermore, \( \lim_{N \rightarrow \infty} \bar{N} = N_\infty \in (0, N_0) \).

We denote by \( M(Z_0) \) the two-dimensional invariant manifold that consists of heteroclinic orbits established in Theorem 6(ii), and define a map

\[
H : (N_0, \infty) \rightarrow (0, N_0), \quad H(N) = \bar{N}
\]

where \( \bar{N} \) is defined by the heteroclinic orbits in Theorem 6(ii). The invariant manifold \( M(Z_0) \) and the map \( H \) will play important roles in our results on relaxation oscillations for model (3) with \( \epsilon > 0 \).
3.2. Persistence of $M(Z_0)$ for $\varepsilon > 0$ small. We are interested in whether the invariant manifold $M(Z_0)$ will persist for $\varepsilon > 0$ small, that is, for $\varepsilon > 0$ small, whether there is an invariant manifold $M_\varepsilon$ for system (3) so that $M_\varepsilon \to M(Z_0)$ as $\varepsilon \to 0$.

Recall that when $\varepsilon = 0$, for each equilibrium $w = \left(\frac{d}{d+p}N, 0, N\right) \in Z_0$, the eigenvalues of the linearization at $w$ are

$$\lambda_1 = 0, \quad \lambda_2 = -(d + p), \quad \lambda_3 = h\left(\frac{d}{d+p}N, N\right) - a.$$ 

Based on the relative size of eigenvalues, the consideration can be divided into two cases.

Case 1: $a = d + \alpha + \gamma < d + p$. It follows that $h\left(\frac{d}{d+p}N, N\right) - a > -(d + p)$ for all $N \geq 0$. At each point $w \in Z_0$, we have $\lambda_1 > \lambda_2$ and $\lambda_3 > \lambda_2$. Applying a center manifold theorem in [5, 6] to the invariant set $Z_0$, we obtain the existence of a two-dimensional center manifold $W^c(Z_0)$. The center manifold $W^c(Z_0)$ is invariant under (5) and contains $Z_0$ and all orbits bounded in the vicinity of $Z_0$. At each $w \in Z_0$, the tangent space $T_w W^c(Z_0)$ is spanned by the eigenvectors associated with $\lambda_1$ and $\lambda_3$ (both are larger than $\lambda_2$). Most importantly, the center manifold theorem guarantees the persistence of $W^c(Z_0)$ for $\varepsilon > 0$ small. In general, a center manifold may not be unique but any center manifold will contain all orbits that are bounded in the vicinity of $Z_0$. Therefore, for this model problem, $W^c(Z_0)$ coincides with $M(Z_0)$ and is unique, and $M(Z_0)$ persists for $\varepsilon > 0$.

Case 2: $a = d + \alpha + \gamma \geq d + p$. In this case, there exists a unique $N_0 \in [0, N_0)$ such that $h\left(\frac{d}{d+p}N, N\right) - a > -(d + p)$ for $N > N_0$ but $h\left(\frac{d}{d+p}N, N\right) - a \leq -(d + p)$ for $N \leq N_0$. The general results on center manifolds in [5, 6] cannot be applied to the whole set $Z_0$ to obtain a two-dimensional center manifold. For any fixed $\delta > 0$, the results in [5, 6] can be applied to the subset $Z_0^\delta := Z_0 \cap \{N \geq N_0 + \delta\}$ but the corresponding center manifold $W^c(Z_0^\delta)$ will only be a proper subset of $M(Z_0)$. It turns out, for $\varepsilon > 0$, that parts of relaxation oscillations could occur outside $W^c(Z_0^\delta)$ for all $\delta > 0$. We take the advantage of a crucial property that the set $\{I = 0\}$ is invariant under system (3) for all $\varepsilon \geq 0$ and show that $M(Z_0)$ persists for $\varepsilon > 0$ small even though it is not normally hyperbolic. This is established in Appendix II. This persistence result appears to be contradictory to Mâné’s result that an invariant manifold is persistent if and only if it is normally hyperbolic [21]. It is not, since the persistence in Mâné’s result is with
respect to all small perturbations, while the perturbations, in our system are special: they leave the set \( \{ I = 0 \} \) invariant. As mentioned above, it is possible that a portion of a relaxation oscillation occurs over the region where \( N < \hat{N} \). In the limit as \( \varepsilon \to 0 \), this portion approaches \( Z_0 \) along the eigenvector associated with \(- (p + q)\) in general.

### 3.3. The map \( H \) near \( N_0 \)

The map \( H : (N_0, \infty) \to (0, N_0) \) defined in Theorem 6 will be a key ingredient for our main result on relaxation oscillations. Detailed global properties of \( H \) seem to be not achievable. On the other hand, it is possible to examine properties of \( H \) near \( N_0 \) based on an approximation of \( W^c(Z_0) \) near \((S_0, 0, N_0)\) or, simply, a center manifold \( W^c(S_0, 0, N_0) \) of the equilibrium \((S_0, 0, N_0)\). Note that the eigenvalues at \((S_0, 0, N_0)\) are \( \lambda_1 = \lambda_3 = 0 > \lambda_2 = -(d + p) \). Thus, for an equilibrium \( w \in Z_0 \) near \((S_0, 0, N_0)\), the corresponding eigenvalues satisfy \( \lambda_1 > \lambda_2 \) and \( \lambda_3 > \lambda_2 \). As a consequence, \( W^c(S_0, 0, N_0) \subset M(Z_0) \) and hence is unique. It should be pointed out that, in general, a center manifold may not be unique.

#### 3.3.1. An approximation of the center manifold \( W^c(S_0, 0, N_0) \)

We look for an approximation of the center manifold \( W^c(S_0, 0, N_0) \) in the vicinity of \((S_0, 0, N_0)\) as the graph of a function

\[
S = \frac{d}{d + p} N + U(N, I) I = \frac{d}{d + p} N + a_0(N) I + a_1(N, I) I^2.
\]

The form is justified by the fact that \( \{ I = 0 \} \) is invariant and \( W^c(S_0, 0, N_0) \cap \{ I = 0 \} \subset Z_0 \).

Taking the derivative of \( S = \frac{d}{d + p} N + U(N, I) I \) with respect to \( t \), we have

\[
S' = \frac{d}{d + p} N' + a_0' I N' + a_1, N I^2 N' + a_0 I' + 2a_1 I I' + a_1, I I^2 I'.
\]

From (5) we have

\[
dN - h \left( \frac{d}{d + p} N + a_0(N) I + a_1(N, I) I^2, N \right) I
\]

\[- (d + p) \left( \frac{d}{d + p} N + a_0(N) I + a_1(N, I) I^2 \right)
\]

\[= - \alpha \left( \frac{d}{d + p} + a_0 I + a_1, N I^2 \right) I
\]

\[+ (a_0 + 2a_1 I + a_1, I I' \prime) \left( h \left( \frac{d}{d + p} N + a_0(N) I + a_1(N, I) I^2, N \right) - a \right) I.
\]

Expanding \( h \) at the point \((bN/(b + p), N)\) we get

\[
dN - h \left( \frac{d}{d + p} N, N \right) I - (d + p) \left( \frac{d}{d + p} N + a_0(N) I \right) + O(I^2)
\]

\[= - \frac{ad}{d + p} I + a_0 \left( h \left( \frac{d}{d + p} N, N \right) - a \right) I + O(I^2).
\]

Comparing coefficients of \( I^0 \) and \( I^1 \) we obtain

\[
a_0(N) = \frac{ad}{d + p} - h \left( \frac{d}{d + p} N, N \right) \frac{d + p + h \left( \frac{d}{d + p} N, N \right) - a}{d + p + h \left( \frac{d}{d + p} N, N \right) - a}.
\]

Note that we restrict the approximation of \( W^c(S_0, 0, N_0) \) near \((S_0, 0, N_0)\). Thus, \( N \) is close to \( N_0 \), and hence, the denominator in the above expression is close to \( d + p > 0 \).
Near equilibrium \((S_0, 0, N_0)\), the center manifold \(W^c(S_0, 0, N_0)\) is given as the graph of the function

\[
S = \frac{d}{d+p} N + a_0(N)I + O(I^2)
\]

\[
= \frac{d}{d+p} N + \frac{ad}{d+p} - h\left(\frac{d}{d+p}N, N\right)I + O(I^2).
\]

On the center manifold \(W^c(S_0, 0, N_0)\) and near \((S_0, 0, N_0)\), system (5) is reduced to a two-dimensional system,

\[
I' = h\left(\frac{dN}{d+p} + a_0(N)I + O(I^2), N\right) I - aI,
\]

\[
N' = -\alpha I.
\]

### 3.3.2. Properties of the map \(H\) near \(N_0\).

**Proposition 7.** The map \(H\) satisfies \(H(N_0) = N_0\), \(H'(N_0) = -1\), and \(H''(N_0) = -\alpha a_0(N_0)h_S(S_0, N_0)\).

**Proof.** Set \(v(t) = N(t) - N_0\). In terms of \((I, v)\), system (7) becomes

\[
I' = h\left(S_0 + \frac{dv}{d+p} + a_0(N_0 + v)I + O(I^2), N_0 + v\right) I - aI,
\]

\[
v' = -\alpha I.
\]

Since \(\{I = 0\}\) is invariant and the map \(H\) is defined through the dynamics where \(I > 0\), we divide the two equations above to get

\[
\frac{dI}{dv} = -\frac{1}{\alpha} \left(h\left(S_0 + \frac{dv}{d+p} + a_0(N_0 + v)I + O(I^2), N_0 + v\right) - a\right).
\]

Expanding the right-hand side at \(v = 0\) leads to

\[
h\left(S_0 + \frac{dv}{d+p} + a_0(N_0 + v)I + O(I^2), N_0 + v\right) - a
\]

\[
= h_S \cdot \left(\frac{dv}{d+p} + (a_0 + a_0'v)I\right) + h_N \cdot v + \frac{1}{2} h_{SS} \cdot \left(\frac{dv}{d+p} + (a_0 + a_0'v)I\right)^2
\]

\[
+ h_{SN} \cdot \left(\frac{dv}{d+p} + (a_0 + a_0'v)I\right) v + \frac{1}{2} h_{NN} \cdot v^2 + O(I^2, v^2 I, I^3),
\]

where the partial derivatives of \(h\) are all evaluated at \((S_0, N_0)\), and \(a_0 = a_0(N_0)\) and \(a_0' = a_0'(N_0)\). Denote

\[
L = \frac{d}{d+p} h_S + h_N \quad \text{and} \quad Q = \frac{d^2}{(d+p)^2} h_{SS} + \frac{2d}{d+p} h_{SN} + h_{NN}.
\]

Equation (8) becomes

\[
\frac{dI}{dv} = -\frac{1}{2\alpha} (2Lv + Qv^2)
\]

\[
- \frac{1}{\alpha} \left(h_S + \left(\frac{d}{d+p} h_{SS} + h_{SN}\right) v\right) (a_0 + a_0'v)I + O(I^2, v^2 I, v^3).
\]
By the existence and smoothness of solutions and smooth dependence of solutions on parameters, for \(v\) small, we look for solutions of the form \(I(v) = c_0 + c_1 v + c_2 v^2 + O(v^3)\). Substituting \(I(v)\) into (10) and comparing terms of like powers in \(v\) we get

\[
\begin{align*}
  c_1 &= -\frac{1}{\alpha} a_0 h_S c_0 + O(c_0^2), \\
  c_2 &= -\frac{1}{2\alpha} L - \frac{1}{2\alpha} \left( a_0 h_S - \frac{1}{\alpha} a_0^2 h_S^2 + \frac{d}{d+p} a_0 h_{SS} + a_0 h_{SN} \right) c_0 + O(c_0^2).
\end{align*}
\]

Thus, near \(v = 0\), the solution of (10) is

\[
I(v) = c_0 - \frac{a_0 h_S v}{\alpha} c_0 - \frac{L}{2\alpha} v^2 + O(c_0 v^2).
\]

To define \(H\), we need the initial condition \(I(v) = 0\) at \(v = N - N_0\) for \(N > N_0\) and \(N - N_0 \ll 1\). We can then determine the value \(c_0\) corresponding to this initial condition. From (12),

\[
0 = c_0 - \frac{a_0 h_S \cdot (N - N_0)}{\alpha} c_0 - \frac{L}{2\alpha} (N - N_0)^2 + O(c_0(N - N_0)^2),
\]

or equivalently,

\[
c_0 \left( 1 - \frac{1}{\alpha} a_0 h_S \cdot (N - N_0) + O(N - N_0)^2 \right) = \frac{L}{2\alpha} (N - N_0)^2.
\]

Thus,

\[
0 = \frac{L}{2\alpha} (N - N_0)^2 + \frac{L a_0 h_S}{2\alpha^2} (N - N_0)^3 + O(N - N_0)^4.
\]

The value of \(H(N)\) satisfies \(I(H(N) - N_0) = 0\). Note that

\[
H(N) - N_0 = H(N) - H(N_0) = H'(N_0)(N - N_0) + \frac{1}{2} H''(N_0)(N - N_0)^2 + O(N - N_0)^3.
\]

It then follows from \(I(H(N) - N_0) = 0\), (11), (12), and (13) that

\[
0 = \frac{L}{2\alpha} (N - N_0)^2 + \frac{L a_0 h_S}{2\alpha^2} (1 - H'(N_0))(N - N_0)^3
- \frac{L}{2\alpha} \left( H'(N_0)(N - N_0) + \frac{1}{2} H''(N_0)(N - N_0)^2 \right)^2 + O(N - N_0)^4.
\]

Comparing \((N - N_0)^2\) terms gives that \(H'(N_0) = -1\) (due also to that \(H\) is decreasing). The \((N - N_0)^3\) terms then yield

\[
\frac{L a_0 h_S}{\alpha^2} + \frac{L}{2\alpha} H''(N_0) = 0.
\]

This completes the proof.

**3.4. A discussion and the link to the main result.** In this section, we summarize the results for system (5), discuss the impact of the sign changing eigenvalue \(h(S, N) - a\), and provide mathematical and biological motivations for our main result.

For \(N < N_0\), \(h(dN/(d+p), N) - a < 0\), and it implies that for an initial state \((S(0), I(0), N(0))\) near the region \(\{I = 0, N < N_0\}\), \(I(t)\) decreases and the solution
converges to an equilibrium in $Z_0$ with $N < N_0$. Biologically speaking, if the total population $N$ is below the critical community size $N_0$ or, equivalently, the number of susceptibles $S$ is below the critical size $S_0 = dN_0/(d + p)$, then the population cannot sustain an epidemic and the disease dies out.

We describe the dynamics for solutions with initial conditions near the other region $\{I = 0, N > N_0\}$ in three stages.

Stage I. For an initial state $(S(0), I(0), N(0))$ with $N > N_0$, $h(dN/(d + p), N) - a > 0$ for small $t > 0$ and $I(t)$ increases initially. In biological terms, if the population size surpasses the critical community size $N_0$, then any initial infection will lead to a disease outbreak.

Stage II. As $I(t)$ increases away from $\{I = 0\}$, the dynamics outside $\{I = 0\}$ become dominant; in particular, $N(t)$ decreases. Once $N(t) < N_0$ (or equivalently $S(t) < S_0$), we know that $h(S, N) - a < 0$ and $I(t)$ begins to decrease. The solution follows a heteroclinic orbit depicted in Figure 2.

Stage III. As time goes on, $I(t)$ continues to decrease. Eventually the solution will enter a vicinity of the region $\{I = 0, N < N_0\}$ and is attracted to an equilibrium in $Z_0$ with $N < N_0$. The disease outbreak leads to an epidemic but the disease eventually dies out.

We see that when $\varepsilon = 0$, model (5) only describes epidemics of the disease; the disease eventually dies out. There is no mechanism for the recurrence of the disease if the population growth is zero. This is parallel to the classical SIR model with no demography and disease-caused death.

When $\varepsilon > 0$, solutions of system (3) with $N(0) > N_0$ and $I(0)$ small go through Stages I and II as described above, but Stage III will no longer be the terminal stage. In this case, the disease-free set $\{I = 0\}$ remains invariant. The half-line $Z_0$ also remains invariant but is no longer a set of equilibria. Instead, $Z_0$ becomes an orbit for which $N$ increases with $t$ with speed of order $O(\varepsilon)$. For this reason, $Z_0$ is called the slow manifold for small $\varepsilon > 0$.

Stage IV. When $\varepsilon > 0$, for a solution in the vicinity of $Z_0$ with $N < N_0$ during Stage III, it will follow an orbit on the slow manifold $Z_0$ by the continuous dependence on initial conditions. As $N(t)$ increases beyond the critical community size $N_0$, the solution enters the region $\{I = 0, N > N_0\}$. As a consequence, $I(t)$ begins to increase and the solution repeats Stages I–III, leading to another epidemic. The period during which the solution moves along the slow manifold is the IEP. We see that when $\varepsilon > 0$, the fall of susceptible population during an epidemic and the recovery of the susceptible population during the IEP produce an oscillating behavior.

In summary, for $\varepsilon > 0$ small, all orbits, except for solutions on $\{I = 0\}$ and $E_\varepsilon$ of system (3), will exhibit oscillating behaviors. Three key conditions are responsible for the mechanism of oscillation:

(C0) the plane $\{I = 0\}$ is invariant for $\varepsilon \geq 0$,

(C1) the assumption on the natural growth $g(N)$ of the total population in the absence of disease, and

(C2) the sign changing assumption of the eigenvalue $\lambda_3 = h(S, N) - a$.

In the language of singular perturbation theory, condition (C2) means that the point $(S_0, 0, N_0)$ at which $h(S_0, N_0) - a = 0$ is a turning point. This point marks the level of $N$ or $S$ that separates the region of disease decline from that of disease rise. Condition (C1) implies that on $Z_0$, with the population growth and increase of susceptibles from newborns, all orbits move from a region of disease decline where $N < N_0$ to a region of disease rise where $N > N_0$. Conditions (C0)–(C2) imply that the turning point $(S_0, 0, N_0)$ is associated with the delay of stability loss [7, 17, 18,
Condition (C0) is the consequence of the biological fact that if the disease is not present at time $t = 0$, it remains absent from the population for $t \geq 0$. We emphasize that while condition (C0) holds true naturally for the specific model we consider, it is, however, highly degenerate in general when turning points are present. Without condition (C0), presence of turning points can make it is, however, highly degenerate in general when turning points are present. We require that

(P1) for $\epsilon = 0$, the heteroclinic orbits are determined by $v = \text{const}$ and the $u$-variable is decreasing from the right branch of slow manifold $v = T(u)$ to the left branch as time increases (see Figure 3);

(P2) for $\epsilon \geq 0$, the set $Z_\epsilon$ corresponds to the curve $\{v = T(u)\}$ for function $T : (0,U) \to (0,V)$ with $T(U) = V$, where $(U,V)$ corresponds to the point $(S,0,N^*+\delta) \in Z$, for arbitrarily fixed $\delta > 0$ independent of $\epsilon$, and hence $\{v = V\}$ corresponds to the heteroclinic orbit from $(S,0,N^*+\delta) \in Z_0$; therefore,

$$\mathcal{R} = \{(u,v) : 0 < u < U, T(u) \leq v < V\};$$
(P3) for \(\varepsilon \geq 0\), the point \((u, v) = (u_0, T(u_0))\) corresponds to the point \((S_0, 0, N_0)\), \((u, v) = (u^0, T(u^0))\) corresponds to \((S^*, 0, N^*)\).

![Diagram](image)

**Fig. 3.** Heteroclinic structure of (15) with \(\varepsilon = 0\).

In terms of \((u, v) \in \mathcal{R}\), suppose that system (3) on the center manifold can be put into the form

\[
\begin{align*}
    u' &= F(u, v; \varepsilon), \\
    v' &= G(u, v; \varepsilon).
\end{align*}
\]

We now examine the properties that the vector field of system (14) must satisfy.

First of all, (P1) implies that \(G(u, v; 0) = 0\), \(F(u, T(u); 0) = 0\), and \(F(u, v; 0) < 0\) for \(v > T(u)\). Thus, we can write \(G(u, v; \varepsilon) = \varepsilon G_1(u, v; \varepsilon)\), \(F(u, v; \varepsilon) = T(u) - v + \varepsilon F_1(u, v; \varepsilon)\). The property (P2) implies that \(G_1(u, T(u); \varepsilon) = T(u) F_1(u, T(u); \varepsilon)\).

System (14) can be rewritten as

\[
\begin{align*}
    u' &= T(u) - v + \varepsilon F_1(u, v; \varepsilon), \\
    v' &= \varepsilon T_u(u) F_1(u, T(u); \varepsilon) + \varepsilon(v - T(u)) G_2(u, v; \varepsilon).
\end{align*}
\]

System (15) is a singularly perturbed problem with \(\varepsilon\) as the singular parameter. As usual, the time \(t\) is called the fast time, which is the physical time of our problem.

In terms of the slow time \(\tau = \varepsilon t\), system (15) becomes

\[
\begin{align*}
    \varepsilon \dot{u} &= T(u) - v + \varepsilon F_1(u, v; \varepsilon), \\
    \dot{v} &= T_u(u) F_1(u, T(u); \varepsilon) + (v - T(u)) G_2(u, v; \varepsilon),
\end{align*}
\]

where the overdot symbol indicates the derivative with respect to \(\tau\).

The slow manifold is

\[\mathcal{Z} = \{ v = T(u) \}.\]

On the slow manifold \(\mathcal{Z}\), the flow is given by

\[u' = \varepsilon F_1(u, T(u); \varepsilon).\]

It has a global sink at \(u = u^0\).

We recall that the set \(\mathcal{Z}\) is invariant under system (15) (or equivalently under system (16)) for all \(\varepsilon \geq 0\). This property is crucial in creating oscillations in the system. In fact, one will see later that there is a turning point on \(\mathcal{Z}\) and, due to the invariance of \(\mathcal{Z}\) for all \(\varepsilon\), the turning point causes the delay of stability loss [7, 17, 18, 23, 24, 25, 27]. We believe that the delay of stability loss is one of the most important mechanisms for the oscillation structure in biological population systems.
To describe the delay of stability loss, we define a map \( P : (0, u_0) \to (u_0, u^0) \) via

\[
\int_u^{P(u)} \frac{T_u(\xi)}{F_1(\xi, T(\xi))} \, d\xi = 0.
\]

Also, for any \( v > \min\{T(u) : u \in (0, \infty)\} \), let \( l(v) \) and \( r(v) \) be the two solutions of \( v = T(\tau) \) for \( u \) with \( l(v) < r(v) \) and set \( v^0 = T(u^0) \).

**Proposition 8** (delay of stability loss). Fix \( \delta > 0 \) small, for \( \varepsilon > 0 \) small, let \((u(\tau; \varepsilon), v(\tau; \varepsilon))\) be the solution of system (16) with the initial condition \( (u(0), v(0)) \), where \( u(0) < u_0 \) and \( v(0) = T(u(0)) + \delta \). Let \( \tau(\varepsilon) > 0 \) be the time such that \( v(\tau(\varepsilon); \varepsilon) = T(u(\tau(\varepsilon))) + \delta \). Then, as \( \varepsilon \to 0 \), \( r(v(\tau(\varepsilon))) \to P(l(v(0))) \).

Note that \( P(l(v^0)) < v^0 \), and hence, \( T(l(v^0)) = v^0 > T(P(l(v^0))) \).

**Theorem 9.** For \( \varepsilon > 0 \) small, either the equilibrium \((u_\varepsilon, T(u_\varepsilon))\) is a global attractor of \( \mathcal{R} \) for system (15) or there is a stable periodic relaxation oscillation. Furthermore,

(i) if there exists \( u_1 \in (l(v^0), u_0) \) such that \( T(u_1) < T(P(u_1)) \), then, for \( \varepsilon > 0 \) small, system (15) has a stable periodic relaxation oscillation whose limiting orbit, as \( \varepsilon \to 0 \), is the union of the heteroclinic orbit from \( (P(u^0), T(P(u^0))) \) to \( (u^0, v^0) \) and the curve on \( \{v = T(\tau)\} \) from \( (u^0, T(u^0)) \) to \( (P(u^0), T(P(u^0))) \) for some \( u^0 \in (l(v^0), u_1) \) satisfying \( T(u^0) = T(P(u^0)) \);

(ii) if for every \( u \in (l(v^0), u_0) \), \( T(u) > T(P(u)) \), then, for \( \varepsilon > 0 \) small, the equilibrium \((u_\varepsilon, T(u_\varepsilon))\) is a global attractor of \( \mathcal{R} \) for system (15).

**Proof.** To prove statement (i), note that the unstable manifold \( W^u(u^0, v^0) \) will approach the left branch of the slow manifold \( \{v = T(\tau)\} \) almost horizontally near the set \( \{v = v^0\} \) toward the point \((l(v^0), v^0)\) and then follow the slow orbit through \((l(v^0), v^0)\) up to near the point \((P(l(v^0)), T(P(l(v^0))))\), and leave the slow manifold almost horizontally near the set \( \{v = T(P(l(v^0)))\} \). Due to the fact that \( T(l(v^0)) = v^0 > T(P(l(v^0))) \), upon leaving the slow manifold at near the point \((P(l(v^0)), T(P(l(v^0))))\), the unstable manifold \( W^u(u^0, v^0) \) stays below its initial portion. Therefore, the unstable manifold spirals inward. By the same argument, the existence of \( u_1 \) with the property \( T(u_1) < T(P(u_1)) \) implies that the forward orbit starting from \((u_1 + \delta, T(u_1))\) for some \( \delta > 0 \) small spirals outward. This orbit together with the unstable manifold \( W^u(u^0, v^0) \) encloses a positively invariant region. By the Poincaré–Bendixon theorem, there is a stable periodic orbit. The above argument also shows that between any numbers \( \hat{u}_1, \hat{u}_2 \in (l(v^0), u_0) \) with \( \hat{u}_2 < \hat{u}_1, T(\hat{u}_1) < T(P(\hat{u}_1)) \), and \( T(\hat{u}_2) > T(P(\hat{u}_2)) \), there is a periodic orbit strictly enclosed by the two orbits through, respectively, the points \((\hat{u}_1 + \delta, T(\hat{u}_1))\) and \((\hat{u}_2 + \delta, T(\hat{u}_2))\) for some small \( \delta \). Therefore, the limiting position of a periodic orbit is exactly as described in the statement.

The proof for the statement (ii) follows from the above argument and we will omit the details here. \( \square \)

**4.2. Statement of the main results for system (3).** To translate Theorem 9 in terms of the original system (3), we recall that \( H : (N_0, \infty) \to (0, N_0) \) is the function defined as follows: for \( \varepsilon = 0 \) and for \((dN/(d+p), 0, N) \in \mathbb{Z}_0 \) with \( N > N_0 \), \((dH(N))/(d+p), 0, H(N)) \in \mathbb{Z}_0 \) is the unique equilibrium so that there is a heteroclinic orbit from \((dN/(d+p), 0, N) \) to \((dH(N))/(d+p), 0, H(N)) \). The map \( P \) defined in
(17) is given by \( P : (0, N_0) \to (N_0, \infty) \) by

\[
\int_N^{P(N)} \frac{h(d\xi/(d+p), \xi) - a}{g(\xi)} d\xi = 0.
\]

**Theorem 10.** Let \( H(N) \) and \( P(N) \) be defined as above. For \( \varepsilon > 0 \) small, either the endemic equilibrium \((S_\varepsilon, I_\varepsilon, N_\varepsilon)\) is a global attractor or there is a stable periodic relaxation oscillation. More precisely,

(i) if there exists \( N_1 \in (H(N^*), N_0) \) such that \( N_1 > H(P(N_1)) \), then for \( \varepsilon > 0 \) small, system (3) has a stable periodic relaxation oscillation whose limiting orbit, as \( \varepsilon \to 0 \), is the union of the heteroclinic orbit from the point \((dP(N^c)/(d+p), 0, P(N^c))\) to the point \((dN^c/(d+p), 0, N^c)\) and the segment on \( Z_0 \) from the point \((dN^c/(d+p), 0, N^c)\) to the point \((dP(N^c)/(d+p), 0, P(N^c))\) for some \( N^c \in (H(N^*), N_1) \) satisfying \( N^c = H(P(N^c)) \);

(ii) if for every \( N \in (H(N^*), N_0) \), \( N < H(P(N)) \), then for \( \varepsilon > 0 \) small, the endemic equilibrium \((S_\varepsilon, I_\varepsilon, N_\varepsilon)\) is a global attractor for system (3).

**Proof.** It suffices to show that for \( \varepsilon > 0 \), \( M_\varepsilon \) attracts all solutions except the equilibria \((0,0,0)\) and \((S^*,0,N^*)\). Since \( M_\varepsilon \) has a region attracting orbits on \( M_\varepsilon \) and \( M_\varepsilon \) is normally stable, there is neighborhood \( U \) of \( M_\varepsilon \) independent of \( \varepsilon \) such that for \( \varepsilon > 0 \) small enough, any solution entering \( U \) is attracted by the attracting region on \( M_\varepsilon \). Therefore, we only need to show that any solution will enter \( U \).

\( N'(t) < 0 \) if \( N(t) > N^* \). Thus, all solutions are attracted by the domain \( D \) and the domain \( D \) is positively invariant. It can be verified that \( M_\varepsilon \) attracts all solutions on \( \{t = 0\} \) except \((0,0,0)\) and \((S^*,0,N^*)\). Now, for a solution \((S(t), I(t), N(t))\) with the initial condition \((S(0), I(0), N(0)) \in D\) and \( I(0) > 0 \), by continuity, for \( \varepsilon > 0 \) small independent of the solution starting in \( D \), the solution will approach a point \((\bar{S}, 0, \bar{N}) \in Z_0\) with \( \bar{N} \leq N_0 \) and then follow the slow orbit through \((\bar{S}, 0, \bar{N}) \in Z_0\). Therefore, it enters a neighborhood of \((S_0, 0, N_0)\) and hence into \( U \).

**4.3. Concrete conditions for the existence of relaxation oscillations of system (3).**

**Proposition 11.** The map \( P \) satisfies \( P(N_0) = N_0, P'(N_0) = -1 \), and

\[
P''(N_0) = \frac{4g'(N_0)L - 2g(N_0)Q}{3g(N_0)L},
\]

where \( L \) and \( Q \) are defined in (9).

**Proof.** It follows from the definition of \( P \) that \( P(N_0) = N_0 \). Differentiating with respect to \( N \) on (18) we get

\[
\frac{h(dP(N)/(d+p), P(N)) - a}{g(P(N))} P'(N) = \frac{h(dN/(d+p), N) - a}{g(N)}.
\]

Note that

\[
P(N) = P(N_0) + P'(N_0)(N - N_0) + \frac{1}{2} P''(N_0)(N - N_0)^2 + O(N - N_0)^3
\]

\[
= N_0 + P'(N_0)(N - N_0) + \frac{1}{2} P''(N_0)(N - N_0)^2 + O(N - N_0)^3,
\]
\[ P'(N) = P'(N_0) + P''(N_0)(N - N_0) + O(N - N_0)^2; \]
\[ g(N) = g(N_0) + g'(N_0)(N - N_0) + \frac{1}{2} g''(N_0)(N - N_0)^2 + O(N - N_0)^3; \]
\[ g(P(N)) = g(N_0) + g'(N_0)(P(N) - N_0) + \frac{1}{2} g''(N_0)(P(N) - N_0)^2 \]
\[ = g(N_0) + g'(N_0) \left( P'(N_0)(N - N_0) + \frac{1}{2} P''(N_0)(N - N_0)^2 \right) \]
\[ + \frac{1}{2} g''(N_0)(P'(N_0))^2(N - N_0)^2 + O(N - N_0)^3; \]

and

\[ h(dN/(d + p), N) - a = L(N - N_0) + \frac{1}{2} Q(N - N_0)^2 + O(N - N_0)^3; \]
\[ h(dP(N)/(d + p), P(N)) - a = L(P(N) - N_0) + \frac{1}{2} Q(P(N) - N_0)^2 + O(N - N_0)^3 \]
\[ = L \left( P'(N_0)(N - N_0) + \frac{1}{2} P''(N_0)(N - N_0)^2 \right) \]
\[ + \frac{1}{2} Q(P'(N_0))^2(N - N_0)^2 + O(N - N_0)^3. \]

Substituting these expansions into (19) and comparing the terms of like powers in \((N - N_0)\) we get

for \(N - N_0\), \(gL(P')^2 = gL \implies P' = -1,\)
for \((N - N_0)^2\), \(-\frac{1}{2} g(LP'' + Q) + g'L - gL'P'' = -g'L + \frac{1}{2} gQ\)
\[ \implies P'' = \frac{4g'L - 2gQ}{3gL}. \]

This completes the proof. \(\square\)

Combining Propositions 7 and 11 we obtain the following result.

**Proposition 12.** The function \(\bar{F} = H \circ P\) satisfies \(\bar{F}(N_0) = N_0, \bar{F}'(N_0) = 1,\)

and

\[ \bar{F}''(N_0) = H''(N_0) - P''(N_0) = \frac{2\Delta_0}{(d + p)g(N_0)} + \frac{2}{3} \frac{g'(N_0)L + g(N_0)Q}{g(N_0)L}, \]

where \(\Delta_0\) is defined in (4), and \(L\) and \(Q\) are defined in (9).

As a direct consequence of Theorem 10 and Proposition 12, we have the following.

**Corollary 13.** If \(\bar{F}''(N_0) < 0,\) then, for \(\varepsilon > 0\) small, there is at least one stable relaxation oscillation.

**Example.** We establish the existence of a stable relaxation oscillation in the case that \(E_\varepsilon\) is stable. More precisely, we take a special case of \(h\) that is biologically plausible and show that for any \(g\) satisfying (A1), there are parameter ranges for \(\beta\) and \(K,\) dependent on all other fixed parameters so that \(\Delta_0 > 0\) in Theorem 4, for which the equilibrium \(E_\varepsilon\) is stable and \(\bar{F}''(N_0) < 0\) holds. This guarantees the existence of a stable relaxation oscillation. Thus, a stable relaxation oscillation may exist even when the equilibrium \(E_\varepsilon\) is stable. In this case, there exists at least an unstable periodic
orbit between the stable relaxation and the equilibrium. In general, the unstable periodic orbit is not necessarily a relaxation oscillation but a small periodic orbit through a subcritical Hopf bifurcation.

Consider \( h(S,N) = \frac{\beta S}{K+S} \) with \( \beta > a \); it can be verified that

\[
S_0 = \frac{aK}{\beta - a} > 0, \quad N_0 = \frac{d + p}{d} \frac{aK}{\beta - a},
\]

\[
L = \frac{d}{d + p} \frac{\beta K}{(K + S_0)^2}, \quad Q = -\frac{d^2}{(d + p)^2 (K + S_0)^3},
\]

\[
\Delta_0 = \left( \frac{a}{\alpha} - \frac{d}{d + p} \right) \frac{\beta K}{(K + S_0)^2} g(N_0) - (d + p)g_N(N_0),
\]

\[
\bar{F}''(N_0) = \frac{2\Delta_0}{(d + p)g(N_0)} + \frac{2g'(N_0)L + g(N_0)Q}{3} \frac{\beta K}{g(N_0)L}
\]

\[
= \frac{2}{(d + p)g(N_0)} \left( \left( \frac{a}{\alpha} - \frac{d}{d + p} \right) \frac{\beta K}{(K + S_0)^2} g(N_0)
\right.
\]

\[
- \frac{2}{3} \left( d + p \right) g_N(N_0) - \frac{2d}{3} \frac{K + S_0}{(K + S_0)^2} g(N_0).
\]

We also note that

\[
\left( \frac{a}{\alpha} - \frac{d}{d + p} \right) \frac{\beta K}{(K + S_0)^2} - \frac{2d}{3} \frac{K + S_0}{(K + S_0)^2} = \left( \frac{a}{\alpha} - \frac{d}{d + p} - \frac{2d}{3(\beta - a)} \right) \frac{\beta K}{(K + S_0)^2}.
\]

Choose \( \beta^* > a \) such that

\[
\frac{a}{\alpha} - \frac{d}{d + p} - \frac{2d}{3(\beta^* - a)} < 0,
\]

and choose \( K^* \) such that for \( N_0 = N_0^* = \frac{d + p}{d} \frac{aK^*}{\beta^* - a} \), \( g_N(N_0^*) = 0 \) holds. Then

\[
\Delta_0 = \left( \frac{a}{\alpha} - \frac{d}{d + p} \right) \frac{\beta^* K^*}{(K^* + S_0)^2} g(N_0) > 0,
\]

\[
\bar{F}''(N_0^*) = \frac{2}{d + p} \left( \frac{a}{\alpha} - \frac{d}{d + p} - \frac{2d}{3(\beta - a)} \right) \frac{\beta K^*}{(K^* + S_0)^2} < 0.
\]

This accomplishes the goal of this example.

We note that the construction of the above example strongly indicates that it may not be rare to have stable relaxation oscillations when the endemic equilibrium \( E_c \) is stable. It is also possible to give a more detailed analysis, for fixed forms of \( h \) and \( g \), on the parameter ranges for such coexistence of stable structures. It may reveal a more comprehensive understanding of the global dynamics of this model.

5. Numerical simulations and biological interpretations. In this section, we provide results from numerical simulations of model (3) that demonstrate and support our theoretical results on the existence of stable periodic solutions of relaxation oscillation type. Unless otherwise stated, we choose

\[
g(N) = N \left( 1 - \frac{N}{N^*} \right) \quad \text{and} \quad h(S,N) = \frac{\beta S}{K+S}.
\]

It can be verified that \( g(N) \) and \( h(S,N) \) satisfy assumptions (A1), (A2), and (A3).
5.1. Existence of relaxation oscillations.

Case 1. Existence of relaxation oscillation when $E_\varepsilon$ is unstable. Choose $d = 0.2$, $p = 0.01$, $\alpha = 0.048$, $\beta = 1$, $\gamma = 0.75$, $K = 0.1$, $\varepsilon = 10^{-4}$, and $N^* = 400$. The endemic equilibrium $E_\varepsilon = (49.9, 0.09555, 52.8489)$ is unstable. In Figure 4, we show that a trajectory starting from $(35, 0.09555, 67)$ approaches a stable relaxation oscillation cycle with IEP $5.6 \times 10^4$.

Case 2. Existence of relaxation oscillation when $E_\varepsilon$ is stable. Choose $d = 0.2$, $p = 0.01$, $\alpha = 0.049$, $\beta = 1$, $\gamma = 0.75$, $K = 0.1$, $\varepsilon = 10^{-4}$, and $N^* = 380$. In Figure 5, we show that a trajectory starting from $(197, 1.47, 204.4)$ approaches a stable relaxation oscillation cycle. We modified the function $h(S, N)$ in a small neighborhood of $E_\varepsilon$ so that it becomes locally asymptotically stable. Such modification does not change the relaxation oscillation cycle since it is far away from $E_\varepsilon$. A trajectory starting from $(150, 1, 160)$ is shown in Figure 5 to approach the stable equilibrium $E_\varepsilon$. We note that there should be a second periodic orbit that is unstable (not shown in Figure 5).

5.2. Dependence of IEP on physical parameters.

1. Dependence of IEP on the intrinsic growth rate $\varepsilon$.

We demonstrate using numerical evidence that the IEP is of order $1/\varepsilon$. We choose $d = 0.2$, $p = 0.01$, $\alpha = 0.048$, $\beta = 1$, $\gamma = 0.75$, $K = 0.1$, and $N^* = 400$, and we vary...
Numerical simulations show the existence of a stable periodic solution when the endemic equilibrium is stable. An oscillatory orbit with a large amplitude is shown to converge to a stable relaxation oscillation cycle, and an orbit with a smaller amplitude converges to the stable endemic equilibrium $E_\varepsilon$.

The IEP increases as the intrinsic growth rate $\varepsilon$ decreases in (a), and the IEP is in proportion to $1/\varepsilon$ in (b).

The disease is in the IEP if the number of the infected individuals is less than $10^{-7}$. Plots of IEP against the values of $\varepsilon$ and $1/\varepsilon$ are shown in Figure 6.

2. Dependence of IEP on parameters $\alpha$ and $\beta$.

In Figure 7, we show that the IEP decreases as the transmission coefficient $\beta$ increases, and the IEP increases as the rate $\alpha$ of disease-caused death increases. For the simulations, we choose $d = 0.2$, $p = 0.01$, $\gamma = 0.75$, $K = 0.1$, $N^* = 400$, and $\varepsilon = 10^{-4}$ and vary values of $\beta$ when $\alpha = 0.048$ in Figure 7(a) or vary the values of $\alpha$ when $\beta = 1$ in Figure 7(b).
Fig. 7. Dependence of IEP on the transmission coefficient $\beta$ and on the rate of disease-caused death $\alpha$.

Appendix I: Technical proofs.

Proof of Theorem 4. To show (i), note that the linearization of system (3) at $(0,0,0)$ is

$$J(0,0,0) = \begin{pmatrix} -(d+p) & 0 & d + \varepsilon g_N(0) \\ 0 & -a & 0 \\ 0 & -\alpha & \varepsilon g_N(0) \end{pmatrix},$$

whose eigenvalues are $-(d+p) < 0$, $-a < 0$, $\varepsilon g_N(0) > 0$, where $\varepsilon g_N(0) > 0$ follows from (A1). Similarly, the linearization at $(S^*,0,N^*)$ is

$$J(S^*,0,N^*) = \begin{pmatrix} -(d+p) & -h & d + \varepsilon g_N(N^*) \\ 0 & h-a & 0 \\ 0 & -\alpha & \varepsilon g_N(N^*) \end{pmatrix},$$

with eigenvalues $-(d+p) < 0$, $h(S^*,N^*)-a > 0$, and $\varepsilon g_N(N^*) < 0$, where $h(S^*,N^*)-a > 0$ follows from (A3) and $\varepsilon g_N(N^*) < 0$ follows from (A1).

The linearization at $E_{\varepsilon}$ is

$$J = J(S_{\varepsilon},I_{\varepsilon},N_{\varepsilon}) = \begin{pmatrix} -(d+p + hSI_{\varepsilon}) & -a & d - hNI_{\varepsilon} + \varepsilon g_N \\ hSI_{\varepsilon} & 0 & hNI_{\varepsilon} \\ 0 & -\alpha & \varepsilon g_N \end{pmatrix},$$

whose characteristic polynomial is given by

$$P_{\varepsilon}(\lambda) = \lambda^3 + \left(d + p + \frac{\varepsilon hSG}{\alpha} - \varepsilon g_N\right)\lambda^2 - \varepsilon \left((d+p)g_N - \frac{ahSG}{\alpha} - hNg - hSGNI_{\varepsilon}\right)\lambda + ahdSI_{\varepsilon} + \alpha(d+p)hNI_{\varepsilon} - \varepsilon(a + \alpha)hSGNI_{\varepsilon}.$$

Hence,

$$\text{tr}(J) = -(d+p) - \frac{\varepsilon hSG}{\alpha} + \varepsilon g_N < 0,$$

$$\text{det}(J) = -ahdSI_{\varepsilon} - \alpha(d+p)hNI_{\varepsilon} + \varepsilon(a + \alpha)hSGNI_{\varepsilon}. $$
where \( a_2 \) is the coefficient of \( \lambda \) in \( P_\varepsilon(\lambda) \), namely, the sum of all \( 2 \times 2 \) principal minors of \( J \).

When \( \varepsilon = 0 \), \( P_0(\lambda) = \lambda^3 + (d + p)\lambda^2 \). It has a negative root, \(-(d + p)\). Therefore, when \( \varepsilon > 0 \) small, \( P_\varepsilon(\lambda) \) has a negative root. We show that the remaining roots of \( P_\varepsilon(\lambda) \) are always complex conjugates. To see this, write \( P_\varepsilon(\lambda) = \lambda^3 - a_1\lambda^2 + a_2\lambda - a_3 \), where \( a_1 = \text{tr}(A) < 0 \), \( a_3 = \text{det}(A) < 0 \), and \( a_2 \) is as above. The larger of the two critical points of \( P_\varepsilon(\lambda) \) is

\[
\lambda_1 = \frac{1}{3}(a_1 + \sqrt{a_1^2 - 3a_2}).
\]

Straightforward calculation leads to

\[
P_\varepsilon(\lambda_1) = \frac{1}{27} \left[ -2a_1^3 + 9a_1a_2 - 27a_3 - 2(a_1^2 - 3a_2)\sqrt{a_1^2 - 3a_2} \right].
\]

It can be verified that \( P_\varepsilon(\lambda_1) > 0 \) if and only if

\[
27a_3^2 + 4a_1^2a_3 + 4a_2^3 - 18a_1a_2a_3 - a_1^2a_2^2 > 0.
\]

When \( \varepsilon = 0 \), we have \( \lambda_1 = 0 \), which is a double root of \( P_0(\lambda) \). Therefore, \( P_\varepsilon(\lambda_1) = 0 \) when \( \varepsilon = 0 \), and thus the sign of the expression in (21) is determined by the \( \varepsilon \) order terms, which is given by

\[
4(d + p)^3g(N_0) \left[ \frac{d}{d + p}h_S(S_0, N_0) + h_N(S_0, N_0) \right] \varepsilon > 0,
\]

by assumption (A3) and continuity. Hence \( P_\varepsilon(\lambda_1) > 0 \). This implies that \( P_\varepsilon(\lambda) \) has only one real root. The signs of the real parts of the complex roots can be determined by the Routh–Hurwitz conditions, which state that all roots of \( P_\varepsilon(\lambda) \) have negative real parts if and only if the following three conditions hold: \( a_1 = \text{tr}(A) < 0 \), \( a_3 = \text{det}(A) < 0 \), and \( a_1a_2 - a_3 < 0 \). From relations in (20), we see that for \( \varepsilon > 0 \) small, if \( \Delta_0 > 0 \), then all three eigenvalues have negative real parts, and if \( \Delta_0 < 0 \), then at least one eigenvalue has positive real parts. This establishes (ii).

**Proof of Theorem 6.** To show (i), we note that for solution \((S(t), I(t), N(t))\) with initial condition \((S, 0, N) \in \mathcal{D}, I(t) \equiv 0, N(t) \equiv N\), and \(S(t) \to dN/(d+p)\) as \(t \to \infty\). Thus, \((S(t), I(t), N(t)) \to (dN/(d+p), 0, N)\) as \(t \to \infty\). Now let \((S(t), I(t), N(t))\) be the solution with the initial condition \((S(0), I(0), N(0)) \in \mathcal{D}\) with \(I(0) > 0\). From system (5), we have \(I(t) > 0\) for all \(t \geq 0\) and hence \(N(t)\) monotonically decreases. Therefore \(N(t) \to N_\ast\) as \(t \to \infty\) for some \(N_\ast\) dependent on the initial condition. We claim that \(N_\ast \leq N_0\).

First of all, note that the equilibrium \((S_0, 0, N_0)\) has two zero eigenvalues and one negative eigenvalue \(-(d + p)\). Locally, there is a two-dimensional center manifold
to a point (\(\bar{\alpha}<\alpha\)).

Since feasible region \(D\) is a two-dimensional stable manifold with positive one-to-one function. Therefore, the monotone decreasing property of \(W\) will be established directly that the eigenvectors associated to the stable eigenvalues \(H\) and \(\lambda_2\) are, respectively,

\[
v_2 = (1, 0, 0) \quad \text{and} \quad v_3 = \left(\frac{\alpha d}{a(d + p - a)}, 1, \frac{\alpha}{a}\right).
\]

Since \(\alpha < a\), the vectors \(v_2\) and \(v_3\) at \((0, 0, 0)\) are pointing toward the exterior of the feasible region \(D\). Therefore, the \textit{local} two-dimensional stable manifold \(W_{\text{loc}}^s(0, 0, 0)\) except \((0, 0, 0)\) stays outside of \(D\). By continuity, for some \(\delta > 0\) small and for any equilibrium \((dN)/(d + p), 0, N)\) with \(N < \delta\), an orbit starting on the local stable manifold \(W_{\text{loc}}^s(dN)/(d + p), 0, N)\) will exit the region \(D\) backward and will stay outside \(D\) in backward time upon the exit due to the positive invariance of \(D\). Hence, \(H(N) \geq \delta\) for any \(N > N_0\), which implies that \(N_{\infty} \geq \delta > 0\).
Appendix II: Persistence of $M(Z_0)$ for $\varepsilon > 0$ small. To establish the persistence of $M(Z_0)$ claimed in Case 2 of section 3.2, we make a change of variables. This change of variables is continuous but not everywhere smooth. Indeed, it is smooth everywhere except on $\{I = 0\}$. Nevertheless, the property that $\{I = 0\}$ is invariant for all $\varepsilon \geq 0$ makes the change of variables work.

Let $m$ be a positive integer so that $a < m(d + p)$. We may assume that $m \geq 2$. Make the change of state variables: $S = x$, $I = y^m$, and $N = N$ for $y > 0$. In terms of the new variables $(x, y, N)$, the equation for $I$ in (3) becomes

$$my^{m-1}y' = (h(x, N) - a)y^m \text{ or, equivalently, } y' = \frac{1}{m}(h(x, N) - a)y.$$ 

The model (3) becomes

$$x' = dN + \varepsilon g(N) - sh(x, N)y^m - dx - px,$$

$$y' = \frac{1}{m}(h(x, N) - a)y,$$

$$N' = \varepsilon g(N) - \alpha y^m. \tag{22}$$

We note that this change of state variables is smooth for $y > 0$ and can be continued to $y = 0$. The new system (22) has exactly the same reduced dynamics on $\{y = 0\}$ as that of (3) on $\{I = 0\}$. We emphasize that the naturally given property that $\{I = 0\}$ is invariant under (3) for $\varepsilon \geq 0$ is crucial for such a change of variables. The biological implications are commented on and illustrated by examples in section 5.

Recall that $m \geq 2$. The set $Z_0$ corresponds, for (22), to

$$S_0 = \left\{ y = 0, x = \frac{d}{d + p}N \right\}.$$ 

Let $M(S_0)$ denote the corresponding invariant manifold $M(Z_0)$.

The linearization at each equilibrium on $S_0$ is

$$\begin{pmatrix}
-(d + p) & 0 & d \\
0 & \frac{1}{m}(h - a) & 0 \\
0 & 0 & 0
\end{pmatrix}$$

with eigenvalues $\lambda_1 = 0$, $\lambda_2 = -(d + p)$, and $\lambda_3 = (b(dN/(d + p), N) - a)/m$. The eigenvector $v_1$ associated with $\lambda_1$ is tangent to $S_0$ and $v_2$ associated with $\lambda_2$ is $(1, 0, 0)$, and $v_1$ and $v_2$ span the plane $\{y = 0\}$. The eigenvector $v_3$ associated with $\lambda_3$ is transversal to the plane $\{y = 0\}$. While the eigenvalue $\lambda_2$ stays negative, the eigenvalue $\lambda_3$ changes sign across $(S_0, 0, N_0) \in S_0$. Nevertheless, $\lambda_1 > \lambda_2$ and $\lambda_3 > \lambda_2$. The center manifold theory in [5, 6] implies that $M(S_0)$ persists under system (22) for $\varepsilon > 0$ small.

REFERENCES


