J. Math. Biology 10, 13 – 32 (1980)

Journal of
Mathematical
Biology

The Effect of Integral Conditions in Certain Equations Modelling Epidemics and Population Growth*

Stavros Busenberg[1] and Kenneth L. Cooke[2]

[1] Harvey Mudd College, Claremont, CA 91711, USA
[2] Pomona College, Claremont, CA 91711, USA

Summary. Models of epidemics that lead to delay differential equations often have subsidiary integral conditions that are imposed by the interpretation of these models. The neglect of these conditions may lead to solutions that behave in a radically different manner from solutions restricted to obey them. Examples are given of such behavior, including cases where periodic solutions may occur off the natural set defined by these conditions but not on it. A complete stability analysis is also given of a new model of a disease propagated by a vector where these integral conditions play an important role.

Key words: Delay differential equations – Epidemic models – Invariant integrals.

1. Introduction

In a certain class of models of epidemics and population growth that lead to delay differential equations one encounters natural integral conditions that are dictated by the interpretation of the model. For example, a population model studied by Cooke and Yorke [3] leads to the delay differential equation

$$x'(t) = g[x(t)] - g[x(t - L)].$$

Here, $x(t)$ is the population size at time t, $g[x(t)]$ is the rate of addition to the population at time t, and since it is assumed that all individuals have a constant length of life L, $g[x(t - L)]$ is the rate of removal at time t. This equation has a first integral

$$x(t) = a + \int_{t-L}^{t} g[x(s)]\, ds,$$

with a an arbitrary constant. However, for the correct interpretation of the model one must take $a = 0$, since the size of the population is equal to the total number of those that were born before time t and have not passed away by time t. The choice

* This work was partially supported by N.S.F. Grant MCS 7903497

0303 – 6812/80/0010/0013/$04.00

$a = 0$ gives the proper integral condition in this example. A major purpose of the present paper is to show some of the effects of such conditions and to provide a number of guiding examples that illustrate the main types of conditions that are mandated by models of epidemics. Most of the examples used here are taken from models that have previously been studied in the literature but often without special attention being given to the proper integral conditions. However, a new model of a disease propagated by a vector is also studied, and the role of the invariant integral conditions is brought out in detail in this case.

The paper is separated into two parts. The first deals with a new model of a disease propagated by a vector in which there is a delay between the initial time of exposure to the disease and the onset of infectiousness. A stability analysis is given for this model and the effect of the integral conditions is discussed. In the second part of the paper, several examples dealing with models already studied in the literature are discussed. The emphasis here is on pointing out the proper integral conditions for these models and on studying the effects of the imposition of these conditions. Finally, the relation between the present approach and an alternate formulation described by Hoppensteadt [11, pp. 47–49] is also discussed. The paper is written so as to allow the two parts to be read independently without loss of coherence. In particular, readers who are not interested in the details of the vector propagated disease model, after perusing section two up to the start of the proof of Theorem 1, may proceed to the third section without loss of continuity.

The first use in the literature on epidemic models of the type of invariant integral condition that concerns us seems to be in the already quoted work of Cooke and Yorke [3]. A subsequent paper of Hale [8] dealt in detail with the behavior of solutions which satisfy such conditions, and in particular studied one of the equations in [3] as a special case. However, there does not seem to be a wide recognition of the basic nature of these conditions and a number of studies in the literature seem to totally neglect them. One of the objects of the present paper is to try to convince the reader that these conditions are often needed in the proper interpretation of mathematical results arising from some models. Most of these models can be formulated from basic principles as integral equations incorporating these side conditions. If these integral equations are studied directly, then the difficulties associated with invariant integral conditions in a delay differential equation model can be avoided. However, there are advantages to the delay differential equation model when one uses Lyapunov functional techniques because most of the literature on such techniques is devoted to the differential equation case. These techniques are often useful in establishing global stability results, and the availability of a well posed delay differential equation version of the model can be valuable.

2. A Disease Propagated by a Vector

Consider a disease propagated by a vector affecting a population of constant size N, and assume the following governing hypotheses.

(a) The disease is not lethal and it imparts no immunity.

(b) The vector population is large, and the number of exposed vectors is directly proportional to the number of infectious persons. This supposes homogeneous mixing of the vector and human populations.

(c) There is a delay $T > 0$ between the time that a person is exposed to the disease and the time that the person becomes infectious.

(d) The infective individuals are cured and return to the susceptible pool at a rate proportional to their number.

The population is partitioned into three classes: those who are susceptible, those who have been exposed, and finally, those who are infectious. Denote by y, z, and x, respectively, the number of individuals in these classes normalized by dividing by the constant total size of the population. Then,

$$x + y + z = 1,$$

and by hypothesis (b) the number of infected vectors is equal to $b_1 x(t)$, $b_1 > 0$. If the variables x, y and z are treated as being continuously differentiable in the time variable t, the following equations describe the dynamics of the above model for $t > 0$.

$$\begin{aligned} \frac{dx(t)}{dt} &= bx(t-T)y(t-T) - cx(t), \\ \frac{dy(t)}{dt} &= cx(t) - bx(t)y(t), \\ \frac{dz(t)}{dt} &= b[x(t)y(t) - x(t-T)y(t-T)], \end{aligned} \tag{1}$$

where $b = b_1 b_2$, and $b_2 > 0$ is the contact proportionality constant between vectors and susceptibles.

It appears from the form of equations (1) that the first two of these may be solved independently of the third, and then the fact that $z = 1 - x - y$ can be used to find the proportion of the population in the exposed class. That this is not the case becomes clear when we note that the solutions of (1) that are significant in the light of the model are those where $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$, $0 \leqslant z \leqslant 1$ and $x + y + z = 1$. Moreover, all solutions initially obeying the above restrictions must do so for all subsequent time. However, we shall show in the sequel that, if $b(1 + bT/4) \leqslant c$, the solution $(x(t), y(t), z(t))$ with initial data $(\frac{1}{2}, \frac{1}{2}, 0)$ on $t \in [-T, 0]$ approaches, as $t \to \infty$, the constant solution $(0, 1 + bT/4, 0)$, thus violating the above restrictions. In this sense, the problem consisting of (1) together with the initial condition that x, y and z lie between zero and one and that they add up to one initially, is not well posed.

In order to remedy this situation we introduce the following restriction on the solutions of (1):

$$x(t) + y(t) = 1 - b\int_{t-T}^{t} x(s)y(s)\,ds. \tag{2}$$

The epidemiological interpretation of this condition is that at any time $t \geqslant 0$, the proportion of individuals in the exposed class should equal those who entered this class in the past and have not yet left it, that is, $z(t) = b\int_{t-T}^{t} x(s)y(s)\,ds$. This restriction plus the requirement that $x + y + z = 1$, yields the condition (2). From the mathematical viewpoint, (2) arises out of the observation that the variable $\omega(t) = x(t) + y(t)$ obeys the differential equation $d\omega(t)/dt = b[x(t-T)y(t-T)$

$-x(t)y(t)]$ which has a first integral $x(t) + y(t) = \omega(t) = a - b\int_{t-T}^{t} x(s)y(s)\,ds$, where a is an arbitrary constant. The choice of the constant $a = 1$ is again dictated by the model. In order to bring out more clearly the effect of the integral condition (2) on our model, we rewrite this equation in the more general form

$$x(t) + y(t) = a - b\int_{t-T}^{t} x(s)y(s)\,ds. \tag{2a}$$

We take $a \geqslant 0$, since in the contrary case $x + y < 0$, a situation that is ruled out by the interpretation of the variables x and y.

An alternative formulation of the model which also avoids the above difficulty uses integral equations for $t \leqslant T$ and the equations (1) for $t \geqslant T$. This approach is described by Hoppensteadt [11, pp. 47–49] and its relation to the present one will be discussed at the end of the next section.

We are now prepared to give our main result on this model where the value of the constant a should be taken equal to one.

Theorem 1. *Let $x_0(t)$, $y_0(t)$ be continuous functions on $[-T, 0]$, satisfying $0 \leqslant x_0(t) \leqslant a$, $0 \leqslant y_0(t) \leqslant a$ for $t \in [-T, 0]$, and*

$$x_0(0) + y_0(0) = a - b\int_{-T}^{0} x_0(s)y_0(s)\,ds,$$

i.e., equation (2a) *at $t = 0$. Then, if $(x(t), y(t))$ is the solution of* (1) *with initial data $(x(t), y(t)) = (x_0(t), y_0(t))$ for $t \in [-T, 0]$, the following holds*:

(I) *$0 \leqslant x(t) \leqslant a$, $0 \leqslant y(t) \leqslant a$ and $(x(t), y(t))$ exists for all $t \geqslant 0$ and satisfies* (2a) *on $t \geqslant 0$.*

(II) *If $0 < ab \leqslant c$, then all solutions of* (1) *lying on the positively invariant set $G_a = \{(x, y)$: with x, y continuous on $t \geqslant 0$; $x \geqslant 0$, $y \geqslant 0$ and (x, y) satisfying* (2a) *on $t \geqslant 0\}$ approach $(0, a)$ as $t \to \infty$.*

(III) *If $0 < c < ab$, then the solution $(0, a)$ is unstable and all other solutions $(x(t), y(t)) \not\equiv (0, a)$ that lie on G_a approach the solution $([ab - c]/b(1 + cT), c/b)$ as $t \to \infty$.*

Before embarking on the proof of this theorem we shall point out its epidemiological implications, paying attention to the effects of the integral condition on the interpretation of the mathematical result. First, if we take $a = 1$, as required by the situation that we are modelling, the above results lead to the following conclusions. All initial conditions that obey the natural restrictions (2) and

$$0 \leqslant x \leqslant 1, \qquad 0 \leqslant y \leqslant 1, \qquad 0 \leqslant z \leqslant 1, \qquad 0 \leqslant x + y + z \leqslant 1, \tag{3}$$

lead to solutions that obey (3) for all $t \geqslant 0$. Moreover, if the cure rate c is large enough: $c \geqslant b > 0$, all such solutions tend to the constant solution $(x, y, z) = (0, 1, 0)$, that is, to a state where the epidemic dies out completely. If, however, the cure rate is smaller than a fixed threshold: $0 < c < b$, then for all solutions that obey (3) initially, with the exception of the solution $(x, y, z) \equiv (0, 1, 0)$, the proportion of susceptibles tends to the constant level $c/b < 1$.

Thus, if $c < b$, there exists a nontrivial endemic level which is approached in all cases except when the population starts totally free of the infection. The endemic level is $(1 - c/b)\tau/(\tau + T)$ where $\tau = 1/c$ is the average length of infectiousness, and $\tau/(\tau + T)$ is the ratio of infectious time to infectious plus incubation time.

We now turn to the effects of the integral condition (2). If (3) is imposed as a condition on the initial data, but (2) is not; for example, if initially $(x, y, z) = (\frac{1}{2}, \frac{1}{2}, 0)$, then (2a) is satisfied at $t = 0$ with $a = 1 + bT/4 > 1$. In this case, our result shows that if $b \leqslant c < b(1 + bT/4) = ba$, the proportion of susceptibles will tend to the constant endemic level c/b, rather than to one as in the case when $c \geqslant b$ and the epidemiologically correct condition G_1 defined by (2) is imposed. If, however, $a < 1$, as is the case if the initial data for (x, y, z) equals $(\frac{1}{3}, 0, \frac{2}{3})$ (here $a = \frac{1}{3}$ from the relation (2a)), the relation between the solutions on G_1 and the solution starting with this initial condition is reversed whenever $ab = b/3 \leqslant c < b$. Here, all solutions on G_1 yield a proportion y of susceptibles that tends to c/b, while the above data yield a solution that has the susceptibles tending to one. Of course, if $c = b/3$, the correct limiting value of the proportion of susceptibles is equal to $\frac{1}{3}$, which is far from the limiting value one for the susceptibles in the solution starting at $(\frac{1}{3}, 0, \frac{2}{3})$.

It may be pertinent to point out the importance of the integral condition in numerical integrations of equations (1). Suppose that one wishes to study the biologically interesting case $a = 1$. However, one does not ordinarily know initial conditions x_0, y_0, z_0 that would be biologically realistic. If x_0, y_0, z_0 are chosen to lie between 0 and 1, but if (2) is not satisfied at $t = 0$ (i.e., (x_0, y_0) are not in G_1) then it is automatically true that (2a) is satisfied at $t = 0$ for some value $a \neq 1$. By Theorem 1, the resulting solutions must satisfy (2a) for all $t \geqslant 0$ and the limiting behavior of such solutions, as indicated above, depends on a. Thus numerical integration of the delay differential equations without attention to the integral condition would lead to incorrect limiting values; or even to the erroneous conclusion that different initial conditions may result in different equilibrium values. In the case of more complicated models, for which rigorous analysis may be impossible, it will be particularly important to take account of the appropriate integral conditions before numerical integration is performed.

At this point, those who are not interested in further details on this epidemic model may, without loss of coherence, proceed to section three. We note however, that following the proof of Theorem 1 we discuss a model that has the above model as well as that discussed in [4] as special subcases.

We next give the proof of Theorem 1. The proof breaks up into two parts. The first gives the invariance of the condition (2a), while the second part deals with the stability analysis of the constant solutions of the differential difference equations.

Proof of Theorem 1. From (1) we see that for $t > 0$,

$$\frac{d}{dt}[x(t) + y(t)] = b[x(t - T)y(t - T) - x(t)y(t)],$$

$$x(t) + y(t) = x(0) + y(0) + b \int_0^t [x(s - T)y(s - T) - x(s)y(s)]\, ds$$

$$= x(0) + y(0) + b\int_{-T}^{0} x(s)y(s)\,ds + b\int_{0}^{t-T} x(s)y(s)\,ds$$

$$- b\int_{0}^{t} x(s)y(s)\,ds$$

$$= a - b\int_{t-T}^{t} x(s)y(s)\,ds,$$

since (2a) is assumed to hold at $t = 0$. So, any solutions which initially satisfy (2a) at $t = 0$ must satisfy (2a) for all $t \geqslant 0$.

Now, from (1) we also have

$$x(t) = e^{-ct}x(0) + b\int_{0}^{t} e^{c(s-t)}x(s-T)y(s-T)\,ds,$$

$$y(t) = y(0)e^{h(t)} + c\int_{0}^{t} e^{h(t)-h(s)}x(s)\,ds, \qquad h(t) = -b\int_{0}^{t} x(s)\,ds.$$

Since $0 \leqslant x(s)$, $0 \leqslant y(s)$ for $s \in [-T, 0]$, these equations imply that $x(t) \geqslant 0$ for all t in $[0, T]$ for which it exists. Now, the expression for $y(t)$ implies that $y(t) \geqslant 0$ for the same values of $t \in [0, T]$. But, $x(t)$, $y(t)$ being non-negative implies from (2a) that they are bounded by a. So, they exist on all of $[0, T]$, and from the above argument, are non-negative on this interval. Repeating this reasoning we see that the same conclusion holds on $[0, kT]$ provided $x(t)$ and $y(t)$ exist and are non-negative on $[0, (k-1)T]$. By induction, $0 \leqslant x(t) \leqslant a$, $0 \leqslant y(t) \leqslant a$ for all $t \geqslant 0$. This completes the proof of (I).

We note, parenthetically, that if $x(t^*) > 0$, $y(t^*) > 0$ for some $t^* \geqslant 0$, then the above argument can be adapted to show that $x(t) > 0$, $y(t) > 0$ for all $t \geqslant t^*$.

Next, we consider the stability of the constant solution $(0, a)$. Letting $u = x$, $v + a = y$ in (1), we get the following equations for u and v.

$$\frac{du(t)}{dt} = abu(t-T) - cu(t) + bu(t-T)v(t-T),$$

$$\frac{dv(t)}{dt} = (c - ab)u(t) - bu(t)v(t). \tag{4}$$

The characteristic quasipolynomial (see [1], pp. 99–102) of the linear part of the equations is $p(\lambda) = \lambda(\lambda - abe^{-\lambda T} + c)$. If $ab \neq c$, $\lambda = 0$ is a simple root of p, and letting $\lambda = \alpha + i\beta$ we see that the other roots must satisfy

$$\alpha + c - abe^{-\alpha T}\cos\beta T = 0, \qquad \beta + abe^{-\alpha T}\sin\beta T = 0.$$

If $ab < c$ the first of these implies that $\alpha < 0$. Now, if $\alpha = 0$ and $ab = c$ we get $\cos\beta T = 1$, hence $\sin\beta T = 0$, and the second of the above relations gives $\beta = 0$. So, when $ab = c$, $\lambda = 0$ is a root and is the only root on the imaginary axis. Moreover, at $\lambda = 0$, $(d/d\lambda)[\lambda - abe^{-\lambda T} + c] = 1 + abT > 0$, so $\lambda = 0$ is a simple root of this factor. So, if $ab = c$ there is a double root of p at $\lambda = 0$, and all other roots have negative real parts. We note that the same result can be obtained by using Hayes' theorem (see [1], p. 444). Now, if $ab = c$, the linearized equations for u and v admit only the arbitrary constant solutions (x^0, y^0) and, hence, the eigenspace corresponding to $\lambda = 0$ is two-dimensional. From all this we can conclude that the zero solution of the linearization of (1) about $(0, a)$ is stable if $ab \leqslant c$. Of course, since $\lambda = 0$ is an eigenvalue, this does not imply that the solution $(0, a)$ of the nonlinear equations is stable, and the stability analysis will be pursued via Lyapunov methods.

Next, if $ab > c$, and if we seek a solution $\lambda = \alpha + i0$ of $p(\lambda) = 0$, we get $\alpha = abe^{-\alpha T} - c$, which always has a solution $\alpha > 0$. So, in this case the constant solution $(0, a)$ of (1) is unstable.

In order to study the local stability of the solution $([ab - c]/b(1 + cT), c/b)$ set $u + x^0 = x$, $v + c/b = y$, $x^0 = [ab - c]/b(1 + cT)$, to get the following equations for u and v:

$$\frac{du(t)}{dt} = c[u(t - T) - u(t)] + bx^0 v(t - T) + bu(t - T)v(t - T),$$

$$\frac{dv(t)}{dt} = -bx^0 v(t) - bu(t)v(t). \tag{5}$$

The characteristic quasipolynomial for the linear part of this system is $p(\lambda) = (\lambda + bx^0)(\lambda - ce^{-\lambda T} + c)$. Clearly, $\lambda = 0$ is a root of the second factor, and setting $\lambda = \alpha + i\beta$ in this factor, we get the following two equations:

$$\alpha + c[1 - e^{-\alpha T}\cos \beta T] = 0 \quad \text{and} \quad \beta + ce^{-\alpha T}\sin \beta T = 0.$$

The first of these implies that $\alpha \leqslant 0$. Just as before, $\lambda = 0$ is seen to be a simple root of this factor. So, the roots of $p(\lambda)$ satisfy the following: $\lambda = 0$ is a simple root, $\lambda = -bx^0 = [c - ab]/(1 + cT)$, and $\operatorname{Re} \lambda < 0$ for all the other roots. Hence, if $c < ab$ the zero solution of the linearization of (1) about $([ab - c]/b(1 + cT), c/b)$ is stable. If $c > ab$, $p(\lambda)$ has a positive real root and the constant solution $([ab - c]/b(1 + cT), c/b)$ of (1) is unstable. When $c < ab$, the presence of $\lambda = 0$ as a root of the characteristic equation precludes the conclusion that $(x^0, c/b)$ is a locally stable solution of the nonlinear equation. The stability of this solution will be analyzed via Lyapunov methods.

We next consider the global behavior of solutions on the set G_a. In order to do this, we employ the invariance principle of LaSalle as formulated in Theorem 3.1, page 119 in Hale [9]. First, assume that $0 < ab \leqslant c$. Let C be the space of continuous functions on $[-T, 0]$ with uniform norm, and use the notation $x_t(s) = x(t + s)$ for $s \in [-T, 0]$. Let $\vec{f} = (f_1, f_2) \in C \times C$ and define the functional V by:

$$V(\vec{f}) = \frac{1}{2c} f_1^2(0) + \frac{1}{2}\int_{-T}^{0} f_1^2(s)\, ds.$$

The derivative $\dot{V}$ of V along solutions of (1) is:

$$\dot{V}(\vec{f}) = -\frac{1}{2}\left[f_1^2(0) - \frac{2b}{c} f_1(0) f_1(-T) f_2(-T) + f_1^2(-T)\right].$$

If this expression is considered as a quadratic in $f_1(0)$ and $f_1(-T)$, it is seen that it is negative definite if $|f_2(-T)| < c/b$. Now, if $0 \leqslant f_i \leqslant a$, $i = 1, 2$, this inequality is satisfied whenever $ab < c$.

Let $G_a^* = \{\vec{f} = (f_1, f_2) \in C \times C: 0 \leqslant f_i \leqslant a$, and $\vec{f}$ satisfies (2a) at $t = 0\} = \{\vec{f}_t: \vec{f} \in G_a\}$. From the proof of part (I) of the theorem G_a^* is seen to be positively invariant under the flow induced by (1). Moreover, if $ab \leqslant c$, $\dot{V}(\vec{f})$ is negative definite for all $\vec{f} \in G_a^* = \bar{G}_a^*$, the closure of G_a^*. Let $S = \{\vec{f} \in G_a^*: \dot{V}(\vec{f}) = 0\}$. From what was shown in the previous paragraph, if $ab \leqslant c$, $S = \{\vec{f} \in G_a^*: f_1(0) = f_1(-T) = 0\}$. By the invariance principle, if $\vec{f} \in G_a^*$, the solution $(x(t), y(t))$ of (1) with $(x_0, y_0) = \vec{f}$ approaches S as $t \to \infty$, in the topology of C. That is, $x_t \to f_1$ with $f_1(0) = f_1(-T) = 0$. But, $x_t(0) = x(t)$, so $x(t) \to 0$ as $t \to \infty$. From (2a) it then follows that, since $0 \leqslant y(t) \leqslant a$, $y(t) \to a$ as $t \to \infty$. So, all solutions that start on G_a^* tend to the solution $(0, a)$ as $t \to \infty$. If $ab = c$, then the set S also includes $\{f \in G_a^*: f_1(0) = \pm f_1(-T)$, and $f_2(-T) = a\}$. By the above arguments, either $x(t) \to 0$ or $y(t) \to a$ as $t \to \infty$. In this latter case, using (2a) as above, we again get $x(t) \to 0$ as $t \to \infty$.

We next assume that $0 < c < ab$. Let G_a^* be as before, and for $\vec{f} \in G_a^*$ define V by: $V(\vec{f}) = \frac{1}{2}[f_2(0) - c/b]^2$. Then, along solutions of (1) we have

$$\dot{V}(\vec{f}) = [f_2(0) - c/b]\dot{f}_2(0) = -b[f_2(0) - c/b]^2 f_1(0),$$

and $\dot{V}(\vec{f}) \leqslant 0$ for $\vec{f} \in G_a^*$. By the invariance principle, if $\vec{f} \in G_a^*$, then the solution (x, y) with $(x(t), y(t)) = (f_1(t), f_2(t))$, $t \in [-T, 0]$, obeys $(x_t, y_t) \to S = \{\vec{f} \in G_a^*: \dot{V}(\vec{f}) = 0\}$ as $t \to \infty$. But, $\dot{V}(\vec{f}) = 0$ implies that either $f_1(0) = 0$ or $f_2(0) = c/b$, that is, $x(t) \to 0$ or $y(t) \to c/b$. We need to consider more closely the first of these two cases.

Now, suppose that $x(t) \to 0$ as $t \to \infty$. Since $0 \leqslant y \leqslant a$, from (2a) we see that $y(t) \to a$ as $t \to \infty$. From (1) we get $y'(t) = x(t)[c - by(t)]$, and given $\varepsilon > 0$ there exists $t^* \geqslant 0$ such that $y(t) \geqslant a - \varepsilon/b$, and hence, $y'(t) \leqslant x(t)[c - ab + \varepsilon]$ for $t \geqslant t^*$. Since $c - ab < 0$, if we let $\varepsilon = (ab - c)/2$ we get $y'(t) \leqslant 0$ for all $t \geqslant t^*$. So, $y(t)$ is monotone non-increasing for $t \geqslant t^*$, $0 \leqslant y(t) \leqslant a$, and $y(t) \to a$ as $t \to \infty$. Thus $y(t) \equiv a$ for $t \geqslant t^*$. So, for $t \geqslant t^*$ we have $y'(t) = 0 = (c - ab)x(t)$; hence since $c - ab \neq 0$, $x(t) = 0$ for $t \geqslant t^*$. From (1) we have $0 = x'(t) = bx(t - T)y(t - T)$ for $t \geqslant t^*$, and hence, $x(t - T)y(t - T) = 0$ for $t \geqslant t^*$. From this and the continuity of y there exists $\eta > 0$ such that $y(t) > 0$ and $x(t) = 0$ for $t \in [t^* - \eta, \infty)$. Since $y'(t) = (c - by(t))x(t)$, we get $y'(t) = 0$, hence, $y(t) = a$ on $t \in [t^* - \eta, \infty)$. From the above argument we see that the supremum of all such $\eta > 0$ is $\eta^* = \min[t^*, T]$. If $\eta^* = t^*$ we have $(x(t), y(t)) = (0, a)$ on $[0, \infty)$, and if $\eta^* = T$, on $[t^* - T, \infty)$. Repeating the above argument starting with all $t \geqslant t^* - T$, then $t \geqslant t^* - 2T$, and so on, we get $(x(t), y(t)) = (0, a)$, on $[0, \infty)$. So, the only solution that tends to $(0, a)$ starting from G_a^* is the one that is identically equal to $(0, a)$ on $[0, \infty)$.

So, if $(x(t), y(t)) \not\equiv (0, a)$, we have $y(t) \to c/b$ as $t \to \infty$. We need to show that $x(t) \to x^0$ in this case. Integrating the second equation in (1) from $t - T$ to t and rearranging, we get

$$y(t) + b\int_{t-T}^{t} x(s)y(s)\,ds = y(t - T) + c\int_{t-T}^{t} x(s)\,ds.$$

From this and (2a), we get

$$Dx_t = x(t) + c\int_{t-T}^{t} x(s)\,ds = a - y(t-T),$$

where $D: C[-T,0] \to \mathbb{R}$ is the linear difference operator defined by $Df = f(0) + c\int_{-T}^{0} f(\theta)\,d\theta$.

Let t^* be a fixed real number and let $u^*(t) = x(t+t^*) - x^0$, $h^*(t) = c/b - y(t+t^*-T)$, and note that since $x^0 = (ab-c)/b(1+cT)$,

$$Du_t^* = h^*(t).$$

The characteristic equation of D (see [9], Section 12.3) is $1 + c(1-\exp(-\lambda T))/\lambda = 0$. Since $cT \neq -1$, $\lambda = 0$ is not a root, and it is readily seen that all roots of this equation obey $\operatorname{Re}\lambda < 0$. Now, in any vertical strip $\alpha < \operatorname{Re}\lambda < \beta$, this equation has only a finite number of zeros, so there exists $\delta > 0$ with $\operatorname{Re}\lambda \leqslant -\delta < 0$ for all zeros λ of this characteristic equation. From Theorem 4.1, page 287 in Hale [9], we can conclude that there exist constants a and b, independent of h^*, such that

$$|u_t^*| \leqslant be^{-at}|u_0| + b \sup_{0\leqslant s\leqslant t} |h^*(s)|.$$

So, as $t \to \infty$

$$\begin{aligned}\overline{\lim}|x(t) - x^0| &= \overline{\lim}|x(t+t^*) - x^0| = \overline{\lim}|u^*(t)| \\ &\leqslant b\sup_{0\leqslant t}|h^*(t)| = b\sup_{t^*\leqslant t}|y(t-T) - c/b| \to 0 \quad \text{as} \quad t^* \to \infty.\end{aligned}$$

Hence, $x(t) \to x^0$ as $t \to \infty$. This completes the proof of (III) and of the theorem.

We note that the above proof can be easily adapted to yield the global behavior of the solutions of a somewhat more general model where the time spent in the exposed class is not a fixed value T, but is given by a probability distribution. The results in this latter case are analogous to those given above.

The model that we have analyzed above is a special case of a slightly more general one where it is assumed that there is a delay $T_1 > 0$ between the time of exposure of the vector carrier to the disease and the time when it becomes infective. This is indeed the case with malaria. This more general model leads to the following system of governing equations for the same variables (x, y, z) as before.

$$\begin{aligned}\frac{dx(t)}{dt} &= bx(t-T-T_1)y(t-T) - cx(t),\\ \frac{dy(t)}{dt} &= cx(t) - bx(t-T_1)y(t),\\ \frac{dz(t)}{dt} &= b[x(t-T_1)y(t) - x(t-T-T_1)y(t-T)].\end{aligned} \tag{6}$$

The correct integral condition now becomes (with $a = 1$):

$$x(t) + y(t) = a - b\int_{t-T}^{t} x(s-T_1)y(s)\,ds, \qquad t \geqslant 0. \tag{7a}$$

When $T_1 = 0$ this reduces to the model we have been discussing. When $T = 0$, (6) reduces to the single delay differential equation $dx(t)/dt = bx(t - T_1)y(t) - cx(t)$, and the algebraic condition $x(t) + y(t) = a$ which is the reduced form of (7a). The integral condition is, hence, not present in this case and we effectively have a single equation $dx(t)/dt = bx(t - T_1)[a - x(t)] - cx(t)$ describing the situation with the subsidiary restriction $0 \leqslant x \leqslant a$. So, it is seen that the condition (7a) is a consequence of the presence of a group of individuals, those who are exposed, which essentially acts as a holding tank for the population for a fixed time T. The equation for $T = 0$ has been analyzed by Cooke [4], while a generalization of the same model allowing for seasonal variations in the vector population, by letting b be a non-constant periodic function of t, was studied by Busenberg and Cooke [2].

For the full model with two delays $T > 0$, $T_1 > 0$ we have the following results.

Theorem 2. *If* $c = 0$, $a \geqslant 0$, *then all solutions of* (6)–(7a) *with initial data obeying* $0 \leqslant x(s)$, $0 \leqslant y(s)$, *for* $s \in [-T_1 - T, 0]$ *and with* (7a) *holding at* $t = 0$ *behave as follows*:

(I) *Either* $x(t) = 0$, $y(t) = a$ *for* $t \geqslant 0$,

(II) *or*, $x(t)$ *approaches* a *monotonically while* $y(t)$ *approaches* 0 *monotonically as* $t \to \infty$.

If $a = 0$, *only alternative* (I) *can hold.*

The proof of this result is elementary and is omitted.

The epidemiological implications of this result are very simple. It says that, if there is no recovery from the disease, then either the whole population is healthy for all time or else the whole population approaches the infective condition as time goes on. The next result is more interesting.

Theorem 3. *Let* $x_0(t)$, $y_0(t)$ *be continuous functions on* $[-T_1 - T, 0]$ *satisfying* $0 \leqslant x_0(t) \leqslant a$, $0 \leqslant y_0(t) \leqslant a$, *for* $t \in [-T_1 - T, 0]$, *and*

$$x_0(0) + y_0(0) = a - b\int_{-T}^{0} x(s - T_1)y_0(s)\,ds,$$

i.e., (7a) *at* $t = 0$. *Then, if* $(x(t), y(t))$ *is the solution of* (6) *with initial data* $(x(t), y(t)) = (x_0(t), y_0(t))$, *for* $t \in [-T - T_1, 0]$, *the following hold*:

(I) $0 \leqslant x(t) \leqslant a$, $0 \leqslant y(t) \leqslant a$, *and* $(x(t), y(t))$ *exists for all* $t \geqslant 0$ *and satisfies* (7a) *on* $t \geqslant 0$.

(II) *If* $0 < ab \leqslant c$, $\lim_{t\to\infty}(x(t), y(t)) = (0, a)$ *for all solutions lying on the invariant set* $G_a = \{(x, y)$: *with* x, y *continuous on* $t \geqslant 0$, $x \geqslant 0$, $y \geqslant 0$ *and* (x, y) *satisfying* (7a) *on* $t \geqslant 0\}$.

(III) *If* $0 < c < ab$ *and* $c \leqslant \min\{1/T, 2/(T_1 + T)\}$, *then the constant solution* $(0, a)$ *is unstable and the trivial solution of the linearization of* (6) *about* $(x^0, y^0) = ((ab - c)/b(1 + cT), c/b)$ *is stable.*

Remarks. The behavior of the equation when $0 < c < ab$ and $c > \min\{1/T, 2/(T_1 + T)\}$ appears to be rather complex. A number of numerical experiments that we have undertaken suggest that periodic solutions may exist in this parameter range. However, our analysis of this model is as yet not complete.

Proof. The proof of (I) proceeds in a manner identical to the proof of (I) in Theorem 1, and so, we omit the details. The proof of (II) is again based on LaSalle's invariance principle. We let $C = C[-T_1 - T, 0]$ be the continuous functions on $[-T_1 - T, 0]$ endowed with the uniform norm. Let $\vec{f} = (f_1, f_2) \in C \times C$ and define the functional V by

$$V(\vec{f}) = \frac{1}{2c} f_1^2(0) + \frac{1}{2} \int_{-T-T_1}^{0} f_1^2(s)\, ds.$$

The derivative $\dot{V}$ of V along solutions of (1) is given by

$$\dot{V}(\vec{f}) = -\frac{1}{2}\left[f_1^2(0) - \frac{2b}{c} f_1(0) f_1(-T_1 - T) f_2(-T) + f_1^2(-T_1 - T) \right].$$

Viewing this expression as a quadratic in $f_1(0)$ and $f_1(-T_1 - T)$, we see that it is negative definite if $|f_2(-T)| \leqslant c/b$. Now, if $0 \leqslant f_i \leqslant a$, $i = 1, 2$, this is satisfied whenever $ab \leqslant c$.

From this point on, the proof proceeds in exactly the same manner as the proof of (II) of Theorem 1, and so, the details are omitted.

The proof of (III) involves the analysis of the characteristic equation of the linearization of (6) about the constant solutions $(x, y) = (0, a)$ and $(x, y) = (x^0, c/b) = ((ab - c)/b(1 + cT), c/b)$. In the first case, the characteristic quasi-polynomial is $p(\lambda) = \lambda[\lambda + c - abe^{-\lambda(T+T_1)}]$, and setting $\lambda = u + 0i$, we get the following equation for the real roots of p: $u[u + c - abe^{-u(T+T_1)}] = 0$. If $ab > c$, this always has a positive solution, hence, the constant solution $(0, a)$ is unstable.

The characteristic quasipolynomial for the linearization about the root $(x^0, c/b)$ is:

$$p(\lambda) = \lambda^2 - \lambda[c(e^{-\lambda(T+T_1)} - 1) - bx^0] - cbx^0[e^{-\lambda T} - 1].$$

Since $ab > c$, $x^0 > 0$ and if $\lambda = u + iv$ the imaginary part of the equation $p(\lambda) = 0$ is

$$2uv - v[c(e^{-u(T+T_1)} \cos v(T + T_1) - 1) - bx^0] + uce^{-u(T+T_1)} \sin v(T + T_1) + cbx^0 e^{-uT} \sin vT = 0.$$

This can be rewritten as

$$u = \{vc(e^{-u(T+T_1)} \cos v(T + T_1) - 1) - bx^0(v + ce^{-uT} \sin vT)\} \times \{2v + ce^{-u(T+T_1)} \sin v(T + T_1)\}^{-1}.$$

Noting that the right-hand side of this expression does not change when v is replaced by $-v$, we consider the case where $v > 0$. Since $c \leqslant \min(1/T, 2/(T + T_1))$, if we assume that $u \geqslant 0$, we have $e^{-u(T+T_1)} \cos v(T + T_1) - 1 \leqslant 0$, $-bx^0(v + ce^{-uT} \sin vT) < 0$, and $2v + ce^{-u(T+T_1)} \sin v(T + T_1) > 0$. So, the right-hand side of the equation is negative, implying that $u \leqslant 0$. This is a contradiction, and hence, all solutions of the quasipolynomial must obey $\operatorname{Re} \lambda \leqslant 0$.

We next note that $\lambda = 0$ is a root of p and that $p'(0) = bx^0[1 + cT_1] > 0$ if $x^0 \neq 0$, which is the case when $ab > c$. So, $\lambda = 0$ is a simple root of p. This completes the proof of (III) and the theorem.

We note that the neglect of the integral condition (7a) leads to the same type of difficulties in this more general model as it did in the case where $T_1 = 0$.

For example, one may have $0 < b \leqslant c$, hence, all solutions on G_1 approach $(0, 1)$ as $t \to \infty$. While, if $ab > c$ $(a > 1)$ the solution $(0, a)$ is unstable.

3. Invariant Integral Conditions

In the first part of the preceding section we saw the form and the effects of an invariant integral condition in an epidemiological model. The type of condition that was encountered there always appears whenever a model contains a compartment (the exposed class in the model of section 2) which holds individuals for a fixed time T. If $y(t)$ denotes the number of individuals in this compartment, then one must have $dy(t)/dt = h(t) - h(t-T)$, as the governing differential difference equation. This equation has a first integral $y(t) = a + \int_{t-T}^{t} h(s)\,ds$, which is the correct invariant restriction. Note that a restriction of this sort is needed for each compartment of the above type. The importance of these restrictions was illustrated in the previous section and will not be discussed further at this point.

The next point that we take up is the question of what is the appropriate integral condition in models that allow for removal, say due to death, from a holding compartment such as the one discussed above. We proceed by first looking at a particular model that has been studied by Z. Grossman [7] who recognized the importance of restricting the initial data in his model. In this model, we consider a population having a birth rate μ, a death rate μ, and broken up into three components: those susceptible (S), those who are infective (I), and finally the immune (M). It is assumed that there is a fixed delay T between the time of entrance in the infective class and the time of departure from it into the immune class. Since the birth and death rates are equal, the population remains constant and the variables S, I, and M are normalized by division by this constant total population to get the relation

$$I(t) + M(t) + S(t) = 1. \tag{8}$$

The equations describing the dynamics of this model are [7]:

$$\begin{aligned} S'(t) &= \mu - \mu S(t) - \beta I(t)S(t), \\ I'(t) &= \beta I(t)S(t) - e^{-\mu T}\beta I(t-T)S(t-T) - \mu I(t), \\ M'(t) &= e^{-\mu T}\beta I(t-T)S(t-T) - \mu M(t). \end{aligned} \tag{9}$$

It would appear from this formulation of the model that the first two equations of (9) can be solved independently of the third, and then (8) used to get M as a function of t. Of course the solutions that one seeks must obey (9) as well as

$$0 \leqslant S(t), \qquad 0 \leqslant I(t), \qquad 0 \leqslant M(t). \tag{10}$$

The difficulty with the above approach is that (8), (9) and (10) do not form a well posed problem. In order to see this, note that, if initially one has $S(t) = \frac{1}{2}$, $I(t) = -t/(2T)$, $M(t) = (1 + t/T)/2$, $t \in [-T, 0]$, then (8) and (10) are satisfied on this initial interval, but from (9) we get $I'(0) = -e^{-\mu T}\beta S(-T)I(-T) = -e^{-\mu T}\beta/4 < 0$, so there exists $\varepsilon > 0$ with $I(t) < 0$ on $(0, \varepsilon)$, violating (10).

In order to remove this difficulty, we note that the second equation in (9) has the first integral, with an arbitrary constant a,

$$I(t) = e^{-\mu t}a + \beta \int_{t-T}^{t} e^{\mu(s-t)} I(s)S(s)\,ds, \tag{11a}$$

We now note that in this equation $a = 0$ is the correct value of the constant. The heuristic way of viewing this condition is the following: The total number of infectives should be exactly equal to the number who have entered the class in the past and have not left either via natural death or because of the fixed residence time T in this class. The correct interpretation of the mathematical results obtained from this model must incorporate this integral condition. We have the following result that shows that the condition (11a) makes the problem well posed.

Theorem 4. *If β, μ and T are positive and if $a \geqslant 0$, then all solutions of* (9) *with initial data satisfying* (8) *at $t = 0$ must satisfy* (8) *for all $t \geqslant 0$ for which they exist. If, moreover, the initial data satisfy* (10) *for $t \in [-T, 0]$, and* (11a) *at $t = 0$, then the solution through that initial data exists and satisfies* (8), (10) *and* (11a) *for all $t \geqslant 0$.*

We postpone the proof of this theorem until the end of this section, and proceed to another model found in the recent literature where integral conditions play an important role. In this situation, the integral condition determines whether or not certain periodic solutions have biological significance. This model, which is studied by D. Green [6], is concerned with an epidemic where the population can be separated into three classes: Those susceptible (S), those who are infective (I), and those who are recovered (R) but have, for limited time, developed immunity to the disease. The model assumes a fixed time σ during which infectiousness lasts and a fixed time ω during which immunity lasts. It is a special case of a model proposed by Cooke [5] and studied in detail by Hoppensteadt and Waltman [10]. The dynamics of the model are described by the following equations which are used in [6],

$$\begin{aligned} S'(t) &= -rI(t)S(t) + rI(t-T)S(t-T), \qquad T = \sigma + \omega, \\ I'(t) &= rI(t)S(t) - rI(t-\sigma)S(t-\sigma). \end{aligned} \tag{12}$$

However, the form that the Hoppensteadt-Waltman equations take in this special case is

$$I(t) = b + r \int_{t-\sigma}^{t} I(s)S(s)\,ds, \tag{13a}$$

$$S(t) + I(t) = a - r \int_{t-T}^{t-\sigma} I(s)S(s)\,ds, \tag{13b}$$

with $a \geqslant 0$ and $b = 0$ being the appropriate values for the constants in the biological interpretation of these relations.

Even though any pair of constants (α, β) solves (12), this is not the case with (13a, b). In fact setting $b = 0$, and assuming $(S, I) \equiv (\alpha, \beta)$, we conclude that

$$(\alpha, \beta) = (a, 0) \qquad \text{or} \qquad (\alpha, \beta) = \left(\frac{1}{r\sigma}, \frac{r\sigma a - 1}{rT}\right).$$

In [6] the case where $r = 0.2$ and $\sigma = 1$ is studied with ω used as a parameter. Values of ω, and (α, β) are sought for Hopf bifurcation to occur and periodic solutions to exist. With the correct integral condition the constant solutions have $\beta = 0$ or $\alpha = 1/(r\sigma) = 5$. However, in all of the cases considered in [6] $\beta \neq 0, \alpha \neq 5$, so the periodic solutions found in that work have no direct relation to the biological situation that is being modelled. However, there do exist periodic solutions that obey the correct conditions. In fact, setting $\omega = \sigma = 1$ and using ra as a parameter one can show, using the methods of Green [6], that at $ra = 1.4$, approximately, Hopf bifurcation occurs at the constant solution $(1/r, 0.2/r)$ which satisfies the correct integral condition. Since the method of establishing this fact is the same as in [6] we shall omit the details.

There is another difficulty involved in writing the equations of the model in the differential form (12). For, the biological interpretation of variables I and S implies that $I(t) \geqslant 0$ and $S(t) \geqslant 0$ for all $t \geqslant 0$. If, however, one chooses the initial data $I(t) = -Nt/(2T)$, $S(t) = N[T + t]/2T$, $t \in [-T, 0]$, then $I(t) \geqslant 0$, $S(t) \geqslant 0$, $I(t) + S(t) \leqslant N$ on $t \in [-T, 0]$, where N is taken as the total population. However, since $I(0) = 0$ and $I(-\sigma)S(-\sigma) = N^2\sigma\omega/(2T)^2 > 0$, we see from (12) that $I'(0) < 0$, and $I(t)$ is, hence, negative for small values of $t > 0$. Again, one can show that the integral conditions remove such difficulties. In fact, we have the following result.

Theorem 5. *If r, σ, ω and a are positive and if $b \geqslant 0$, then all solutions of* (12) *with non-negative initial data satisfying* (13a, b) *at $t = 0$ exist for all $t \geqslant 0$, satisfy* (13a, b) *on $t \geqslant 0$, and are non-negative and satisfy $S(t) + I(t) \leqslant a$ for all $t \geqslant 0$.*

The proof of this result is very similar to that of Theorem 4, and so will be omitted.

We next consider the model of Cooke and Yorke [3] that was mentioned in the introduction. Our treatment relies on the methods and the discussion of the same model that was given by Hale [8]. The equation that we treat is

$$x(t) = a + \int_{t-L_1-L_2}^{t-L_1} g(x(s))\,ds = a + L(g(x)) \tag{14a}$$

where $a \geqslant 0$ is a constant. In the population model considered in [3], the correct value for a is zero. The function g is assumed to be three times continuously differentiable and to satisfy $g(x) = 0$ for $x \leqslant 0$, and for $x \geqslant 1$, while $g(x) > 0$ on $0 < x < 1$. If $x = c$ is a constant solution of (14a) it must satisfy $c = a + L(g(c))$, and we denote by c_a any solution of this equation. Linearizing (14a) about such a constant solution c_a, we have the following equation for $u = x - c_a$:

$$u(t) = g'(c_a)\int_{t-L_1-L_2}^{t-L_1} u(s)\,ds.$$

The characteristic equation of this equation is

$$\lambda = b[e^{-L_1\lambda} - e^{-(L_1+L_2)\lambda}], \tag{15}$$

where $b = g'(c_a)$. The roots of this quasipolynomial have been studied by Cooke and Yorke [3]. One of their results yields the following: If $0 < b < L_2^{-1}$, there is one simple negative root, the root at $\lambda = 0$, and no other real roots. Moreover, all complex roots have negative real parts.

In the discussion that follows we shall need the following lemma.

Lemma 1. *There is an interval of values of* b *of the form* $b_0 < b < 0$ *with the following property: If* $b \in (b_0, 0)$, *equation* (15) *has a simple root* $\lambda = 0$ *and all other roots of* (15) *satisfy* $\operatorname{Re} \lambda < 0$. *At* $b = b_0$, (15) *has a single pair of simple pure imaginary roots, the root* $\lambda = 0$, *and all other roots satisfy* $\operatorname{Re} \lambda < b^* < 0$ *for some* $b^* < 0$.

The proof of this lemma requires a lengthy but elementary consideration of equation (15) and is not given here. Finally, we need the following result from Hale [8].

Lemma 2 (Hale). *Suppose that the roots of* (15) *satisfy the following condition. For some fixed value of* b, *say* $b = b_0$, *there are two complex conjugate pure imaginary roots and all other roots* λ *of* (15) *satisfy* $\operatorname{Re} \lambda < b^* < 0$ *for some* $b^* < 0$. *Then, there are functions* g^* *close to* g *for which the equation* $x'(t) = g^*(x(t - L_1)) - g^*(x(t - L_1 - L_2))$ *has a nonconstant periodic solution. Here the functions* g, g^* *are of class* C^3 *on* $[-L_1 - L_2, 0]$, *and* $\|g\| = \sup\{\sum_{j=0}^{3} \|g^{(j)}(f)\| : f \in C[-L_1 - L_2, 0]$, $\|g^{(0)}(f)\|$ *denotes the supremum norm of* $g(f(s))$ *on* $s \in [-L_1 - L_2, 0]$, *and the other norms* $\|\ \|$ *denote norms of linear and multilinear operators on* $C[-L_1 - L_2, 0]\}$.

After all this, we can return to our example. Here, if $b = b_0$, then the conditions of Hale's result hold. So, there are functions close to g for which non-trivial periodic solutions of the delay differential equation exist. Recalling that $b = g'(c_a)$ depends on the value of a entering in the integral condition (14a), we note that this condition plays a role in deciding whether or not periodic solutions exist.

We now conclude by giving the proof of Theorem 4.

Proof of Theorem 4. From (9) we get

$$S'(t) + I'(t) + M'(t) = \mu[1 - S(t) - M(t) - I(t)],$$

and letting $x = S + I + M$, we have $x(0) = 1$, $x'(t) = \mu(1 - x(t))$. So, $x(t) = 1$ for $t \geqslant 0$, and (8) holds for all $t \geqslant 0$. We note also that the solution $x \equiv 1$ is asymptotically stable.

Now, if the initial data satisfy (11a) at $t = 0$, then for all $t > 0$ where I, S and M exist we have from the second equation in (9)

$$\begin{aligned} I(t) &= e^{-\mu t} I(0) + \beta \int_0^t e^{\mu(s-t)} [I(s)S(s) - e^{-\mu T} I(s - T)S(s - T)]\, ds \\ &= e^{-\mu t} I(0) + \beta \int_0^t e^{\mu(s-t)} I(s)S(s)\, ds - \beta \int_{-T}^{t-T} e^{\mu(s-t)} I(s)S(s)\, ds \\ &= e^{-\mu t} I(0) + \beta \int_{t-T}^t e^{\mu(s-t)} I(s)S(s)\, ds - \beta \int_{-T}^{0} e^{\mu(s-t)} I(s)S(s)\, ds. \end{aligned}$$

Since (11a) is satisfied at $t = 0$, we have

$$\beta \int_{-T}^{0} e^{\mu(s-t)} I(s)S(s)\, ds = e^{-\mu t}[I(0) - a],$$

and using this in the above equation, we get

$$I(t) = e^{-\mu t}a + \beta \int_{t-T}^{t} e^{\mu(s-t)} I(s)S(s)\, ds.$$

So, (11a) is satisfied for all such $t > 0$. From this, we see that $I(t) \geqslant 0$ whenever $I(s)S(s)$ exists and is non-negative for $s \in [t - T, t]$.

From the first equation in (9), we have

$$S(t) = \exp\left[-\int_0^t (\mu + \beta I(s))\, ds \right] \left[S(0) + \int_0^t \mu \exp\left\{ \int_0^s (\mu + \beta I(u))\, du \right\} ds \right].$$

So, $S(t) \geqslant 0$ for all $t > 0$ such that $I(s)$ exists for $s \in [0, t]$.

From the third equation in (9) we get

$$M(t) = e^{-\mu t}M(0) + \beta \int_0^t e^{\mu(s-t-T)} I(s - T)S(s - T)\, ds.$$

So, $M(t) \geqslant 0$ for all $t > 0$ such that $I(s)S(s)$ exists and is non-negative for $s \in [t - T, t]$.

We now consider two separate cases: $a > 0$ and $a = 0$. First, let $a > 0$, and suppose that $I(t) = 0$ for some $t \geqslant 0$. Then from (11a) we see that $\int_{t-T}^{t} I(s)S(s)\, ds < 0$, so $I(s)S(s) < 0$ for some $s \in [t - T, t]$. But, $S(s) \geqslant 0$ from what was said before, so $I(s) < 0$ for some $s < t$. It is clear from the above arguments that the infimum s^* of all such s must be nonpositive. Now, since $a > 0$, and $I(s)S(s) \geqslant 0$ on $[-T, 0]$, $I(0) > 0$ from (11a). So, $s^* < 0$, and $I(s^*) < 0$ contradicting the non-negativity of the initial data. So, $I(t) > 0$ for all $t > 0$ where it exists.

Next, consider the case $a = 0$. Suppose that $I(t) < 0$ for some $t \geqslant 0$. From (11a) we have $I(s)S(s) < 0$ for some $s \in [t - T, t]$, hence, as before $I(s) < 0$ for some $s < t$. Again, the infimum of all such s is nonpositive. Now, if $I(0) = 0$ we have from (11a) $I(s)S(s) = 0$ for $s \in [-T, 0]$. So, we must have from (9) $I'(t) = \beta I(t)S(t) - \mu I(t)$ for all $t \in [0, T]$, implying that, $I(t) \equiv 0$ for $t \in [0, T]$. Repeating this argument we see that $I(t) \equiv 0$ on $[0, \infty)$. Here, we note that $S(t)$ exists and is nonnegative for all t where $I(t)$ exists.

So, $I(t) \geqslant 0$ on its whole interval of existence $t \in [0, \alpha)$. Hence, $S(t)$ and $M(t)$ both exist and are nonnegative on this same interval. But, $I(t) + S(t) + M(t) = 1$, so they are all bounded by one on this interval. From standard continuation results we see that $\alpha = \infty$, and the theorem is proved.

The proof of Theorem 4, and other similar results, would be greatly simplified if one could conclude that nonnegative initial data lead to nonnegative solutions if (11a) is satisfied. Of course, if (11a) does not hold, we have shown already that nonnegative data need not have this property. It would be useful to have available general results, along the lines of those of Seifert [12], that yield the nonnegativity of solutions when invariant integral conditions like (11a) are imposed. The nature of (11a) seems to rule out the application of the results in [12], while those extensions of such results that we have been able to establish do not reduce the labor needed in proving propositions similar to Theorem 4.

We next consider the relation between the present invariant integral approach to this type of model and the approach described by Hoppensteadt [11, pp.

47–49]. For the model which is described by equations (1)–(2), this alternate formulation leads to the following equations:

$$
\begin{aligned}
z(t) &= z^0(t) + b\int_0^t x(s)y(s)\,ds,\\
x(t) &= x(0)e^{-ct} - \int_0^t \frac{dz^0(s)}{ds}e^{-c(t-s)}\,ds,\\
y(t) &= 1 - x(t) - z(t),
\end{aligned}
\tag{16}
$$

when $t \in [0, T]$, and $z^0(t)$ is non-increasing on $[0, T]$ and zero on $[T, \infty)$. Here $z^0(t)$ is viewed as the proportion of those who are initially exposed and are still exposed at time t. For $t > T$ the following equations are used

$$
\begin{aligned}
x'(t) &= bx(t-T)y(t-T) - cx(t),\\
z'(t) &= b[x(t)y(t) - x(t-T)y(t-T)],\\
y(t) &= 1 - x(t) - z(t).
\end{aligned}
\tag{17}
$$

This formulation of the model does not seem to imply the integral condition (2). However, (17) implies that (1) holds for $t > T$ while if we define initial data $(x_T(s), y_T(s), z_T(s)) = (x(s+T), y(s+T), z(s+T))$, $s \in [-T, 0]$, where (x, y, z) satisfy (16), we see that (since $z^0(T) = 0$)

$$x(T) + y(T) = 1 - b\int_0^T x(s)y(s)\,ds,$$

and condition (2) is satisfied at $t = T$, hence, for all $t \geqslant T$. So, solutions of the formulation via (16)–(17) must also satisfy the invariant integral condition on $t \geqslant T$. Hence, the study of the asymptotic behavior of such solutions involves the same considerations as those we have been discussing. Incidentally, the observation that (2) is satisfied at $t = T$ provides, through Theorem 1, a proof that this alternate formulation of the model is well-posed in the sense that solutions remain non-negative and satisfy $x(t) + y(t) + z(t) = 1$ for all $t \geqslant 0$.

We next show that solutions of (1)–(2) are also solutions of (16)–(17) for appropriately chosen z^0. In fact, let (x, y, z) satisfy (1)–(2) and define

$$
z^0(t) = \begin{cases} 0 & \text{if } t > T\\ z(0) - b\displaystyle\int_0^t x(s-T)y(s-T)\,ds, & t \in [0, T]. \end{cases}
\tag{18}
$$

That is, view $z^0(t)$ as the proportion that is exposed at $t = 0$ minus those leaving the exposed class before $t \leqslant T$. Clearly, $z^0(t)$ is non-increasing on $[0, T]$, and $z^0(T) = z(0) - b\int_0^T x(s-T)y(s-T)\,ds = 1 - x(0) - y(0) - b\int_{-T}^0 x(s)y(s)\,ds = 0$, by condition (2). So, $z^0(t) \geqslant 0$ in $[0, T]$ and satisfies the hypothesis imposed on (16)–(17). Since we have already shown that (1)–(2) implies that (17) will hold for all $t > 0$, we only need to show that (16) holds for $t \in [0, T]$. Now, from (18) $dz^0(t)/dt = -bx(t-T)y(t-T)$ when $t \in [0, T]$, and hence from the equation in (1) we get

$$x(t) = x(0)e^{-ct} + b\int_0^t e^{-c(t-s)}x(s-T)y(s-T)\,ds$$
$$= x(0)e^{-ct} - \int_0^t e^{-c(t-s)}\frac{dz^0(s)}{ds}\,ds.$$

Thus, the second equation in (16) holds for $t\in[0,T]$. Next, from (2) we have

$$z(t) = 1 - x(t) - y(t) = b\int_{t-T}^t x(s)y(s)ds$$
$$= b\int_0^t x(s)y(s)\,ds + b\int_0^T x(s-T)y(s-T)\,ds - b\int_0^t x(s-T)y(s-T)\,ds,$$

so, using the fact that $b\int_0^T x(s-T)y(s-T)\,ds = z(0) - z^0(T) = z(0)$,

$$z(t) = b\int_0^t x(s)y(s)\,ds + z(0) - b\int_0^t x(s-T)y(s-T)\,ds$$
$$= b\int_0^t x(s)y(s)\,ds + z^0(t),$$

for $t\in[0,T]$. That is, the first equation in (16) is satisfied. The last equation in (16) follows from (2) and the fact that $z(t) = b\int_{t-T}^t x(s)y(s)\,ds$. It should be noted that the model formulated via equations (16)–(17) satisfies (1)–(2) on $t\geqslant T$ but need not satisfy condition (1)–(2) on $t\in[0,T]$. This is due to the fact that the choice of the initial function z^0 is fairly arbitrary and may not reflect any of the specific dynamics of the situation being modelled. It is hence interesting to note that the additional conditions needed to have (1)–(2) satisfied on $t\geqslant 0$ are rather simple. In fact, if we assume that, in addition to (16)–(17), $x'(t)$ and $y'(t)$ are continuous at $t=T$, z^0 is continuous at $t=0$ and dz^0/dt is integrable on $[0,T]$ (it, of course, exists almost everywhere and is measurable since z^0 is monotone), then it follows that (1)–(2) are satisfied on $[0,\infty)$. To see this, note that in this case, (16)–(17) imply that

$$x'(T) = -cx(T) - \frac{dz^0(T)}{dt} = bx(0)y(0) - cx(T),$$

that is,

$$\frac{dz^0(T)}{dt} = -bx(0)y(0). \tag{19}$$

Now define $(x_0(t), y_0(t), z_0(t))$ for $t\in[-T,0]$ by

$$x_0(t)\geqslant 0,\qquad y_0(t)\geqslant 0,\qquad -bx_0(t)y_0(t) = \frac{dz^0}{ds}(T+t),$$
$$x_0(0) + y(0) = 1 - z^0(0). \tag{20}$$

The conditions that (20) imposes on x_0 and y_0 at zero are satisfiable since (19) is assumed to hold. Now the second equation in (16) yields for $t\in(0,T)$,

$$x(t) = x(0)e^{-ct} + b\int_0^t e^{-c(t-s)}x_0(s-T)y_0(s-T)\,ds,$$

hence, $x'(t) = bx_0(t-T)y_0(t-T) - cx(t)$, since $-bx_0(t-T)y_0(t-T) = dz^0/dt\,(t+T)$ was assumed to be integrable. The first equation in (16) yields

$$z'(t) = \frac{dz^0(t)}{dt} + bx(t)y(t) = -bx_0(t-T)y_0(t-T) + bx(t)y(t),$$

and using the third equation in (16) we now get

$$y'(t) = cx(t) - bx(t)y(t), \qquad t \in (0, T].$$

This shows that (1) holds on $(0, T]$ with initial data given by (20).

We next verify that (2) holds at $t = 0$. From what was just shown, on $t \in (0, T]$,

$$\frac{d}{dt}[x(t) + y(t)] = b[x_0(t-T)y_0(t-T) - x(t)y(t)],$$

hence, using the continuity of x and y at $t = 0$,

$$\begin{aligned} x(0) + y(0) &= x(T) + y(T) - b\int_0^T x_0(s-T)y_0(s-T)\,ds + b\int_0^T x(s)y(s)\,ds \\ &= x(T) + y(T) - b\int_{-T}^0 x_0(s)y_0(s)\,ds + z(T) - z^0(T) \\ &= 1 - b\int_{-T}^0 x_0(s)y_0(s), \end{aligned}$$

since $z^0(T) = 0$ and $x(T) + y(T) + z(T) = 1$. So, from Theorem 1, (2) must hold for all $t \geqslant 0$, hence (1)–(2) hold on $(0, \infty)$.

The above arguments can be adapted to show that these two approaches to all the models we have been considering are essentially equivalent. When the approach in [11] is taken, the transient behavior on $[0, T)$ may differ from that obtained via the invariant integral formulation. However, the behavior on $[T, \infty)$ is the same. In particular, in studying stability questions, periodic solutions, or other behavior for $t \geqslant T$, the analysis must be restricted to solutions obeying the proper invariant integral conditions.

Acknowledgement. We wish to thank Jack Carr, Jack Hale and Herbert Hethcote for a number of useful comments on this work.

References

1. Bellman, R., Cooke, K.: Differential-difference equations. New York: Academic Press 1963
2. Busenberg, S., Cooke, K.: Periodic solutions of a periodic non-linear delay differential equation. SIAM J. Applied Math. **35**, 704–721 (1978)
3. Cooke, K., Yorke, J.: Some equations modelling growth processes and gonorrhea epidemics. Math. Biosciences **16**, 75–101 (1973)
4. Cooke, K.: Stability analysis for a vector disease model. Rocky Mountain Math. J., **9**, 31–42 (1979)

5. Cooke, K.: Functional-differential equations: Some models and perturbation problems. Differential equations and dynamical systems, pp. 167 – 183. New York: Academic Press 1967
6. Green, D.: Self-oscillations for epidemic models. Math. Biosciences **38**, 91 – 111 (1978)
7. Grossman, Z.: Oscillatory phenomena in a model of infectious diseases, preprint
8. Hale, J.: Behavior near constant solutions of functional differential equations. J. Diff. Eq. **15**, 278 – 294 (1974)
9. Hale, J.: Theory of functional differential equations. New York: Springer 1977
10. Hoppensteadt, F., Waltman, P.: A problem in the theory of epidemics. Math. Biosciences **9**, 71 – 91 (1970)
11. Hoppensteadt, F.: Mathematical theories of populations: Demographics, genetics and epidemics, Philadelphia: SIAM 1975
12. Seifert, G.: Positively invariant closed sets for systems of delay differential equations. J. Diff. Eq. **22**, 292 – 304 (1976)

Received May 16, 1979/Revised February 28, 1980