Endemic Disease in Environments with Spatially Heterogeneous Host Populations*

W. M. POST, D. L. DeANGELIS, AND C. C. TRAVIS*
Environmental Sciences Division, Oak Ridge National Laboratory,
Oak Ridge, Tennessee 37830
Received 29 March 1982; revised 18 October 1982

ABSTRACT

The main interest in epidemic models stems from their use in uncovering certain qualitative features of epidemic processes. A deterministic model of a general epidemic in a population with an arbitrary number of separate population centers is presented. The mixing within each center is assumed to be homogeneous, and the usual threshold theorem holds for each population. The mixing between centers is nonhomogeneous. This model is used to identify the necessary and sufficient conditions under which a disease will become endemic in the general population when each population center is below the threshold required for establishment of the disease and does not mix with other centers. These conditions depend critically on the concavity of the infection rate function with respect to the length of exposure time. The application of these results to host-vector models is discussed.

I. INTRODUCTION

Mathematical models have been used to show that an isolated host population must have a critical population size for certain contagious diseases to be maintained in the population over a long period of time. Bartlett [2,3] suggested that an isolated human community must have at least 250,000 to 300,000 inhabitants for measles to be maintained. Maintenance of a disease, however, depends not only on population size, but also on the spatial arrangement of settlement. In a densely compacted population, most mem-

*Research supported by the National Science Foundation's Ecosystem Studies Program under Interagency Agreement No. DEB 77-25781 with the U.S. Department of Energy under contract W-7405-eng-26 with Union Carbide Corporation. Publication No. 2065, Environmental Sciences Division, ORNL.

*Health and Safety Research Division, Oak Ridge National Laboratory.
bers of the population will soon be exposed to the infection, which may then be maintained in the population. When the population is spatially heterogeneous, on the other hand, the contagion may not become established because the transmittal rate between population centers is too low.

In this report we look at a deterministic model of a host population in which spatial heterogeneity is taken into account by dividing the population into several groups which interact slightly. The degree of mixing between groups can be specified, relaxing the assumption of uniform mixing of all individuals. We shall consider only the case where the disease is directly contagious, that is, where no intermediate host is necessary.

Lajmanovich and Yorke [9] and Hethcote [8] have analyzed epidemic models for a nonhomogeneous population which can be divided into homogeneous subgroups. They establish necessary and sufficient conditions so that the disease will become established in the entire population. These conditions are stated in terms of the maximum real parts of the eigenvalues of a certain matrix. One contribution of the present paper is the use of $M$-matrix theory [13] to obtain equivalent conditions that are more useful than the eigenvalue criterion. We present alternative equivalent conditions that are easily interpreted in terms of the model parameters and that may be verified in a small number of simple arithmetical calculations.

Using these conditions, we establish that under the usual assumptions on interactions between susceptible and infectives, a disease cannot become endemic in the entire population unless it is endemic in some isolated subpopulation. We then examine the effect of assuming that infectibility is a nonlinear function of the fraction of time populations are in contact. That is, we assume that visitation between population centers increases the risk of contracting a disease. Under this assumption we establish sufficient conditions for a disease to become endemic in a heterogeneous population even though each subpopulation is below the required threshold when in isolation.

2. EFFECTS OF SPATIAL HETEROGENEITY

The incorporation of spatial heterogeneity into epidemic models can have two opposite effects. Models that incorporate continuous spatial distribution of individuals which interact strongly with neighbors and weakly with more distant individuals demonstrate a damping effect of geographic dispersion due to an effective lowering of the rate of infection [1, 11]. On the other hand, the fadeout of a disease in a subpopulation below the threshold size may be countered by reintroduction from other subpopulations [4]. Thus, interaction between subpopulations can effectively raise the infection rate so that the disease will persist in the total population even when it would fade out of each separate isolated subpopulation. It is toward this latter possibility that we focus our attention.
Let us assume that the total population exists in a region of $m$ population centers, where the $i$th center has a constant population, including susceptibles, infecteds, and immunes, of $N_i$. The members of each center make short visits to at least some of the other centers. To model this realistically, we could introduce complex model equations, but we shall attempt to simplify matters as much as possible so that our basic point can be made clearly. Actually, only certain fractions of each population will visit certain other population centers. However, we shall make the simplifying assumption that all members of each center spend the same amount of time visiting other centers, though the time spent visiting depends on the center. While visiting the other centers, infected visitors will have the chance of transmitting the disease to susceptibles in the visited center, while susceptible visitors have a chance of acquiring the disease from infected members of the visited population center.

We point out, here, that we are not modeling the migration of individuals from one population to another, since we assume that visitors return to their proper subpopulation. We are, in a sense, modeling the migration of the disease itself, rather than the migration of hosts.

We can represent these interactions by two sets of $m$ equations,

$$\frac{dx_i}{dt} = a_iN_i - a_ix_i - \sum_{j=1}^{m} f(x_i, y_j, T_{ij}),$$

(1)

$$\frac{dy_i}{dt} = \sum_{j=1}^{m} f(x_i, y_j, T_{ij}) - (a_i + r_i)y_i,$$

(2)

The number of susceptibles and infecteds in each subpopulation are represented by $x_i, y_i$, respectively. Since we assume that each subpopulation has a constant total number of individuals, $N_i$, we do not need equations for the number of immunes. The constant $a_i > 0$ represents the reproductive rate, assumed, for simplicity, the same for each subpopulation category. Each population category undergoes natural mortality. However, since we are assuming constant subpopulation size, these mortality rates are equal to the reproductive rate $a_i$. The rate at which infected hosts recover, thereby joining the category of immunes, is $r_i$. The functions $f(x_i, y_j, T_{ij})$ represent the rate of contagion in subpopulation $i$ resulting from contact with infected individuals of subpopulation center $j$. We assume that $f(x_i, y_j, T_{ij})$ is an increasing function of $x_i$ and $y_j$ and that $\frac{\partial f}{\partial x_i} = 0$ when $y_j = 0$. The parameter $T_{ij}$ represents the time of contact as a fraction of some unit of time. This is incorporated to express the idea that the rate of contagion between two subpopulation centers should increase with the fraction of time members from each center are in contact.
We now want to determine conditions under which interactions between subpopulation centers will cause a disease to become established when each center, in isolation, is incapable of supporting the disease. This is determined by examining local stability of the equilibrium point \( \bar{N} = (N_1, N_2, \ldots, N_m, 0, 0, \ldots, 0) \). The perturbed equations obtained from (1) and (2) in the neighborhood of \( \bar{N} \) are

\[
\begin{bmatrix}
\frac{dx}{dt} \\
\frac{dy}{dt}
\end{bmatrix} = \begin{bmatrix}
A_{11}(\bar{N}) & A_{12}(\bar{N}) \\
A_{21}(\bar{N}) & A_{22}(\bar{N})
\end{bmatrix} \begin{bmatrix}
x \\
y
\end{bmatrix},
\]

where \( \frac{dx}{dt}, \frac{dy}{dt}, x, \) and \( y \) are \( m \times 1 \) vectors and \( A_{11}, A_{12}, A_{21}, \) and \( A_{22} \) are \( m \times m \) matrices of the form

\[
A_{11}(\bar{N}) = \text{diag} \left[ -a, -\sum_{j=1}^{m} \frac{\partial f(x_j, y_j, T_{j1})}{\partial x_j} \right],
\]

\[
A_{12}(\bar{N}) = \begin{bmatrix}
-\frac{\partial f(x_1, y_2, T_{11})}{\partial y_1} & -\frac{\partial f(x_1, y_2, T_{12})}{\partial y_2} & \cdots & -\frac{\partial f(x_1, y_m, T_{1m})}{\partial y_m} \\
-\frac{\partial f(x_2, y_1, T_{21})}{\partial y_1} & -\frac{\partial f(x_2, y_2, T_{22})}{\partial y_2} & \cdots & -\frac{\partial f(x_2, y_m, T_{2m})}{\partial y_m} \\
\vdots & \vdots & \ddots & \vdots \\
-\frac{\partial f(x_m, y_1, T_{m1})}{\partial y_1} & -\frac{\partial f(x_m, y_2, T_{m2})}{\partial y_2} & \cdots & -\frac{\partial f(x_m, y_m, T_{mm})}{\partial y_m}
\end{bmatrix},
\]

\[
A_{21}(\bar{N}) = \text{diag} \left[ \sum_{j=1}^{m} \frac{\partial f(x_j, y_j, T_{j1})}{\partial x_j} \right],
\]

\[
A_{22}(\bar{N}) = \begin{bmatrix}
-(a_i + r_i) + \frac{\partial f(x_1, y_1, T_{i1})}{\partial y_1} & \frac{\partial f(x_1, y_2, T_{i2})}{\partial y_2} & \cdots & \frac{\partial f(x_1, y_m, T_{im})}{\partial y_m} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial f(x_m, y_1, T_{mi})}{\partial y_1} & \frac{\partial f(x_m, y_2, T_{m2})}{\partial y_2} & \cdots & -(a_i + r_i) + \frac{\partial f(x_m, y_m, T_{mm})}{\partial y_m}
\end{bmatrix}.
\]

A disease can become established if the vector \( y = (y_1, y_2, \ldots, y_m) \) can increase when \( \bar{y} = (0, 0, \ldots, 0) \) and the vector \( x = (x_1, x_2, \ldots, x_m) \) is equal to \( \bar{x} = (N_1, N_2, \ldots, N_m) \). This is equivalent to requiring that the linear system (3) be unstable at the equilibrium point \( \bar{N} \).

The diagonal submatrix \( A_{21}(\bar{N}) \) contains elements that express the effect of infected individuals on susceptible individuals at the equilibrium point \( \bar{N} \). At this point, however, there are no infected individuals. Since we assume that \( \frac{\partial f}{\partial x_i} = 0 \) when \( y_i = 0 \), this submatrix is zero \( [A_{21}(\bar{N}) = 0] \). The total matrix \( A(\bar{N}) \) is therefore decomposable. The set of eigenvalues of \( A(\bar{N}) \) is the union of the subsets of eigenvalues of \( A_{11}(\bar{N}) \) and \( A_{22}(\bar{N}) \). By inspection, the
submatrix $A_{11}(\hat{N})$ contributes eigenvalues with negative real parts. If the equilibrium point $\hat{N}$ is to be unstable, then the submatrix $A_{22}(\hat{N})$ must have at least one eigenvalue with a positive real part. We have established the following theorem:

**Theorem (Hethcote [8])**

A disease will become established in a spatially heterogeneous population if and only if the submatrix $A_{22}(\hat{N})$ of Equation (3) has at least one eigenvalue with a positive real part.

Hethcote [8] first established the above theorem under the assumption of a mass-action law. However, in its present form, the theorem is difficult to interpret. In particular, it is not readily apparent how to establish threshold conditions for heterogeneous populations. We shall use a result from linear algebra to reformulate the theorem in a form which is more amenable to biological interpretation. First notice that the effect of infected individuals on the rate of infection will always be positive, that is,

$$\frac{\partial f(x_i, y_j, T_{ij})}{\partial y_j} |_{\hat{N}} > 0.$$  \hspace{1cm} (4)

This results in the matrix $A_{22}(\hat{N})$ having a special form that permits the use of simple criteria to determine stability (rather than computing eigenvalues). To establish these criteria, we introduce the notion of an $M$-matrix (see Plemmons [13] for a thorough review).

**Definition**

A $k \times k$ matrix $M = (m_{ij})$ ($1 \leq i, j \leq k$) is said to be an $M$-matrix if $m_{ij} \leq 0$ for all $i \neq j$ and if any one of the following equivalent statements is true:

(i) all the principal minors of $M$ are positive;
(ii) all eigenvalues of $M$ have positive real parts;
(iii) $M$ is nonsingular and $M^{-1} \geq 0$;
(iv) there is a vector $u > 0$ such that $Mu > 0$;
(v) there is a vector $v > 0$ such that $M^Tv > 0$.

All of the off-diagonal terms of $A_{22}(\hat{N})$ are nonnegative, that is, $a_{ij} \geq 0$ for $i \neq j$. Considering the matrix $-A_{22}(\hat{N})$, we see from the above definition that all the eigenvalues of $-A_{22}(\hat{N})$ have positive real parts if and only if $-A_{22}(\hat{N})$ is an $M$-matrix. Alternatively, all the eigenvalues of the matrix
$A_{22}(\vec{N})$ have negative real parts if and only if $-A_{22}(\vec{N})$ is an $M$-matrix. This establishes the following corollary to the theorem:

**COROLLARY**

*A disease will become established in a spatially heterogeneous population if and only if $-A_{22}(\vec{N})$ is not an $M$-matrix.*

The strength of this corollary lies in the usefulness of the several equivalent conditions which define an $M$-matrix. Conditions (iv) and (v) allows us to express, in a biologically intuitive fashion, the conditions under which a disease can become established. Condition (iv) when applied to $A_{22}(\vec{N})$ implies that $-A_{22}(\vec{N})$ is not an $M$-matrix if and only if for every $u = (u_1, u_2, \ldots, u_m) > 0$ there exists a subpopulation $i$ such that

$$\sum_{j=1}^{m} u_j \frac{\partial f(x_i, x_j, T_{ij})}{\partial y_j} |_{\vec{N}} > u_i(a_i + r_i).$$

A rough interpretation is that for at least one subpopulation the rate at which susceptible individuals contract the disease must exceed the rate at which infected individuals are removed from the subpopulation through recovery ($r_i$) or death ($a_i$). This intuitive condition, applicable to any number of interacting subpopulations, parallels the threshold condition for the single isolated population.

Similarly, from condition (v) we can derive an inequality which can be interpreted as stating that for at least one subpopulation $i$, the rate at which infected individuals spread the disease among various subpopulations must be greater than the rate at which the infected of subpopulation $i$ are removed. This condition is not as obvious as the previous one, since it does not equate the creation and removal of infecteds in a single subpopulation. This condition is similar to the concept of infectious contact number $\sigma$ of Hethcote [8] and Nold [12], that is, the average number of individuals contacted by an infective during his or her infectious period. Condition (i) of an $M$-matrix allows one to determine with a finite number of arithmetical calculations whether or not a particular set of model parameters satisfies the theorem.

3. **LINEAR RATE OF CONTAGION**

So far we have not specified the functional form of the rate of contagion. In this section we shall explore the relationship between the functional form of the rate of contagion and whether or not a disease can become established in a population.

Let the rate of contagion have the form

$$f(x_i, y_j, T_{ij}) = b(T_{ij}) x_i y_j,$$  \hspace{1cm} (5)
where \( b(T_{ij}) \), the infectibility, is a function of \( T_{ij} \), and \( \sum_{i=1}^{m} T_{ij} = 1 \). If a population \( i \) does not contact any other population, then \( T_{ii} = 1, b(T_{ii}) = b(1) = b_0, \) and \( T_{ij} = 0, b(T_{ij}) = b(0) = 0, i \neq j \). Suppose that the infectibility is linearly proportional to the fraction of time two populations are in contact, that is,

\[
b(T_{ij}) = b_0 T_{ij}.
\] (6)

By the corollary, a disease will become established if and only if the following matrix is not an M-matrix:

\[
-A_22(\hat{N}) =
\begin{bmatrix}
(a_1 + r_1) - b_0 T_{11} N_1 & -b_0 T_{12} N_1 & \cdots & -b_0 T_{1m} N_1 \\
-b_0 T_{11} N_2 & (a_2 + r_2) - b_0 T_{22} N_2 & \cdots & -b_0 T_{2m} N_2 \\
\vdots & \vdots & \ddots & \vdots \\
-b_0 T_{m1} N_m & -b_0 T_{m2} N_m & \cdots & (a_m + r_m) - b_0 T_{mm} N_m
\end{bmatrix}
\] (7)

Property (v) of an M-matrix allows us to establish easily that this is never the case if each population center itself is not capable of supporting the disease. Choosing \( v = (1/N_1, 1/N_2, \ldots, 1/N_m) \), the vector \(-A_22(N)v\) consists of elements of the form

\[
\frac{a_i + r_i}{N_i} - \sum_{j=1}^{m} b_0 T_{ij}.
\] (8)

Since \( \sum_{j=1}^{m} T_{ji} = 1 \), the condition for \(-A_22(\hat{N})\) to be an M-matrix becomes

\[
(a_i + r_i) - b_0 N_i > 0, \quad i = 1, 2, \ldots, m.
\] (9)

This is exactly the condition for determining when a disease cannot become established in an isolated population [7]. Thus, in the case when the infectibility is linearly proportional to the fraction of time subpopulation centers are in contact, the threshold condition for maintenance of a disease in an entire population is identical to the threshold condition for the maintenance of the disease in each of the isolated subpopulation centers. In other words, the disease cannot become endemic in the entire population unless it is endemic in some isolated subpopulation.

This important property of disease in heterogeneous populations is not a consequence of the assumption that the rate of contagion is of "mass action" type as given by (5)–(6). In fact, it is easily established that the same
principle holds when the rate of contagion has the form

\[ f(x, y, T_{ij}) = b_0 T_{ij} x_i g(y_j), \]

where \( g > 0, g(0) = 0, \) and \( g'(x) > 0. \) Capasso and Serio [5] have suggested such an interaction term to explain a cholera epidemic spread in Bari.

4. NONLINEAR RATE OF CONTAGION

If the basic infectibility \( b_0 \) is the same in each subpopulation, we see that the dynamics are the same as if each subpopulation were separate. This is intuitively correct, because in this case, it does not matter whether a susceptible contracts the disease at home or while traveling if the probability of becoming infected is the same everywhere.

For migration to enhance the probability of an outbreak, it is necessary that travel increase the risk of contracting the disease. We have no theoretical or empirical basis for the intuitive concept. Most treatments of the geographical spread of infectious disease involve the migration of individuals between population centers. Thus, the size of each population center, \( N_i, \) does not remain constant. We, however, are assuming that the traveling hosts return to their respective subpopulations, so that we are, in effect, modeling the migration of the disease rather than the migration of the host population. One plausible interpretation of the phenomenon that travel increases the risk of contracting the disease is that visiting individuals actually experience higher contact rates. Traveling to population centers is usually for the purpose of contacting individuals not usually contacted, and there is also increased contact with fellow travelers at transportation and lodging centers. Other factors may also contribute, such as decreased resistance due to the rigors of travel or exposure to slightly different strains of the disease to which the individual is not as immune. While it is possible to incorporate such effects into a mathematical model from first principles, such a refinement is unnecessary to our basic point. We shall incorporate this effect into our model in a more general, phenomenological way. One way in which this can come about is if the infection rate is a concave function of the length of time an individual spends in a given community, rather than in linear function as was assumed above. Figure 1 shows a hypothetical example of such a concave curve which has the necessary features required for an epidemic to occur. The slope of the concave curve differs most strongly from that of the constant infection rate at the origin. This reflects the idea that when an individual arrives in a community, the chance that the disease will be contracted is far greater per unit time than if that individual had been there for a considerable time without contracting the disease.
ENDEMIC DISEASE IN A HETEROGENEOUS POPULATION

As a simple example consider the case of two population centers which cannot support the disease in isolation and in which a fraction of the populations migrate back and forth between centers. Let

$$f(x_i, y_j, T_{ij}) = b(T_{ij})x_i y_j.$$  \hfill (10)

Then the matrix $A_{22}(\hat{N})$ appears as

$$\begin{bmatrix}
N_1 b(T_{11}) - (r_1 + a_1) & N_1 b(T_{12}) \\
N_2 b(T_{21}) & N_2 b(T_{22}) - (r_2 + a_2)
\end{bmatrix}.$$  \hfill (11)

Recall that $T_{11}$ is the time resident susceptible and infecteds of center 1 are in contact, $T_{12}$ is the time susceptibles of center 1 and infecteds of center 2 are in contact, etc. We have $T_{11} + T_{21} = 1$ and $T_{12} + T_{22} = 1$.

According to the corollary, the condition that $y$ increases from zero is that (11) is not the negative of an $M$-matrix. Using property (i) of an $M$-matrix, this is equivalent to the criterion that the determinant of (11) is negative, or

$$(N_1 b(T_{11}) - (r_1 + a_1))(N_2 b(T_{22}) - (r_2 + a_2)) - b(T_{12}) b(T_{21}) N_1 N_2 < 0.$$  \hfill (12)
It is difficult, using this formula, to evaluate directly whether a disease can become established through contact between subpopulation centers. Migration has two effects on (12). As the time that a visitor spends in another subpopulation increases, \( b(T_{11}) \) or \( b(T_{22}) \) decreases, which makes (12) more difficult to achieve. On the other hand, as the visiting times \( T_{12}, T_{21} \) increase, \( b(T_{11}) \) and/or \( b(T_{22}) \) increases, thereby making (12) easier to satisfy. We can reduce the four variables \( b(T_{11}), b(T_{12}), b(T_{21}), b(T_{22}) \) with divergent effects into two new variables, \( \Delta_1, \Delta_2 \), which express the increase in the rate of infection a subpopulation experiences due to migration or contact with other subpopulation centers. These variables may be stated concisely as follows (see Figure 1):

\[
\Delta_1 = b(T_{11}) + b(T_{21}) - b_0 = b_0 T_{11} + \Delta_{11} + b_0 T_{21} + \Delta_{21} - b_0 = \Delta_{11} + \Delta_{21},
\]

\[
\Delta_2 = b(T_{12}) + b(T_{22}) = b_0 T_{12} + \Delta_{12} + b_0 T_{21} + \Delta_{22} - b_0 = \Delta_{12} + \Delta_{22}.
\]

The quantities \( \Delta_1, \Delta_2 \) are nonlinear functions of the \( T_{ij} \)'s, but have the advantage of relating directly to the rate of infection and the stability of equations under study.

Choosing model parameters and a particular form of the infectibility function, we can evaluate (12) with \( \Delta_1 \) and \( \Delta_2 \) to determine the effect of migration on the progress of the disease. Figure 2 summarizes the results of such a calculation for various values of \( b_0 \) and an infection rate function of the form \( b(T) = 2b_0 T/(1 + T) \). Other functional forms would give similar results. Each subpopulation center was assumed to have the same population size \( (N_1 = N_2 = 200) \) and the same rate of removal of infecteds from the infected class \( (r_1 + a_1 = r_2 + a_2 = 0.8) \). If the point specified by \( \Delta_1 \) and \( \Delta_2 \) lies above the hyperbola for the appropriate value of \( b_0 \), the determinant of (11) is negative; that is, (12) is satisfied. If the point lies below the hyperbola, then the determinant is positive. The threshold condition necessary for each subpopulation center to support the disease in the absence of interaction is \( b_0 > 0.004 \). If \( b_0 \) is less than this critical rate, then the disease can become established only if there is sufficient interaction between subpopulation centers. This will occur when \( \Delta_1 \) and \( \Delta_2 \) are large enough for the point \( (\Delta_1, \Delta_2) \) to lie above the appropriate hyperbola.

There are two other important facts that Figure 2 makes clear. First, the shape of the curve separating the stable from the unstable region suggests that this transition occurs when the product \( \Delta_1 \Delta_2 \) exceeds some constant. This requires that both subpopulations must mix with each other for the
Fig. 2. Each curve defines the enhancement in infectibility through mixing that is required for a disease to become endemic in a population of two interacting subpopulations. See text for additional explanation.

disease to become established. Secondly, only a small section of each hyperbola is drawn. This is due to the fact that $\Delta_1$ and $\Delta_2$ have maximum values which depend on the infectibility function $b(T)$ and the magnitude of $b_0$. As a result, there is an infection rate (in the present case $b_0 = 0.003$) below which a disease cannot become established, regardless of the degree of mixing between the subpopulations.

5. DISCUSSION

Several models similar to the one introduced here have been considered by others. Rushton and Mautner [14] present solutions to a special case of a model of a simple epidemic in many communities. Watson [16] developed a stochastic model to evaluate the severity of an outbreak in a population divided into $M$ subpopulations. A deterministic analog of Watson’s model is similar to a special case of the model presented in this paper with linear rates
of contagion and without reproduction. Using computer simulations, he determined that when the total population size became large, the probability of a major outbreak involving most subpopulations increased to certainty. The population size required, however, was large enough to ensure that most subpopulations were larger than the threshold size required to support an epidemic in isolation. This is in concordance with the analytical results presented in the previous section.

Lajmanovich and Yorke [9] have analyzed equations similar to (1) and (2) with linear infection rate functions analogous to (5). They show that, depending on parameter values, either the disease will disappear from the population or the disease will be endemic with the infective and susceptible levels approaching unique constant values, regardless of the initial conditions. That is, they establish uniqueness and global stability conditions for the equilibrium. This result holds for epidemic models where the population size in each center may be assumed to be constant and the rate of infection is proportional to the product of the susceptible and infected populations. Specifically, their result holds for both forms of the infection rate function we have discussed [Equations (5), (6)]. However, their model is an S-I-S model which does not have a recovered class, so their results do not apply directly to our model.

The theorem and corollary of Section 3, on the other hand, may be applied to more general epidemic models than those presented above and in [8], [12] and [14]. For example, Capasso and Serio [5] introduce an interaction term in which the dependence upon the number of infectives incorporates a nonlinear bounded function which may express saturation or psychological effects. In this sense, the theorem of this report has wider application than the theorem of Lajmanovich and Yorke [8], even though our theorem determines the conditions for local instability of the population equilibrium without infected individuals.

The introduction of spatial heterogeneity into epidemic models does not by itself alter the general qualitative behavior found in single population models with homogeneous mixing. Bailey [1] points out that deterministic models incorporating spatial heterogeneity display the same type of asymptotic behavior as deterministic models with homogeneous mixing. Our model is not an exception. Computer simulations of the model (1)–(2) with \( b(T) = 2b_0 T/(1 + T) \) show that the nonzero equilibrium is globally stable when (12) is satisfied. Because of this difficulty, most investigators have resorted to stochastic models or time delays to describe the periodic behavior exhibited by such diseases as measles and influenza [2]. However, there can be no doubt that nonhomogeneous mixing affects the quantitative behavior of epidemics. The current work demonstrates a pronounced effect of spatial heterogeneity on threshold values. This effect depends on the nature of the interaction between infected and susceptible populations. As a further exam-
ENDEMIC DISEASE IN A HETEROGENEOUS POPULATION

ple, Rushton and Mautner [14], Watson [16], and others have discussed the effect of nonhomogeneous mixing between populations on disease dynamics during the fadeout process.

There is a formal similarity between models for a directly communicated disease with two interacting host populations and parasitic disease which involve two populations (intermediate and definitive hosts). Parasitic diseases are characterized by the existence of an intermediate host which transfers, directly or indirectly, the parasites to the definitive host, where sexually reproducing forms of the parasite accumulate. The definitive host in turn infects the intermediate host to complete the cycle. Several deterministic models describing this general process have been analyzed for the presence of threshold behavior [2, 6, 10]. These models are special cases of our model for two population centers. Since host populations cannot infect themselves directly, $b_{11} = b_{22} = 0$. The threshold condition which must be met for the disease to become endemic is easily derived from (12); it is

$$b(T_{12}) b(T_{21}) > \frac{(r_1 + a_1)(r_2 + a_2)}{N_1 N_2}.$$  

(15)

The details of a particular parasitic disease may be quite complicated, involving several forms of the parasite in each host as well as the possible multiplication of parasites outside the host populations. Various complications can be added to the equations to account for realistic aspects of a disease that are considered important without altering the applicability of the theorem and corollary of the previous section. For example, the general analysis presented in this paper is applicable to models incorporating age-specific interactions or interactions between multiple definitive and alternate host populations [15].

REFERENCES