How may infection-age dependent infectivity affect the dynamics of HIV/AIDS?

by

Horst R. Thieme
and
Carlos Castillo-Chavez

BU-1102-MA

July 1991

How may infection-age dependent infectivity affect the dynamics of HIV/AIDS?

Horst R. Thieme
Department of Mathematics
Arizona State University
Tempe, AZ 85787-1804

Carlos Castillo-Chavez
Biometrics Unit, Center for
Appl. Math., and Population
and Development Program
Cornell University
Ithaca, NY 14853-7801

Abstract

Epidemiological and behavioral factors crucial to the dynamics of HIV/AIDS include long and variable periods of infectiousness, variable infectivity, and the processes of pair formation and dissolution. Most of the recent mathematical work on AIDS models has concentrated on the effects of long periods of incubation and heterogeneous mixing in the transmission dynamics of HIV. In this paper we explore the role of variable infectivity in combination with a variable incubation period in the dynamics of HIV transmission in a homogeneously mixing population. We keep track of an individual's infection-age, that is of the time which has passed since infection, and assume a nonlinear functional relationship between mean sexual activity and the size of the sexually active population which saturates at high population sizes. We identify a basic reproductive number R_0 and show that the disease dies out if $R_0 < 1$, whereas if $R_0 > 1$ the disease persists in the population and the incidence rate converges to or oscillates around a uniquely determined non-trivial equilibrium. Though we find conditions for the endemic equilibrium to be locally asymptotically stable, undamped oscillations cannot be excluded in general and may occur in particular if the variable infectivity is highly concentrated at certain parts of the incubation period. Whether undamped oscillations can also occur for the reported one early peak and one late plateau of infectivity observed in *HIV*-infected individuals has to be a subject of future numerical investigations.

Abbreviated title: HIV/AIDS and variable infectivity

Key words: HIV, AIDS, variable infectivity, infection-age, class-age, population-size-dependent contact rate, basic reproductive number, endemic equilibrium, disease-free equilibrium, disease-persistence, stability change, undamped oscillations, characteristic equation, abstract differential equation, Volterra integral equation

AMS-Classification: 34G20, 35B40, 45D05, 92A15

Introduction

Most epidemiological models for the transmission of infectious diseases have assumed that all infectious individuals are equally infectious during their period of infectivity. This assumption has proved to be reasonable in the study of the dynamics of communicable diseases such as influenza (see Castillo-Chavez et al. 1988, 1989 e, and references therein) or in the study of sexually transmitted diseases such as gonorrhea (see Hethcote and Yorke 1984, and references therein).

In their classical work, Kermack and McKendrick present and analyze both epidemic (McKendrick, 1926, Kermack&McKendrick, 1927) and endemic models (Kermack&McKendrick, 1932, 1933) where the infectivity is allowed to depend on infection-age (that is the time that has passed since the moment of infection). The Kermack&McKendrick model in this general form was largely neglected until the 1970's (Reddingius, 1971, Waltman, 1974, Hoppensteadt, 1974, 1975).

The interest in the role played by variable infectivity in disease transmission dynamics has been considerably increased by the HIV/AIDS epidemic. The early infectivity experiments reported in Francis et al. (1984) and the measurements of HIV antigen and antibody titers (Salahuddin et al., 1984, Lange et al., 1986, Pedersen et al., 1987) have supported the possibility of an early infectivity peak (a few weeks after exposure) and a late infectivity plateau (one year or so before the onset of "full-blown" AIDS) for HIV-infected individuals. Presently one cannot necessarily identify virus titers and infectivity levels and one has to be aware that alternative patterns of HIV infection have been found (Imagawa et al., 1989, Haseltine, 1989). Nevertheless, these findings are reason enough to study the possible effects of variable infectivity on the transmission dynamics of HIV/AIDS.

In this paper we pursue the question whether variable infectivity can cause undamped oscillations in AIDS incidence. This possibility is of importance for the interpretation of incidence data, namely as to whether a decline in disease incidence can be attributed to effective behavioral changes and preventive measures or may be partially due to an oscillatory behavior inherent in the disease dynamics.

Early analyses of models without variable infectivity (see Castillo-Chavez et al., 1989 a, b, c ,d) have shown that long and variable incubation periods (time from infection to appearance of symptoms) of HIV-infection cannot excite oscillations on their own (at least not by a *Hopf* bifurcation). Numerical simulations of models that incorporate variable infectivity (see May and Anderson, 1989, Hyman and Stanley 1988, 1989; Blythe and Anderson 1988 b, Anderson et al., preprint) demonstrate that the initial (transient) dynamics are very sensitive to the shape and timing of the first infectivity peak, but that the long time dynamics show the same qualitative behavior (even in the presence of heterogeneity in sexual behavior): a steady approach to an endemic equilibrium.

The model in this paper extends the model of Castillo-Chavez et al. (1989 b) by incorporating infection-age dependent infectivity to study analytically whether convergence to the endemic equilibrium is a general feature or whether undamped oscillations can occur under specific circumstances.

Various mechanisms have been found to cause undamped oscillations in time-autonomous endemic models (see Hethcote et al., 1981, and Hethcote and Levin, 1989, for surveys). One of the most commonly found reasons is return of infectives into the susceptible class with or without having spent a period of temporary immunity. Anderson et al. (1981) have found sustained oscillations in a fox rabies model, which were generated by the combined effects of a rapid turnover of the fox population and the relatively long latency period and the high fatality of fox rabies. Models showing similar phenomena have been considered by Brauer (1989, to appear a,b) and Pugliese (preprint). Liu et al. (1986, 1987) have shown, in their work on influenza, that generalized nonlinear incidence rates can also generate sustained oscillations. The work by Chastillo-Chavez et al. (1988, 1989 e) and Andreasen (thesis, 1989) on influenza strongly suggests that the interaction between related multiple viral strains of influenza type A, the host immune system (cross-immunity), and age-dependent host mortality are needed to generate sustained oscillations. Endemic models which incorporate infection-age have so far exhibited sustained oscillations only when infective individuals are allowed to return to the susceptible class; it depends on the form of the infection-age dependent infectivity curve and the distribution of the length of the immunity period whether or not undamped oscillations actually occur (see Diekmann and Montijn, 1982, Diekmann and van Gils, 1984, Gripenberg, 1980, 1981, and Hethcote and Thieme, 1985).

All these mechanisms found to generate sustained oscillations do not operate in the case of *HIV* dynamics. Most of the models described above (the rabies model and the models considered by Brauer and Pugliese are the exceptions) assume that the disease is essentially non-fatal and the population size is constant, assumptions that are not realistic in HIV modeling. Even if one assumes a constant recruitment rate into the sexually active population (as we do), the population size will vary with time due to the disease fatalities. This makes it necessary to model the functional relationship between the per capita number of sexual contacts $C = C(T)$ and the number of sexually active individuals T. For sexually transmitted diseases it seems reasonable to assume a saturation effect for partner acquisition, namely that $C(T)$ becomes largely independent of T if the population size T is large.

In our model the possible occurrence of undamped oscillations will depend mainly on the interaction of the following epidemiological entities: the saturation of the functional relationship C, the length distribution of the sexual activity period of exposed individuals, and the distribution of infectivity over that period. Undamped oscillations (caused by

an unstable endemic equilibrium) can be ruled out if the probability that an infected individual is still sexually active is a convex function of infection-age or if the infectivity is sufficiently evenly distributed over the activity period (as it is already suggested by the findings of Castillo-Chavez et al. (1989 a, b, c, d) or if $C(T)/T$ is close to a constant, i.e. there is not saturation in partner acquisition. Undamped oscillations may occur if all the following conditions are satisfied simultaneously:

(i) The probability that an infected individual is still sexually active is sufficiently far away from being a convex function of infection age.

(ii) There is sufficient saturation in partner acquisition, i.e. the number $C(T)$ of actual partners per capita is largely independent of slight changes in the number of available partners T.

(iii) The period of sexual activity is not too short for infected individuals in relation to uninfected ones.

(iv) The fraction of infected individuals in the sexually active population is neither too low nor too high.

(v) The infection-age-distributed infectivity is highly concentrated at specific parts of the incubation period.

The conditions (i) to (iii) appear to be realistic for AIDS. Condition (v) is rather extreme though several infectivity peaks are allowed (the timing of which has to be interrelated in a certain way which presumably is not consistent with reality). We emphasize that this paper completely relies on analytical methods. Future numerical work has to show whether or not undamped oscillations also occur for infection-age distributions with an early peak and a late plateau.

In other respects we can show the same phenomena as for the model by Castillo-Chavez et al. (1989 b) with constant infectivity. We can identify a basic reproductive number R_0 in terms of the model parameters such that, for $R_0 < 1$, the disease dies out and, for $R_0 > 1$, persists in the population. In the latter case there is a unique endemic equilibrium which is locally asymptotically stable for R_0 slightly larger than 1, but which may lose stability if R_0 increases. Even if unstable, the endemic equilibrium may be some indication of the severity of the disease because (as we show) the incidence rate, for example, oscillates around its equilibrium value.

The work of Castillo-Chavez et al. (1989 c, d), Huang (thesis), and Huang et al. (in press) has shown that multiple endemic equilibria may exist in epidemic models (of the type illustrated in this article) for heterogeneously mixing populations, even if the infectivity does not depend on infection age.

The paper is organized as follows: Section 1 introduces our model with infection-age dependent infectivity, Section 2 discusses the existence of endemic stationary states

(in relation to the basic reproductive number), Section 3 relates the basic reproductive number to disease extinction or persistence. Section 4 discusses the epidemiological content of our stability results, while Section 5 and 6 establish the validity of our claims. An example is studied in Section 7, and Section 8 discusses the significance of our results and projects future work. Most of the results of this paper were stated, without proofs and not completely correct, in Thieme and Castillo-Chavez (1989).

1. A model with infection-age dependent infectivity

The transmission dynamics of *HIV* in a homogeneously mixing male homosexual population is modeled through the incorporation of the following ingredients:

- A nonlinear functional relationship between mean sexual per capita activity and the size of the sexually active population, T. We assume that it increases linearly for small population sizes while saturating for large values of T.
- A stratification of the infected part of the sexually active population according to infection age, i.e., time since the moment of being infected
- An infection-age-dependent rate of leaving the sexually active part of the population by force of the disease
- An infection-age dependent infectivity.

The model considered in this paper shares the first three features with the models considered by Castillo-Chavez et al. (1989 a, b, c, d) though the stratification according to infection age is not explicit there. The key modification, infection-age dependent infectivity, has been added in order to study its effect in combination with the other mechanisms. The model does not include heterogeneities other than infection-age dependent infectivity. By restricting itself to the homosexual part of a population which is replenished by constant recruitment, it does not reflect the joint effects of *HIV* dynamics and the demographic dynamics of the population (see May and Anderson, 1989, May et al., 1988, 1989, Busenberg et al., to appear).

We divide the homosexually active population under consideration into three groups: S (uninfected, but susceptible), I (*HIV* infected with hardly any symptoms), and A (fully developed *AIDS* symptoms). A-individuals are assumed to have been sexually inactivated by the disease. Individuals that are still sexually active (S and I) are supposed to choose their partners at random (although the rate of partnership change depends on the size of the active population $T = S + I$.)

Further t denotes time, whereas τ denotes time since the moment of being infected, i.e., infection-age. As time unit we choose the average length of the period of sexual activity for healthy individuals. We assume that individuals are recruited into the sexually active population at a constant rate Λ and that the length of the sexually active period is exponentially distributed such that healthy individuals become sexually inactive at a constant rate μ. As we have chosen the average length $1/\mu$ of the activity period to be 1, $\mu = 1$. Infected individuals with infection-age τ become sexually inactive by force of the disease (and enter the A class) at a rate $\alpha(\tau)$; consequently the proportion of those individuals that are still sexually active given that they were infected τ time units ago is

given by

$$P_{\alpha+1}(\tau) = \exp\left(-\tau - \int_0^\tau \alpha(\rho)d\rho\right).$$

We stratify the infected part of the population, I, according to age of infection:

$$I(t) = \int_0^\infty i(t,\tau)d\tau$$

where $i(t,\cdot)$ denotes the infection-age density at time t.

The proportion of sexually active infected individual with infection-age τ in the age-interval $[\tau, \tau + \Delta\tau]$ is

$$\frac{i(t,\tau)}{T(t)}\Delta\tau$$

with

$$T = S + I$$

being the size of the sexually active population. We assume that a typical susceptible (under homogeneous mixing everybody is typical) contracts the disease from an infected partner with age of infection τ at a mean risk $\lambda(\tau)$. Consequently the per capita rate at which susceptible individuals are infected at time t (given that they have a sexual contact at that time) is provided by

$$\frac{W(t)}{T(t)},$$

where

$$W(t) = \int_0^\infty \lambda(\tau)i(t,\tau)d\tau.$$

We assume that $C = C(T)$, the mean number of sexual contacts that a typical individual has per unit of time, is a function of the size of the sexually active population $T = S + I$.

Combining these considerations, we arrive at the following expression for the incidence rate of infections (number of new cases of infection per unit of time), B:

$$B(t) = C\Big(T(t)\Big)S(t)\frac{W(t)}{T(t)}.$$

The complete dynamical model with infection-age dependent infectivity can now be formulated as follows:

$$\frac{d}{dt}S(t) = \Lambda - B(t) - S(t), \tag{1}$$

$$(\frac{\partial}{\partial t} + \frac{\partial}{\partial \tau})i(t,\tau) = -\Big(1 + \alpha(\tau)\Big)i(t,\tau), \tag{2}$$

$$i(t,0) = B(t) = S(t)C\Big(T(t)\Big)\frac{W(t)}{T(t)}, \tag{3}$$

where

$$T = I + S \tag{4}$$

$$I(t) = \int_0^\infty i(t,\tau)d\tau \tag{5}$$

$$W(t) = \int_0^\infty \lambda(\tau)i(t,\tau)d\tau \tag{6}$$

$$\frac{d}{dt}A(t) = \int_0^\infty \alpha(\tau)i(t,\tau)d\tau - \nu A(t),$$

where ν denotes the rate at which an individual with fully developed AIDS symptoms dies from the disease.

In spite of the fact that the A equation — for the number of individuals with fully developed AIDS symptoms (that are supposed to play no further role in the dynamics of the disease) — is not needed to make the model well-posed, we include it here because it provides an entity that can be directly compared to data.

Note that, in contrast to earlier models by Anderson and May (1987), Blythe and Anderson (1988 a), and Castillo-Chavez et al. (1989 a,b) this model does not assume that, at the moment of infection, an individual is determined to follow either a severe or a mild course of the disease. By assuming that $\int_0^\infty \alpha(\tau)d\tau < \infty$, however, this model, albeit by a different mechanism, can take into account the possibility that not all individuals will develop "full-blown" *AIDS*.

Throughout this paper we make the following mathematical assumptions:

$\alpha(\tau)$ is a non-negative measurable function, $\lambda(\tau)$ is a non-negative bounded function of infection age, and $C(T)$ is a non-decreasing function of T. Further we assume that $C(T)$ is continuously differentiable and $C(T) > 0$ whenever $T > 0$.

Later we will also assume that

$$M(T) = \frac{C(T)}{T}$$

is a non-increasing function of T, i.e., C increases in a sublinear way that reflects a saturation effect.

There are various ways of handling problem (1), ..., (6) each of which has its definite advantages. One may deal with this system as an abstract ordinary differential equation, see Thieme (1990), in particular section 7. This approach provides a dynamical system in terms of S and i and is useful to prove instability of equilibria and persistence of solutions. The same dynamical system can be obtained by integrating (2) along characteristic lines

(Webb, 1985). Or one uses integration along characteristic lines to reduce (1), ..., (6) to the following system of integral equations:

$$S = \Lambda - B * P_1 + f_1, \tag{7}$$

$$I = B * P_{\alpha+1} + f_2, \tag{8}$$

$$W = B * Q + f_3, \tag{9}$$

$$B = SM(S + I)W. \tag{10}$$

Here we have used the following notation:

$$P_\alpha(\tau) = \exp\left(-\int_0^\tau \alpha(s)ds\right) \tag{11}$$

$$Q(\tau) = \lambda(\tau)P_{\alpha+1}(\tau) \tag{12}$$

$$(B * P)(t) = \int_0^t B(t-s)P(s)ds \tag{13}$$

$$f_1(t) = (S(0) - \Lambda)e^{-t} \tag{14}$$

$$f_2(t) = \int_t^\infty i(0, \tau - t)\frac{P_{\alpha+1}(\tau)}{P_{\alpha+1}(\tau - t)}d\tau \tag{15}$$

$$f_3(t) = \int_t^\infty i(0, \tau - t)\lambda(\tau)\frac{P_{\alpha+1}(\tau)}{P_{\alpha+1}(\tau - t)}d\tau \tag{16}$$

$$M(T) = \frac{C(T)}{T}. \tag{17}$$

P_1 and $P_{\alpha+1}$ are defined in analogy to P_α. One easily observes that

$$f_j(t) \to 0, \qquad t \to \infty, \quad j = 1, 2, 3. \tag{18}$$

Some of the expressions just defined above have an intuitive biological interpretation. For example, $P_1(s) = e^{-s}$ gives the proportion of healthy individuals that remain sexually active s time units after having entered the active population. $P_{\alpha+1}(\tau)$ can be identified with the proportion of those infected individuals of infection-age τ that are still sexually active. $Q(\tau) = \lambda(\tau)P_{\alpha+1}(\tau)$ in Equation (12) can be interpreted as the effective infectivity of an individual with infection age τ.

By substituting (10) into (7), (8), (9) we obtain a system of Volterra integral equations of convolution type for which a well-developed theory is available (see Miller, 1971, or Gripenberg et al., 1990). Alternatively one can substitute (7), (8), (9) into (10) and obtain a single integral equation which is not of common Volterra type.

From Equation (1) we realize that $S(t)$ remains positive (non-negative) if $S(0)$ has the corresponding property. It is then easily checked from the various equations that

non-negativity is preserved under the solution flow. Integrating Equation (2) over τ and combining it with (1) and (4) yields the differential inequality

$$\frac{d}{dt}T \leq \Lambda - T, \tag{19}$$

that provides us with the a priori estimate

$$S(t), I(t) \leq T(t) = S(t) + I(t) \leq \Lambda + (T(0) - \Lambda)e^{-t}. \tag{20}$$

Using the theory in Webb (1985) or Thieme (1990) or applying standard fixed point arguments to (7), ..., (10) one easily shows that the model is well-posed, i.e., there is a unique non-negative solution for non-negative initial conditions. Furthermore, the mathematical theory shows that the solutions depend continuously on the initial conditions and the functions S, I, W, B are continuous and satisfy the estimate given by the inequality (20).

2. Stationary states and the basic reproductive number

In this section we concentrate on the study of the existence of stationary states, that is, equilibria or time-independent solutions. These special solutions are important because they are candidates for the asymptotic behaviour of the disease dynamics. It will turn out that there are only two equilibria: the infection-free state and the endemic state.

Until recently, it was common belief that most epidemic models had at most these two equilibria. Earlier studies (see Hethcote and Yorke, 1984, Castillo-Chavez et al., 1988, 1989e, and reference therein) supported this conjecture even for heterogeneously mixing populations. Recent studies (see Castillo-Chavez et al. 1989c,d; Huang, thesis, and Huang et al., in press) have shown, however, that multiple endemic equilibria exist.

The existence of endemic equilibria is usually intimately connected to the basic reproductive number, R_0, which can be determined in terms of the model parameters. We will show in this section that the disease-free state is the only equilibrium if $R_0 < 1$. In the next section we will show that in this case the disease becomes extinct. When $R_0 > 1$, there exists a unique endemic equilibrium (as we will see in this section) which is locally asymptotically stable provided that R_0 is slightly larger than 1, but which may lose stability if R_0 increases (see Section 4). Even if unstable the endemic equilibrium provides us with an indicator of the severity of the disease because, as we show in Section 3, the incidence rate fluctuates around its endemic equilibrium value. So it can serve as a starting point for the development and evaluation of control measures — an approach that has been useful in the management of other diseases, specifically of gonorrhea (see Hethcote and Yorke, 1984, and references therein).

Clearly, the system (1), ..., (6) always has the disease-free equilibrium

$$S_0 = \Lambda,\ I_0 = 0,\ W_0 = 0,\ B_0 = 0,\ i_0 = 0. \tag{21}$$

To determine the existence of an endemic equilibrium of (1), ..., (6) we have to look for solutions of the following nonlinear system of algebraic equations:

$$S^* = \Lambda - B^* \tag{22}$$

$$I^* = B^* \hat{P}_{\alpha+1}(0) \tag{23}$$

$$W^* = B^* \hat{Q}(0) \tag{24}$$

$$B^* = \frac{S^*}{T^*} C(T^*) W^*, \quad T^* = S^* + I^*. \tag{25}$$

(23) and (24) are obtained by integrating (2) for time-independent i and substituting the result into (5) and (6) respectively. Alternatively one can consider time-independent

solutions of the limiting equations (for $t \to \infty$) associated with (7),...,(10). (Note (18).) In (23), (24) we have used the *Laplace* transform notation

$$\hat{Q}(z) = \int_0^\infty e^{-za} Q(a)da, \tag{26}$$

$$\hat{P}_{\alpha+1}(z) = \int_0^\infty e^{-za} P_{\alpha+1}(a)da. \tag{27}$$

We substitute (24) into (25) and divide by B^* (which is supposed to be positive, as otherwise the equilibrium is not endemic):

$$1 = \frac{S^*}{T^*} C(T^*)\hat{Q}(0), \quad T^* = S^* + I^*. \tag{28}$$

It will be convenient to formulate the endemic equilibrium equation in terms of the fraction of infected individuals

$$\xi = \frac{I^*}{T^*}. \tag{29}$$

Obviously

$$\frac{S^*}{T^*} = 1 - \xi.$$

Further, by (22), (23), and (29),

$$T^* = S^* + I^* = \Lambda - B^* + I^* = \Lambda - \left(\frac{1}{\hat{P}_{\alpha+1}(0)} - 1\right) I^*$$

$$= \Lambda - \left(\frac{1}{\hat{P}_{\alpha+1}(0)} - 1\right) \xi T^*.$$

We solve this equation for T^*:

$$T^* = \frac{\Lambda}{1 + \left(\frac{1}{\hat{P}_{\alpha+1}(0)} - 1\right)\xi}. \tag{30}$$

We substitute $S^*/T^* = 1 - \xi$ and (30) into (28). This results in the following reformulation of Equation (28) in terms of ξ:

$$1 = (1-\xi)C\left(\frac{\Lambda}{1 + \left(\frac{1}{\hat{P}_{\alpha+1}(0)} - 1\right)\xi}\right)\hat{Q}(0). \tag{31}$$

Alternatively we can also consider

$$1 = \frac{\Lambda - B^*}{\Lambda - B^*\left(1 - \hat{P}_{\alpha+1}(0)\right)} C\left(\Lambda - B^*\left(1 - \hat{P}_{\alpha+1}(0)\right)\right)\hat{Q}(0) \tag{32}$$

which we obtain from (28) by substituting $T^* = S^* + I^*$ and (22), (23).

We realize from (31) that the existence of a unique endemic equilibrium is intimately connected to the properties of $C(T)$. Since $C(T)$ is a monotone non-decreasing function of T and $1 > \hat{P}_{\alpha+1}(0)$, the right hand side of (31) is strictly decreasing in $\xi \geq 0$ such that there is at most one solution $\xi > 0$. For the same reason we can conclude that there is no solution $\xi > 0$ if the right hand side of (31) is smaller than or equal to 1 for $\xi = 0$. We notice that the right hand side of (31) is 0 if $\xi = 1$. The intermediate value theorem implies that there is a solution $\xi, 0 < \xi < 1$ of (31), if the right hand side of (31) is strictly larger than 1 for $\xi = 0$. We conclude that the right hand side of (31), evaluated for $\xi = 0$, plays a crucial role for the existence of an endemic equilibrium. Hence we define

$$R_0 = C(\Lambda)\hat{Q}(0). \tag{33}$$

As we see from (33), R_0 can epidemiologically be interpreted as the total average number of secondary cases an infective individual can produce if it is introduced into the disease-free population (at its equilibrium size, Λ). ($\hat{Q}(0) = \int_0^\infty Q(\tau)d\tau$ is the total infectivity of an average infective individual.) R_0 is called the *basic reproductive number* of the disease.

Consequently we have proved the following threshold result for the existence of an endemic equilibrium:

Theorem 1. *If $R_0 \leq 1$, there exists only the disease-free equilibrium. If $R_0 > 1$, there is a unique endemic equilibrium.*

3. Disease extinction or persistence? The basic reproductive number once more

Theorem 1 is a static result that does not provide us with a relation between basic reproductive number and the actual disease dynamics. It only gives us information about the existence of a very special state in which the disease persists. In this section we relate the basic reproductive number to the extinction and persistence of the disease.

Mathematically we use both the model formulation as an abstract differential equation and as a system of Volterra integral equations of convolution type. For exploiting the second formulation Fatou's lemma will be an important tool. The following notation will be useful. For a bounded real-valued function f defined on $[0,\infty)$ we set

$$f_\infty = \liminf_{t\to\infty} f(t), \quad f^\infty = \limsup_{t\to\infty} f(t).$$

The following theorem connects the basic reproductive number to the extinction of the disease.

Theorem 2. *Let $R_0 < 1$. Then the disease free equilibrium is globally attractive. In particular we have*

$$B(t), V(t), W(t) \to 0, \quad S(t) \to \Lambda \quad \text{for} \quad t \to \infty.$$

Proof: We use the model formulation (7),...,(10). First we apply Fatou's lemma to (10) and use the estimate (20) and the assumption that C is non-decreasing:

$$B^\infty \le C(\Lambda) W^\infty.$$

Next we apply Fatou's lemma to (9):

$$W^\infty \le B^\infty \hat{Q}(0).$$

We substitute the second inequality into the first and use the definition of R_0 in (33):

$$B^\infty \le R_0 B^\infty.$$

As $R_0 < 1$, B^∞ has to be 0, i.e., $B(t) \to 0, t \to \infty$. The remaining parts of our assertion now follow by applying Lebesgue's theorem of dominated convergence to (7), (8), (9).

For our model it is not possible in general to obtain a global convergence result if $R_0 > 1$. One can show, however, that if a trajectory is not attracted to the endemic equilibrium, it has to oscillate around it.

From now on we assume that $M(T) = \frac{C(T)}{T}$ is a monotone non-increasing function of T.

Theorem 3. *Let $R_0 > 1$. Then the following holds:*

a) $$B_\infty \leq B^\star, I_\infty \leq I^\star.$$

b) Let $\lambda(a)$ be strictly positive on a non-empty open interval. Let $a_+ \leq \infty$ be the smallest $\tilde{a}$ such that $\lambda(a) = 0$ for a.a. $a \geq \tilde{a}$ and assume that

$$\int_0^{a_+} I(0,a)da > 0.$$

Then

$$B^\infty \geq B^\star, I^\infty \geq I^\star.$$

In the proof of Theorem 3 we will use the following inequality several times:

Lemma. *For $0 < S_1 < S_2, I_1 \geq I_2 \geq 0$,*

$$S_1 M(S_1 + I_1) < S_2 M(S_2 + I_2).$$

Proof: As C is non-decreasing we have

$$S_1 M(S_1 + I_1) = \frac{S_1 C(S_1 + I_1)}{S_1 + I_1} < \frac{S_2 C(S_2 + I_1)}{S_2 + I_1} = S_2 M(S_2 + I_1).$$

As $M(T) = C(T)/T$ is non-increasing, the assertion follows.

Proof of Theorem 3: a) From (10):

$$B^\infty \leq \limsup_{t\to\infty} \frac{S(t)C\Big(S(t) + I(t)\Big)}{S(t) + I(t)} W^\infty.$$

By the Lemma,

$$B^\infty \leq \frac{S^\infty C(S^\infty + I_\infty)}{S^\infty + I_\infty} W^\infty.$$

We apply Fatou's lemma to (9):

$$W^\infty \leq B^\infty \hat{Q}(0).$$

We substitute this inequality into the previous one and divide by B^∞ which we can assume to be strictly positive because otherwise $B_\infty \le B^\infty = 0$ and the proof is finished:

$$1 \le \frac{S^\infty C(S^\infty + I_\infty)}{S^\infty + I_\infty}\hat{Q}(0). \qquad (\heartsuit)$$

We apply Fatou's lemma to (7) and (8) and use (18):

$$S^\infty \le \Lambda - B_\infty, \quad I_\infty \ge B_\infty \hat{P}_{\alpha+1}(0).$$

We now suppose that $B_\infty > B^*$ and derive a contradiction. First, by (22) and (23), $S^\infty < S^*, I_\infty > I^*$. By the Lemma, we can substitute these inequalities into ($\heartsuit$):

$$1 < \frac{S^*C(S^* + I^*)}{S^* + I^*}\hat{Q}(0) = \frac{S^*C(T^*)}{T^*}\hat{Q}(0).$$

But, by (28), the right hand side of this inequality is 1, a contradiction.

b) From the Lemma we have

$$\liminf_{t\to\infty} S(t)M(T(t)) \ge S_\infty M(S_\infty + I^\infty). \qquad (\diamondsuit)$$

We apply Fatou's lemma to (7) and (8) and use (18):

$$S_\infty \ge \Lambda - B^\infty, \quad I^\infty \le B^\infty \hat{P}_{\alpha+1}(0).$$

We suppose that $B^\infty < B^*$ and derive a contradiction. First we obtain from the last inequality and (22), (23) that

$$S_\infty > S^*, I^\infty < I^*.$$

By the Lemma we can substitute these inequalities into ($\diamondsuit$) and obtain

$$\liminf_{t\to\infty} S(t)M(T(t)) > S^*M(S^* + I^*).$$

Hence we find $\tilde{s} \ge 0$ and $\beta > S^*M(S^* + I^*)$ such that

$$S(t+s)M(T(t+s)) \ge \beta, \quad s \ge \tilde{s}, t \ge 0.$$

By (28),

$$\beta\hat{Q}(0) > 1.$$

From (10) and (9),

$$B(t+s) \ge \beta\left(\int_0^t B(t+s-a)Q(a)da + \int_0^s B(s-a)Q(a+t)da\right)$$

for $t \geq 0, s \geq \tilde{s}$. It follows from the assumption in b) that the initial infectives produce secondary cases. Actually one can show that there is some $\tilde{t}$ such that

$$B(t) > 0 \text{ for } \quad t \geq \tilde{t}.$$

Hence by choosing s large enough we can achieve that

$$g(t) = \beta \int_0^s B(s-a)Q(a+t)da > 0 \text{ for some } t > 0.$$

We fix s and set $u(t) = B(t+s)$. This way we obtain the following renewal inequality for u:

$$u(t) \geq \beta \int_0^t u(t-a)Q(a)da + g(t)$$

with

$$\beta\hat{Q}(0) > 1, g \geq 0, g \not\equiv 0.$$

We now apply a standard comparison argument for Volterra integral equations — see Gripenberg et al., 1990, p. 344, e.g. — and the celebrated renewal theorem — formulated by Sharpe and Lotka (1911) and first proved rigorously by Feller (1941). (See Webb (1985), theorem 4.10, for a proof and further references.) This yields that $u(t)$ and hence $B(t)$ tend to infinity as $t \to \infty$, in contradiction to our assumption $B^\infty < B^*$.

Analogous statements as for B can now be derived for S, I, W by using Fatou's lemma and the equilibrium equations.

Theorem 3 b) implies that the disease weakly persists in the population (independently of the initial condition except of the requirement that there are secondary cases) if the basic reproductive number strictly exceeds 1. It does not settle the questions, however, whether I, the total number of infected individuals, will be bounded away from zero if $R_0 > 1$ and whether this bound does or does not depend on the initial conditions. In order to address these questions — the strong persistence or even uniform persistence of the disease — we switch from the Volterra integral formulation (7),...,(10) to the abstract differential equation formulation (1),...,(6) the solutions of which induce a dynamical system (as follows from the theory developed by Webb (1975) or the theory developed by Thieme (1990)). We use the persistence theory which Hale and Waltman (1989) elaborated for infinite-dimensional dynamical systems, in particular their Theorem 4.2. In checking the assumptions of this theorem we use the terminology by Hale and Waltman. Our state space X is given by elements (S, i) where S is a non-negative number and i a non-negative integrable function on $[0, a_+)$ where a_+ is the number specified in our Theorem 3 b). We choose the "boundary" ∂X of our state space to be formed by the elements of the form

$(S,0)$. In our case the flow on ∂X is extremely simple: all solutions in ∂X converge to the state $(\Lambda,0)$. This is a much stronger property than assumption (iv) in Theorem 4.2. It follows from our Theorem 3 b) that $I^\infty > I^\star$. This means in particular that the distance from ∂X to any solution starting outside ∂X does not converge to 0. Thus assumption (4.2) by Hale and Waltman is satisfied. From our estimate (20) we see that the dynamical system induced by (1),...,(6) has a bounded attractor, i.e. it is point dissipative (assumption (ii) in Theorem 4.2). From (20) we realize as well that the orbit of any bounded set is bounded (assumption (iii) in Theorem 4.2). Assumption (i), that the dynamical system is asymptotically smooth (i.e. the trajectory of every forward invariant bounded set is attracted by a compact set), can be proved in the same way as Proposition 3.16 by Webb (1985).

Theorem 4. *Let $\lambda(a)$ be positive on a non-empty open interval. Further let a_+ be the smallest $\tilde{a}$ such that $\lambda(a) = 0$ for a.a. $a \geq \tilde{a}$ and assume that*

$$\int_0^{a_+} I(0,a)da > 0.$$

Then, if $R_0 > 1$,

$$\liminf_{t\to\infty} I(t) > \epsilon > 0$$

with ϵ not depending on the initial conditions.

Dynamical systems persistence theory does not give us information, however, as to whether B and W are bounded away from zero, too.

4. Stability of the endemic equilibrium: The epidemiology

The stability of the endemic equilibrium is of epidemiological interest for at least two reasons: In the case of locally asymptotic stability, in many instances it is the ultimate state of the epidemic, though only a global stability analysis would provide the definite answer to this question. In our model, this is a definite possibility, as there is only one endemic equilibrium and the disease-free equilibrium becomes a repeller as soon as the endemic equilibrium comes into existence. On the other hand, if an endemic equilibrium is unique and unstable, undamped oscillations of the disease dynamics around this equilibrium are very likely. Recall Theorem 3 which states that the disease dynamics either converge to the endemic equilibrium or oscillate around it.

Intuitively local asymptotic stability means that once the course of the disease comes close to the endemic equilibrium it remains close and finally approaches it. The precise definition can be given most nicely in reference to the model formulation (1), ..., (6).

Definition. a) *The endemic equilibrium* S^*, I^*, W^*, B^*, i^* *of (1), ..., (6) with*

$$i^*(\tau) = B^* P_{\alpha+1}(\tau)$$

is locally asymptotically stable if and only if the following two properties hold:
(i) For any $\epsilon > 0$ *there is some* $\delta > 0$ *such that*

$$|S(0) - S^*| + \int_0^\infty |i(0,a) - i^*(a)|da \leq \delta$$

implies that

$$|S(t) - S^*| + \int_0^\infty |i(t,a) - i^*(a)|da \leq \epsilon \quad \text{for all} \quad t \geq 0.$$

(ii) There exists $\delta_0 > 0$ *with the following property: If*

$$|S(0) - S^*| + \int_0^\infty |i(0,a) - i^*(a)|da \leq \delta_0,$$

then

$$|S(t) - S^*| + \int_0^\infty |i(t,a) - i^*(a)|da \quad \longrightarrow 0 \quad \text{for } t \to \infty.$$

b) The endemic equilibrium is called unstable if the following holds: There exists a sequence of solutions S_n, i_n *to (1),..., (6) and a sequence of times* $t_n \to \infty$ *and a positive number* $\epsilon_0 > 0$ *such that*

$$|S_n(0) - S^*| + \int_0^\infty |i_n(0,a) - i^*(a)|da \quad \longrightarrow 0 \quad \text{for} \quad n \to \infty$$

but

$$|S_n(t_n) - S^*| + \int_0^\infty |i_n(t_n, a) - i^*(a)|da \geq \epsilon_0 \quad \text{for all } n \in \mathbb{N}.$$

The discussion of the stability and instability of the endemic equilibrium is facilitated by switching from the original parameters of the model to the following non-dimensional ones:

$$\xi = \frac{I^*}{T^*} = \frac{I^*}{S^* + I^*}, \tag{34}$$

$$\gamma := -\frac{T^* M'(T^*)}{M(T^*)} \tag{35}$$

and

$$\sigma := \frac{1}{\hat{P}_{\alpha+1}(0)}. \tag{36}$$

Note that ξ gives the fraction of infected individuals in the sexually active population and therefore is a dimensionless parameter satisfying

$$0 < \xi < 1.$$

Even though all values of ξ in the interval $0 < \xi < 1$ are feasible (as one can see from (31) and (33) by choosing $R_0 > 1$ accordingly), not all of them may be realistic. Note that $\frac{1}{\sigma} = \hat{P}_{\alpha+1}(0)$ can be interpreted as the average length of the effective sexually active period of infected individuals (relative to the average length of the sexually active period of healthy individuals, our time unit). Hence it is intuitively clear — and follows from the definition of $P_{\alpha+1}$ in (11) — that $\sigma > 1$. The average duration of infection has been estimated to be about 10 years (see May and Anderson, 1989, and references therein). If we assume that the mean length of the sexually active period of healthy individuals lies between 15 and 30 years, we obtain values of σ in the interval [1.5, 3].

We observe that

$$\gamma = -\frac{T^* M'(T^*)}{M(T^*)}$$

is a dimensionless parameter also. As $M(T) = \frac{C(T)}{T}$ is non-increasing and C is non-decreasing we have that

$$0 \leq \gamma \leq 1.$$

The following choices for $C(T)$ may give us a feeling for a reasonable range for γ.

(a) *Mass action type contact law*

The classical epidemiological contact law is $C(T) = \beta T$. This implies that $M = \text{const}$, hence $\gamma = 0$.

This contact law is more appropriate for communicable diseases such as influenza etc. (see Castillo-Chavez et al., 1988, 1989 e), but not for sexually transmitted diseases.

(b) $C = \text{const}$

This may be a good approximation if the number of available partners is large enough and everybody could make more contacts than is practically feasible. In this case, $\gamma = 1$.

(c) *Michaelis-Menten type contact law*

The Michaelis-Menten type contact law (or Holling functional response type 2) combines the two previous approaches by assuming that, if the number of available partners T is low, the number of actual per capita partners $C(T)$ is proportional to T whereas, if the number of available partners is large, there is a saturation effect which makes the number of actual partners constant. Specifically it has the form

$$C(T) = \frac{\beta T}{1 + \kappa T}.$$

This law was first formulated by Michaelis and Menten (1913) for enzyme-catalyzed reactions. It was used by Monod (1942) to describe the nutrient uptake by bacteria. Later Holling (1966, p. 11) derived it to model the functional response of an invertebrate predator to the available amount to prey. A similar derivation can be made relating the number of sexual contacts to the number of available partners.

For the Michaelis-Menten contact law we obtain from definition (35) that

$$\gamma = \frac{\kappa T^*}{1 + \kappa T^*}.$$

Consequently γ covers the whole range from 0 to 1 when T^* covers the range from 0 to ∞. From (30) we have that

$$T^* = \frac{\Lambda}{1 + (\sigma - 1)\xi},$$

such that any value of T^* between 0 and ∞ — i.e. any value of γ between 0 and 1 — is possible (though not necessarily realistic) by choosing Λ accordingly.

In view of this discussion we call γ the ***saturation index*** of the number of partners at the endemic equilibrium. If $\gamma = 0$ there is no saturation at all because the number of actual partners is proportional to the number of available partners. If $\gamma = 1$ there is complete saturation because the number of actual partners hardly changes if the number of available partners does.

In sections 5 and 6 we prove the following result.

Theorem 5. *The endemic equilibrium is locally asymptotically stable if one of the following holds:*

a) ξ *is sufficiently close to 0 or to 1.*

b) σ *is sufficiently large.*

c) γ *is sufficiently close to 0.*

d) $\lambda = const.$

e) $P_{\alpha+1}$ *is convex.*

Thus the endemic equilibrium is locally asymptotically stable if the fraction of infected individuals is either low or high, or if the mean length of the sexually active period of infected individuals is short compared with the length of the sexually active period of susceptible individuals, or the saturation index is low. Further we have local stability if the infectivity is evenly distributed over the period of sexual activity. $P_{\alpha+1}$ may be convex, for example, if the length of the sexually active period of infected individuals is exponentially distributed (in particular if the inactivation rate α is constant); but this is presumably not the case for AIDS.

The fact that constant infectivity implies the stability of the endemic equilibrium has already been proved in Castillo-Chavez et al. (1989 b). In the model with variable infectivity we can specify how much the infectivity may deviate from its mean value without destroying the stability of the endemic equilibrium. To this end we first have to introduce an appropriate mean infectivity. We recall that $\lambda(\tau)$ is actually the potential infectivity at infection-age τ because it does come into effect if an individual has already retired from sexual activity. Remember that $P_{\alpha+1}(\tau)$ gives the proportion of individuals that are still sexually active at time τ after infection. We make it a probability density by setting

$$p(\tau) = \frac{P_{\alpha+1}(\tau)}{\hat{P}_{\alpha+1}(0)}.$$

Hence

$$\bar{\lambda} = \int_0^\infty \lambda(\tau)p(\tau)d\tau$$

can be interpreted as the *effective mean infectivity.* Note that $\lambda = \bar{\lambda}$ if λ is constant. As a measure for the deviation from the effective mean infectivity we introduce

$$\int_0^\infty \frac{|\lambda(\tau) - \bar{\lambda}|}{\bar{\lambda}} p(\tau)d\tau.$$

Notice that this expression does not change if λ is replaced by a constant multiple of itself.

Theorem 6. *The endemic equilibrium is locally asymptotically stable if*

$$\int_0^\infty \frac{|\lambda(\tau) - \bar{\lambda}|}{\bar{\lambda}} p(\tau)d\tau \leq 1.$$

Apparently we can expect the endemic equilibrium to loose local stability only if the infectivity distribution is sufficiently spiky. This can happen indeed.

Theorem 7. *Let* $0 < \gamma \leq 1$ *and*

$$\frac{1}{1+y^2}(1-\gamma)+\gamma\int_0^\infty \cos(sy)P_{\alpha+1}(s)ds < 0$$

for some $y > 0$. *Then there exists an infectivity distribution* λ *with arbitrarily many peaks such that the endemic equilibrium is unstable.*

Remark. *It is actually only the multiplicative combination* Q *of* λ *and* $P_{\alpha+1}$ *in (12) that is required to have a peaked distribution. The peaks are concentrated at points of the form* $\tau_0, \ldots, \tau_m$ *where* $0 < \tau_0 < \frac{\pi}{2y}$ *has to be chosen appropriately and* $\tau_j = \tau_0 + \frac{2k_j}{y}$ *with arbitrary numbers* $k_j \in \mathrm{N}$.

The condition in Theorem 6 concerning $P_{\alpha+1}$ is the easier to satisfy the closer γ is to 1. So, apparently, the saturation index has a destabilizing effect. The requirement that λ is concentrated at certain parts of the incubation period, i.e. that the infectivity occurs in peaks, emphasizes the importance of an infection-age-dependent infectivity. The Remark shows that the endemic equilibrium can be unstable for infectivity distributions with an early peak if the condition in Theorem 6 can be satisfied for large y. This is the case for the following example of an inactivation rate α which is discussed in detail in Section 7:

$$\alpha(\tau) = \begin{cases} \rho_1, & 0 \leq \tau \leq \tau_\diamond \\ \rho_2, & \tau > \tau_\diamond \end{cases}$$

with $0 \leq \rho_1 < \rho_2, \tau_\diamond > 0$.

ρ_1 can be interpreted as the rate at which infected individuals stop sexual activity because they have tested HIV-positive, whereas the larger rate ρ_2 also incorporates inactivation by AIDS symptoms. $\tau_\diamond$ is the infection-age at which symptoms start appearing.

Example 1. *Let the inactivation rate* α *be as just described. If* $\gamma, 0 < \gamma \leq 1$, *is sufficiently close to 1, then there exist an infectivity distribution* λ *with arbitrarily many peaks — including early peaks — such that the endemic equilibrium is unstable.*

The peaks of $\lambda(\tau)$ *are concentrated at points of the form* $\tau_0, \ldots, \tau_m$ *with*

$$0 < \tau_0 < \frac{\tau_\diamond}{2s+4n}, \quad 1 < s < 2,$$

$$\tau_j = \tau_0 + \frac{2k_j}{s+2n}\tau_\diamond, \quad k_j \in \mathrm{N},$$

where $n \in \mathrm{N}$ has to be chosen sufficiently large.

In this example the time moments at which the infectivity is concentrated are related in a certain way among each other and also to the time $\tau_\diamond$ after infection at which symptoms start to appear. It is doubtful whether this mathematical relation is compatible with the biological relation between the time at which AIDS symptoms occur for the first time and the time at which the second infectivity rise occurs. Moreover the empirical data suggest that the second infectivity rise has rather the form of a plateau than of a peak. Though the multiplicative combination Q in (12) of the infectivity distribution λ with the probability $P_{\alpha+1}$ of being still sexually active brings the plateau-like shape of λ finally down again, Q is presumably too spread out to be of the form described in Theorem 6 and the subsequent Remark.

In order to illustrate the possible effect of a physiologically sensible coupling between the infectivity distribution and the inactivation rate we consider the following example where we consider only a late infectivity plateau. We assume that the infectivity is 0 until the moment where symptoms occur and a positive constant thereafter:

$$\lambda(\tau) = \begin{cases} 0, & 0 \le \tau < \tau_\diamond \\ \lambda_\diamond, & \tau > \tau_\diamond \end{cases}$$

It is suggestive that the moment where the symptoms start to appear and the infectivity rises are related in time because both effects are caused by the breakdown of the immune system. The above choice is the easiest way to mimic this relation.

Example 2. *Let α and λ be given as just described. Then the endemic equilibrium is locally asymptotically stable.*

Though the function $Q = \lambda P_{\alpha+1}$ gets a sharp peek at $\tau = \tau_\diamond$ if the parameter ρ_2 in the definition of α becomes large, the endemic equilibrium stays locally asymptotically stable.

In a next step one would like to combine a late infectivity plateau as described in example 2 with an early peak as in example 1. If the early peak alone would destroy the local stability, the combined distribution will have an unstable endemic equilibrium as well, provided that the early peak dominates the late plateau. Conversely, if the late plateau dominates the early peak, the endemic equilibrium will be stable. Further (presumably numerical) investigations have to clarify how much the early peak has to dominate the late plateau for instability to be possible.

5. Stability of the endemic equilibrium and the characteristic equation

Using results from the theory of evolution equations (abstract differential equations), specifically, Corollary 4.3 and section 7 in Thieme (1990) or theorem 4.13 in Webb (1985), it is possible to approach the stability of the endemic equilibrium for (1), ..., (6) in the same way as for a finite system of ordinary differential equations. Hence we set

$$S = S^* + s,\ i = i^* + u,\ I = I^* + v,\ W = W^* + w$$

and consider the variational equations for s, u, v, w related to (1), ..., (6). In other words we linearize (1), ... ,(6) around the endemic equilibrium:

$$\frac{d}{dt}s(t) = -u(t,0) - s(t), \tag{40}$$

$$(\frac{\partial}{\partial t} + \frac{\partial}{\partial a})u(t,a) = -(\alpha(a)+1)u(t,a), \tag{41}$$

$$u(t,0) = s(t)M(T^*)W^* + (s+v)S^*M'(T^*)W^* + wS^*M(T^*), \tag{42}$$

where

$$v(t) = \int_0^\infty u(t,a)da \tag{43}$$

$$w(t) = \int_0^\infty \lambda(a)u(t,a)da. \tag{44}$$

To study the stability of this linear system we look for solutions to (40), ..., (44) of the exponential form

$$s(t) = e^{zt}\tilde{s}, u(t,a) = e^{zt}\tilde{u}(a) \tag{45}$$

with a complex number z and $\tilde{s} \neq 0$ or $\tilde{u} \not\equiv 0$.

The endemic equilibrium will be locally asymptotically stable provided that for all solutions of this form the real part of z is strictly negative. It will be unstable if there is at least one such solution with the real part of z being strictly positive.

Substituting (45) into (40), ..., (44) yields

$$z\tilde{s} = -\tilde{u}(0) - \tilde{s} \tag{46}$$

$$z\tilde{u}(a) + \frac{d}{da}\tilde{u}(a) = -(\alpha(a)+1)\tilde{u}(a) \tag{47}$$

$$\tilde{u}(0) = \tilde{s}W^*M(T^*) + (\tilde{s}+\tilde{v})S^*M'(T^*)W^* + \tilde{w}S^*M(T^*) \tag{48}$$

$$\tilde{v} = \int_0^\infty \tilde{u}(a)da \tag{49}$$

$$\tilde{w} = \int_0^\infty \lambda(a)\tilde{u}(a)da. \tag{50}$$

From (46) we obtain

$$\tilde{s} = -\frac{\tilde{u}(0)}{1+z}. \tag{51}$$

We solve (47) for $\tilde{u}$ and fit the result into (49) and (50):

$$\tilde{v} = \tilde{u}(0)\hat{P}_{\alpha+1}(z), \tag{52}$$

$$\tilde{w} = \tilde{u}(0)\hat{Q}(z). \tag{53}$$

Note that $\tilde{u}(0)$ has to be different from 0 because otherwise both $\tilde{s} = 0, \tilde{u} \equiv 0$. Fitting (51), (52), and (53) into (48) and dividing by $\tilde{u}(0) \neq 0$ yields the *characteristic equation*

$$1 = -\frac{W^*}{1+z}\Big(M(T^*) + S^*M'(T^*)\Big) + S^*M'(T^*)W^*\hat{P}_{\alpha+1}(z) + S^*M(T^*)\hat{Q}(z). \tag{54}$$

We note that the same characteristic equation is obtained by linearizing the limiting equations associated with (7), ..., (10) around the endemic equilibrium and looking for solutions of exponential form.

Finally, we arrive at the following relation between the stability of the endemic equilibrium and the roots of the characteristic equation (54).

Theorem 8. a) *The endemic equilibrium is locally asymptotically stable if all the roots z of the characteristic equation (54) have strictly negative real parts.*
b) *The endemic equilibrium is unstable if the characteristic equation (54) has at least one root z with strictly positive real part.*

6. Analysis of the characteristic equation

The difficulty in analyzing (54) along the lines proposed by theorem 8 consists in the fact that W^*, S^*, T^* are related by the steady state equations (22), ..., (25). In a next step we incorporate these relations into (54) and reformulate (54) with epidemiologically meaningful parameters which can be modified independently of each other.

From (25), (17), (23), (29) we obtain

$$W^* M(T^*) = \frac{B^*}{S^*} = \frac{I^*}{S^*} \frac{1}{\hat{P}_{\alpha+1}(0)} = \frac{\xi}{1-\xi} \frac{1}{\hat{P}_{\alpha+1}(0)}.$$

From (24), (28), (17), (23), (29) we get

$$S^* W^* = S^* B^* \hat{Q}(0) = \frac{B^*}{M(T^*)} = \frac{I^*}{M(T^*)\hat{P}_{\alpha+1}(0)} = \frac{T^*}{M(T^*)} \frac{\xi}{\hat{P}_{\alpha+1}(0)},$$

and from (28)

$$S^* M(T^*) = \frac{1}{\hat{Q}(0)}.$$

We substitute these relations into the characteristic equation (54):

$$1 = -\frac{1}{1+z} \frac{\xi}{\hat{P}_{\alpha+1}(0)} \left(\frac{1}{1-\xi} + \frac{T^* M'(T^*)}{M(T^*)} \right) + \xi \frac{T^* M'(T^*)}{M(T^*)} \frac{\hat{P}_{\alpha+1}(z)}{\hat{P}_{\alpha+1}(0)} + \frac{\hat{Q}(z)}{\hat{Q}(0)}. \tag{55}$$

In order to simplify the characteristic equation further we define the probability densities

$$p(s) = \frac{P_{\alpha+1}(s)}{\hat{P}_{\alpha+1}(0)}, \tag{56}$$

and

$$q(s) = \frac{Q(s)}{\hat{Q}(0)}, \tag{57}$$

and re-introduce the parameter

$$\gamma := -\frac{T^* M'(T^*)}{M(T^*)}. \tag{58}$$

As $P_{\alpha+1}(0) = 1$ by (11), we see from (56) that

$$\frac{1}{\hat{P}_{\alpha+1}(0)} = p(0). \tag{59}$$

With these definitions equation (55) takes the form

$$1 = -\frac{p(0)\xi}{1+z}\left(\frac{1}{1-\xi} - \gamma\right) - \xi\gamma\hat{p}(z) + \hat{q}(z). \tag{60}$$

From the definition of $P_{\alpha+1}$ — see (11) — it follows that $p(0) > 1$. As $\frac{1}{p(0)}$ can be interpreted as the mean length of the effective sexually active period of infected individuals, it is more realistic to require $p(0) > 2$ because the duration of the sexually active period of healthy individuals seems to be more than twice that of infected individuals. Note that p and q are probability densities in dimensionless time related to the average duration of the sexually active period of healthy individuals.

As we have explained in Section 4, one can interpret the dimensionless parameter

$$\gamma = -\frac{T^* M'(T^*)}{M(T^*)}$$

as the saturation index of partner acquisition at the endemic equilibrium. As $M(T) = \frac{C(T)}{T}$ is non-increasing and C is non-decreasing we have that $0 \leq \gamma \leq 1$ (see the discussion in Section 4 concerning the dependence of γ on the functional relation C between the numbers of actual and potential partners and on the equilibrium value T^*.) We recall that $\xi = \frac{I^*}{T^*}, 0 < \xi < 1$, is the fraction of infected individuals at the endemic equilibrium. We know from (31) and (33) that each $0 < \xi < 1$ is feasible (though not necessarily realistic) by choosing $R_0 > 1$ accordingly.

In order to study the position of the roots $z \in \mathbf{C}$ of (60) we let $z = x + iy$ and separate (60) into real and imaginary part:

$$1 - \int_0^\infty e^{-xs}\cos(sy)q(s)ds = -\frac{(1+x)p(0)\xi}{(1+x)^2+y^2}\left(\frac{1}{1-\xi} - \gamma\right) - \xi\gamma\int_0^\infty e^{-xs}\cos(sy)p(s)ds \tag{61}$$

and

$$\int_0^\infty e^{-xs}\sin(sy)q(s)ds = \frac{yp(0)\xi}{(1+x)^2+y^2}\left(\frac{1}{1-\xi} - \gamma\right) + \xi\gamma\int_0^\infty e^{-xs}\sin(sy)p(s)ds. \tag{62}$$

We can solve for ξ by multiplying (61) by y and (62) by $1+x$ and adding the two equations together:

$$\xi = \frac{y\left(1 - \int_0^\infty e^{-sx}\cos(sy)q(s)ds\right) + (1+x)\int_0^\infty e^{-sx}\sin(sy)q(s)ds}{\gamma\left((1+x)\int_0^\infty e^{-sx}\sin(sy)p(s)ds - y\int_0^\infty e^{-sx}\cos(sy)p(s)ds\right)} \tag{63}$$

As q is a probability density, we see that the left hand side of (61) is strictly positive. Further, by the Riemann&Lebesgue Lemma,

$$\int_0^\infty e^{-sx}\sin(sy)p(s)ds \to 0,\ \int_0^\infty e^{-sx}\sin(sy)q(s)ds \to 0,\ |y|+x\to\infty, x\geq 0. \qquad (64)$$

provided that λ is of bounded variation, and, of course, the same holds if sine is replaced by cosine. Note further that

$$\int_0^\infty e^{-sx}\sin(sy)p(s)ds > 0, \quad x\geq 0, \qquad (65)$$

because p is non-increasing. This implies that the roots of (61), (62) satisfy $x<0$ if $\xi>0$ is small enough. As $0<\xi<1, 0\leq\gamma\leq 1$ we have $\frac{1}{1-\xi}-\gamma>0$. Hence there exist no roots with $x\geq 0, y=0$.

Now suppose that there exists some $0<\xi<1$ such that (61) and (62) can be solved by $x,y>0$. Since the roots of the characteristic equation depend continuously on ξ by Rouché's theorem and lie in the left half plane for small $\xi>0$, they must cross the imaginary axis as ξ increases. Note that $y\to\infty$ is excluded by (64). Hence, for some $0<\xi<1$ (different from the one we started with), (61) and (62) are solved with $x=0, y>0$, i.e.

$$1-\int_0^\infty \cos(sy)q(s)ds = -\frac{1}{1+y^2}p(0)\xi\left(\frac{1}{1-\xi}-\gamma\right)-\xi\gamma\int_0^\infty\cos(sy)p(s)ds \qquad (66)$$

and

$$\int_0^\infty \sin(sy)q(s)ds = \frac{y}{1+y^2}p(0)\xi\left(\frac{1}{1-\xi}-\gamma\right)+\xi\gamma\int_0^\infty\sin(sy)p(s)ds. \qquad (67)$$

In this case — see (63) —

$$\xi = \frac{y\left(1-\int_0^\infty\cos(sy)q(s)ds\right)+\int_0^\infty\sin(sy)q(s)ds}{\gamma\left(\int_0^\infty\sin(sy)p(s)ds - y\int_0^\infty\cos(sy)p(s)ds\right)}. \qquad (68)$$

Taking into account that $0<\xi<1, 0\leq\gamma\leq 1$ and that $p(0)>1$ we have the following:

Proposition 1. *There are no roots of (61), (62) with $x\geq 0$ if one of the following holds:*

a) ξ is sufficiently close to 0 or to 1.

b) $p(0)$ is sufficiently large.

c) γ is sufficiently close to 0.

d) There is no $y > 0$ satisfying the following simultaneously:

$$\int_0^\infty \cos(sy)q(s)ds > 0,$$

$$\int_0^\infty \sin(sy)q(s)ds > 0,$$

$$\int_0^\infty \cos(sy)p(s)ds < 0,$$

$$\begin{aligned} 0 < \quad & y\left(1 - \int_0^\infty \cos(sy)q(s)ds\right) + \int_0^\infty \sin(sy)q(s)ds \\ < & \gamma\left(\int_0^\infty \sin(sy)p(s)ds - y\int_0^\infty \cos(sy)p(s)ds\right). \end{aligned}$$

e) $\lambda = const.$

f) p convex.

Proof: a) If (a) does not hold, we have sequences ξ_j, x_j, y_j satisfying (61) and (62) with $0 < \xi_j < 1$, $x_j, y_j \geq 0$ and $\xi_j \to 0$ or $\xi_j \to 1$ for $j \to \infty$. As we have argued above, we actually have $y_j > 0$. Recall that $\frac{1}{1-\xi} - \gamma > 0$ are strictly positive.

We first consider the case $\xi_j \to 0, j \to \infty$. As the left hand side of (61) is strictly positive for $y > 0$ and bounded away from 0 for $y \to \infty$ by (64), this can only occur if $x_j, y_j \to 0$. We now divide (62) (with $y = y_j$) by $y_j > 0$. Taking the limit for $j \to \infty$, we obtain $\int_0^\infty sq(s)ds > 0$ on the left hand side, whereas, on the right hand side, we obtain 0. This contradicts (62).

In case that $\xi_j \to 1, j \to \infty$, the left hand side remains non-negative whereas the right hand side of (61) converges towards $-\infty$ (at least for a subsequence), unless $y_j \to \infty$. In case that $y_j \to \infty, j \to \infty$, we have from (64) that the left hand side of (61) converges to 1 for $j \to \infty$, whereas the right hand side is non-positive in the limit. Hence (61) cannot hold for large j, a contradiction.

The statements b) , c) are follows from (61), (62), and (64) in a similar way as a).

d) As we have argued above, the existence of a root with $x \geq 0$ for some $\xi, 0 < \xi < 1$, implies the existence of a root with $x = 0$ for some (presumably different) $\xi, 0 < \xi < 1$. For such a root it follows from (67) and (65) that $\int_0^\infty \sin(sy)q(s)ds > 0$. As the left hand side of (66) is strictly positive, it follows that $\int_0^\infty \cos(sy)p(s)ds < 0$. If $\int_0^\infty \cos(sy)q(s)ds \leq 0$, the left hand side of (66) is larger than 1 whereas the right hand side is strictly smaller than 1. Recall that p is a probability density. The last inequality in d) follows from (68) and $0 < \xi < 1$.

The statements e) and f) are consequences of d). If $\lambda = \text{const}$, then $q(s) = \lambda p(s)$, so the relative integrals cannot have different signs. If p is convex, then $\int_0^\infty \cos(sy)p(s)ds = \int_0^\infty \frac{1}{y}\sin(sy)(-p'(s))ds > 0$.

Proposition 1 e) — λ =const — is the case considered in Castillo-Chavez et al. (1989 b). We want to specify how much the infection-age dependent infectivity λ may deviate from its mean value without destroying the stability of the endemic equilibrium. To this end we introduce the effective mean infectivity

$$\bar{\lambda} = \int_0^\infty \lambda(\tau)p(\tau)d\tau.$$

As p is a probability density, $\lambda(\tau) = \bar{\lambda}$ in case that λ is constant. The following result implies Theorem 6 via Theorem 8.

Proposition 2. *All roots z of the characteristic equation (60) have strictly negative real part provided that*

$$\int_0^\infty \frac{|\lambda(\tau) - \bar{\lambda}|}{\bar{\lambda}} p(\tau)d\tau \leq 1.$$

Proof: Let us suppose that there is a root z with non-negative real part. As we have argued before, we then have a purely imaginary root for some $\xi, 0 < \xi < 1$. In particular Equation (66) holds. By (57), (12), (56), and the above definition of $\bar{\lambda}$ we have

$$q(\tau) = \frac{Q(\tau)}{\hat{Q}(0)} = \frac{\lambda(\tau)P_{\alpha+1}(\tau)}{\int_0^\infty \lambda(s)P_{\alpha+1}(s)ds} = \frac{\lambda(\tau)p(\tau)}{\bar{\lambda}}.$$

From (66) we now obtain

$$0 > 1 - \int_0^\infty \cos(sy)\frac{\lambda(s)}{\bar{\lambda}}p(s)ds + \xi\gamma\int_0^\infty \cos(sy)p(s)ds.$$

By Proposition 1 d) we can assume that

$$\int_0^\infty \cos(ys)p(s)ds < 0.$$

As $\xi\gamma < 1$ we can continue the above inequality by

$$0 > 1 - \int_0^\infty \cos(sy)\frac{\lambda(s) - \bar{\lambda}}{\bar{\lambda}}p(s)ds \geq 1 - \int_0^\infty \frac{|\lambda(s) - \bar{\lambda}|}{\bar{\lambda}}p(s)ds.$$

This inequality contradicts our assumption.

There are so many constraints for a root of the characteristic equation to have non-negative or even positive real part that one might conjecture that there are none with this property. It is difficult, indeed, to show for given probability densities p, q that the characteristic equation has roots with positive real part. This task is facilitated by considering a family of probability densities q_c rather than a specific density q. We will give conditions for a fixed probability density p and a family of probability densities q_c which guarantee that the characteristic equation has a root with positive real part for at least one member of the family. Actually, in order to apply this condition, it is convenient to formulate it for probability measures.

Assumptions. *(a) Let p be a non-increasing probability density. Assume there is some $y > 0$ such that*

$$\frac{1}{1+y^2} p(0)\,(1-\gamma) + \gamma \int_0^\infty \cos(sy) p(s) ds < 0.$$

Further assume that $q_c, c_1 \le c \le c_2$, is a family of probability measures with the following properties:

(b) $\hat{q}_c(z)$ is continuous in $c, c_1 \le c \le c_2$, for every $z \in \mathbf{C}$ with non-negative real part.

(c) $\hat{q}_c(z)$ is continuous in $z \in \mathbf{C}, \Re z \ge 0$, uniformly in $c, c_1 \le c \le c_2$.

(d)
$$\int_0^\infty \sin(sy) q_c(s) ds$$

is strictly positive for $c > c_1$, c close to c_1, and is 0 for $c = c_1$.

(e)
$$\frac{1 - \int_0^\infty \cos(sy) q_c(s) ds}{\int_0^\infty \sin(sy) q_c(s) ds} \to 0 \qquad c \to c_1.$$

(f) If ξ_c is defined by (68) with $q = q_c$, then $0 < \xi_c < 1$ for $c_1 < c < c_2$ and $\xi_c = 1$ for $c = c_2$.

Proposition 3. *Let the Assumptions be satisfied. Then the characteristic equation (60) has roots $z = x \pm \sqrt{-1} y$ with $x > 0$ for at least one $\xi \in (0,1)$ and one probability measure $q = q_c, c_1 < c < c_2$.*

Proof: The strategy of the proof is the following: We consider Equation (61) with ξ being given by Equation (63) where q is replaced by q_c. We show that for some $\tilde{c}_1$ close to c_1 the left hand side of (61) is smaller than the right hand side, whereas for some $\tilde{c}_2$ close to c_2 the left hand side of (61) is larger than the right hand side, and that further, for all values of c between $\tilde{c}_1$ and $\tilde{c}_2$, the number ξ given by (63) is strictly between 0 and 1. Then we apply the intermediate value theorem.

In a first step we rather consider (66) and (68) than (61) and (63). It follows from the Assumptions (d), (e) that $\xi_c \to 0, c \to c_1$, and that the following limit exists:

$$\lim_{c \to c_1} \frac{\int_0^\infty \sin(sy) q_c(s) ds}{\xi_c} \in (0, \infty). \tag{69}$$

We divide (66) by ξ_c:

$$\frac{1 - \int_0^\infty \cos(sy) q_c(s) ds}{\xi_c} = -\frac{1}{1+y^2} p(0) \left(\frac{1}{1-\xi_c} - \gamma \right) - \gamma \int_0^\infty \cos(sy) p(s) ds \tag{70}$$

By (e) and (69), the left hand side of (70) tends to 0 for $c \to c_1$ while the right hand side converges towards

$$-\frac{1}{1+y^2} p(0) (1-\gamma) - \gamma \int_0^\infty \cos(sy) p(s) ds.$$

This expression is positive because of (a) in the Assumptions. So we find some $\tilde{c}_1$ close to c_1 such that the right hand side of (70), or equivalently of (66), is strictly larger than the left hand side. If $c \to c_2$, $\xi_c \to 1$, hence the right hand side of (66) goes to $-\infty$, whereas, by (b), the left hand side of (66) tends to some finite limit. Hence we find some $\tilde{c}_2$ between $\tilde{c}_1$ and c_2 such that the right hand side of (66) is strictly smaller than the left hand side. By (c) of the Assumptions this still holds if (66) is replaced by (61) and $x > 0$ is sufficiently close to 0. For the same reason we have $\xi_c(x) \in (0,1), \tilde{c}_1 < c < \tilde{c}_2$ if $\xi_c(x)$ is given by (63) instead of (68). (b) and the intermediate value theorem imply that, for any $x > 0$ which is sufficiently close to 0, there is some c between $\tilde{c}_1$ and $\tilde{c}_2$ such that (61) is satisfied with $\xi = \xi_c(x)$ being given by (63), $0 < \xi < 1$. Equivalently (60) with $\xi = \xi_c(x)$ has a root $z = x + \sqrt{-1}y$.

We can use this Proposition to show that the characteristic equation can have roots with positive real part indeed. The following result implies Theorem 7 (recall (59)).

Proposition 3. *Let p be a decreasing probability density. Assume there is some $y > 0$ such that*

$$\frac{1}{1+y^2} p(0) (1-\gamma) + \gamma \int_0^\infty \cos(sy) p(s) ds < 0.$$

Then there exists a probability density q with arbitrarily many peaks such that the characteristic equation (60) has a root z with strictly positive real part.

Remark. *The peaks of q are concentrated at points of the form $c, cs_1, \ldots, cs_m$, $0 < c < \frac{\pi}{2y}$, such that*

$$s_j = 1 + \frac{2k_j \pi}{cy}$$

with $k_j \in \mathrm{N}, k_j > 0, j = 1, \ldots, m$.

Proof: Let $q_c, c \geq 0$, be a probability measure which is concentrated at the points cs_j as indicated in the Remark, $s_0 = 1$. In other words

$$q_c = \sum_{j=0}^{m} \kappa_j \delta_{cs_j}$$

with $\kappa_j > 0, \sum_{j=0}^{m} \kappa_j = 1$ and δ_s denoting the Dirac measure concentrated at the point s. The choice of the points cs_j implies that

$$\int_{[0,\infty)} \cos(sy) q_c(ds) = \cos(cy), \quad \int_{[0,\infty)} \sin(sy) q_c(ds) = \sin(cy).$$

By (68),

$$\xi_c = \frac{y\Big(1 - \cos(cy)\Big) + \sin(cy)}{\gamma\Big(\int_0^\infty \sin(sy)p(s)ds - y\int_0^\infty \cos(sy)p(s)ds\Big).}$$

For $c = \frac{\pi}{2y}$ we have

$$\xi_c = \frac{y+1}{\gamma\Big(\int_0^\infty \sin(sy)p(s)ds - y\int_0^\infty \cos(sy)p(s)ds\Big)} \geq \frac{1}{\gamma} \geq 1,$$

as $0 < \int_0^\infty \sin(sy)p(s)ds, -\int_0^\infty \cos(sy)p(s)ds \leq 1$. So we find some $c_2, 0 < c_2 \leq \frac{\pi}{2y}$, such that, with $c_1 = 0$, part (f) of the Assumptions is satisfied. The other parts are now checked easily.

7. Examples

We give an example for inactivation rates $\alpha(\tau)$ such that the assumption of Proposition 3 in Section 6 is satisfied for $P_{\alpha+1}$:

$$\alpha(\tau) = \begin{cases} \rho_1, & 0 \le \tau \le \tau_\diamond \\ \rho_2, & \tau > \tau_\diamond \end{cases} \tag{71}$$

with $0 \le \rho_1 < \rho_2, \tau_\diamond > 0$.

ρ_1 can be interpreted as the rate at which infected individuals stop sexual activity because they have tested HIV-positive, whereas the larger rate ρ_2 also incorporates inactivation by AIDS symptoms. $\tau_\diamond$ is the infection-age at which symptoms start appearing.

By (11),

$$P_{\alpha+1}(\tau) = \begin{cases} e^{-(1+\rho_1)\tau}, & 0 \le \tau \le \tau_\diamond \\ e^{-(1+\rho_1)\tau_\diamond} e^{-(1+\rho_2)(\tau-\tau_\diamond)}, & \tau \ge \tau_\diamond \end{cases} \tag{72}$$

Hence the Laplace transform is given by

$$\hat{P}_{\alpha+1}(z) = \frac{1}{1+\rho_1+z}\left(1 - e^{-(1+\rho_1+z)\tau_\diamond}\right) + \frac{1}{1+\rho_2+z} e^{-(1+\rho_1+z)\tau_\diamond}. \tag{73}$$

In order to check the assumption of Proposition 3 we set $z = \sqrt{-1}y$ and take the real part:

$$\begin{aligned} &\int_0^\infty \cos(sy) P_{\alpha+1}(s) ds \\ &= \frac{1+\rho_1}{(1+\rho_1)^2+y^2}\left(1 - e^{-(1+\rho_1)\tau_\diamond}\cos(y\tau_\diamond)\right) + \frac{1+\rho_2}{(1+\rho_2)^2+y^2} e^{-(1+\rho_1)\tau_\diamond}\cos(y\tau_\diamond) \\ &\quad + y\left(\frac{1}{(1+\rho_1)^2+y^2} - \frac{1}{(1+\rho_2)^2+y^2}\right) e^{-(1+\rho_1)\tau_\diamond}\sin(y\tau_\diamond). \end{aligned} \tag{74}$$

As $\rho_2 > \rho_1$, we find that

$$\int_0^\infty \cos(sy) P_{\alpha+1}(s) ds < 0$$

if, e.g.,

$$y\tau_\diamond = (s+2n)\pi, \quad 1 < s < 2,$$

with s being kept fixed and with n being chosen sufficiently large. By (59) the condition in Proposition 3 takes the form

$$\frac{1}{1+y^2}(1-\gamma) + \gamma \int_0^\infty \cos(sy) P_{\alpha+1}(s) ds < 0.$$

and thus is satisfied if

$$y\tau_\diamond = (s+2n)\pi, \quad 1 < s < 2,$$

with n being sufficiently large and $\gamma, 0 < \gamma \leq 1$, being close enough to 1.

Corollary 1. *Let the inactivation rate $\alpha(\tau)$ be given by (71). If $\gamma, 0 < \gamma \leq 1$, is sufficiently close to 1, then there exist a probability density q with arbitrarily many peaks such that the characteristic equation (60) has roots z with strictly positive real part.*

From the Remark following Proposition 3 we can obtain information about the position of the peaks:

Remark. *The peaks of q are concentrated at points of the form $c_0, \ldots, c_m$ with*

$$0 < c_0 < \frac{\tau_\diamond}{2s + 4n}, \quad 1 < s < 2,$$

$$c_j = c_0 + \frac{2k_j}{s + 2n}\tau_\diamond, \quad k_j \in \mathrm{N},$$

where $n \in \mathrm{N}$ has to be chosen sufficiently large.

The peaks of q satisfy some kind of resonance relation among each other and also to $\tau_\diamond$. It is questionable whether such a relation is satisfied in reality. In order to illustrate that such a resonance condition is necessary we now consider a case of variable infectivity without a first peak, but with a late plateau. One can argue that the infectivity should rise at about the same time where symptoms do occur. For simplicity we let the infectivity be zero up to time $\tau_\diamond$ and be positive and constant thereafter. This is an extreme idealization, but not totally artificial:

$$\lambda(\tau) = \begin{cases} 0, & 0 \leq \tau \leq \tau_\diamond \\ \lambda_\diamond, & \tau > \tau_\diamond \end{cases} \tag{75}$$

By (12) and (57),

$$q(\tau) = \begin{cases} 0, & 0 \leq \tau \leq \tau_\diamond \\ (1 + \rho_2)e^{-(1+\rho_2)(\tau - \tau_\diamond)}, & \tau > \tau_\diamond \end{cases} \tag{76}$$

If ρ_2 is large, q has a rather sharp peak at $\tau_\diamond$. Nevertheless the characteristic equation (60) has roots with strictly negative real parts only.

Corollary 2. *Let α, λ be given by (71), (75). Then all roots z of the characteristic equation (60) have strictly negative real part.*

Proof: By (76)

$$\hat{q}(z) = \frac{1 + \rho_2}{1 + \rho_2 + z} e^{-z\tau_\diamond}.$$

We set $z = \sqrt{-1}y$ and separate into real and imaginary part:

$$\int_0^\infty \cos(sy)q(s)ds = \frac{1+\rho_2}{(1+\rho_2)^2+y^2}\left((1+\rho_2)\cos(y\tau_\diamond) - y\sin(y\tau_\diamond)\right), \tag{77}$$

$$\int_0^\infty \sin(sy)q(s)ds = \frac{1+\rho_2}{(1+\rho_2)^2+y^2}\left(y\cos(y\tau_\diamond) + (1+\rho_2)\sin(y\tau_\diamond)\right). \tag{78}$$

Let us suppose that the characteristic equation (60) has roots with non-negative real part. From Proposition 1 (d) we can conclude:

$$(1+\rho_2)\cos(y\tau_\diamond) - y\sin(y\tau_\diamond) > 0, \tag{79}$$

$$y\cos(y\tau_\diamond) + (1+\rho_2)\sin(y\tau_\diamond) > 0. \tag{80}$$

This implies that

$$\cos(y\tau_\diamond) > 0.$$

From (80),

$$\sin(y\tau_\diamond) > -\frac{y}{1+\rho_2}\cos(y\tau_\diamond).$$

We substitute this inequality into (74):

$$\begin{aligned}
&\int_0^\infty \cos(sy)P_{\alpha+1}(s)ds \\
&= \frac{1+\rho_1}{(1+\rho_1)^2+y^2}\left(1 - e^{-(1+\rho_1)\tau_\diamond}\cos(y\tau_\diamond)\right) + \frac{1+\rho_2}{(1+\rho_2)^2+y^2}e^{-(1+\rho_1)\tau_\diamond}\cos(y\tau_\diamond) \\
&\quad - y\left(\frac{1}{(1+\rho_1)^2+y^2} - \frac{1}{(1+\rho_2)^2+y^2}\right)e^{-(1+\rho_1)\tau_\diamond}\frac{y}{1+\rho_2}\cos(y\tau_\diamond) \\
&= \frac{1+\rho_1}{(1+\rho_1)^2+y^2}\left(1 - e^{-(1+\rho_1)\tau_\diamond}\cos(y\tau_\diamond)\right) \\
&\quad + \frac{1}{1+\rho_2}\left(1 - \frac{y^2}{(1+\rho_1)^2+y^2}\right)e^{-(1+\rho_1)\tau_\diamond}\cos(y\tau_\diamond).
\end{aligned} \tag{81}$$

As $\cos(y\tau_\diamond) > 0$, this implies

$$\int_0^\infty \cos(sy)P_{\alpha+1}(s)ds \geq 0,$$

in contradiction to Proposition 1 d) and (56).

8. Conclusions

Several mathematical studies of epidemic models have identified both mechanisms capable and incapable of generating sustained oscillations (see Hethcote et al., 1981, and Hethcote and Levin, 1989, for surveys), and as discussed in the introduction of this paper, most of these mechanisms are inadequate in the case of HIV.

In our model, the saturation of mean per capita sexual activity interacts with an infection-age-dependent removal rate (from sexual activity by the disease) and an infection-age-dependent infectivity. We have shown in this paper that the unique endemic equilibrium can lose its stability (thus presumably generating sustained oscillations) by a rather unique combination of conditions (see the Introduction and Sections 4, 6). The endemic equilibrium is locally asymptotically stable if any of the reasonable conditions (i)–(iv) in the Introduction are not satisfied. Condition (v) — the infection-age-distributed infectivity is highly concentrated at certain parts of the incubation period — emphasizes the possible relevance of variable infectivity on the dynamics of an HIV epidemic. Whereas the endemic equilibrium is locally asymptotically stable as long as the infectivity is sufficiently evenly distributed over the activity period (Theorem 6), we have shown the possibility of sustained oscillations if the infectivity distribution has sufficiently sharp peaks which are suitably situated. Example 1, Section 4, suggests that the first peak can be early. Example 2 shows, however, that — if we restrict to infectivity distributions with one late peak or one late plateau — the endemic equilibrium may be stable if the peak (or plateau) is related to the inactivation rate in an epidemiologically sensible way. This leaves the stability question open for distributions with one early peak and one late plateau when the late plateau is realistically linked with the inactivation rate. Undamped oscillations may occur if the early peak sufficiently dominates the late plateau whereas the endemic equilibrium is stable if the plateau is the dominating part.

Though these results show that the stability of the endemic state can definitely not be taken for granted, they allow the cautious conjecture that the endemic equilibrium is locally asymptotically stable for realistic inactivation rates and infectivity distributions. This conjecture is supported by the numerical simulations of models that incorporate variable infectivity by Anderson and May (1989), Hyman and Stanley (1988, 1989), Blythe and Anderson (1988b), Anderson et al. (preprint). Our model ignores that not only the infectivity, but also the timing of the first infectivity peak and of the late infectivity plateau may be highly variable. This variability has the effect that, on the average, the early peak and the late plateau are spread out; this presumably increases the odds that the endemic equilibrium is locally stable.

Whereas the results in this paper completely rely on analytical techniques, we plan to clarify numerically how much the early peak has to dominate the late plateau for undamped oscillations to occur. Our stability criterion can be used to determine numerically in

which parameter range the endemic equilibrium is unstable. Though simulations of the full model will be indispensable for showing whether the amplitudes of the oscillations are large enough to be epidemiologically significant, we feel that they need to be guided by the previous exploration of the critical parameter range.

The uncertainty of whether or not the endemic equilibrium is stable raises the question of whether or not it should be discarded as some kind of measure of the severity of the disease when unstable. However, its usefulness has been established by our proof that the incidence rate either converges to its uniquely determined endemic equilibrium (provided that it exists) or it fluctuates around it. Of course, it would be more useful to know whether or not the time averages converge (as time tends to infinity) towards the endemic equilibrium.

We remark that variable infectivity is just one of the important factors involved in HIV dynamics. Heterogeneity in sexual behavior is also of crucial importance in the transmission dynamics of HIV (see Blythe and Castillo-Chavez, 1989; Castillo-Chavez and Blythe, 1989; Busenberg and Castillo-Chavez, 1989, in press; Castillo-Chavez et al., in press; Castillo-Chavez and Busenberg, 1990; Sattenspiel and Castillo-Chavez, 1990; and references therein). The analytical study of a model that incorporates variable infectivity and heterogeneous mixing looks like a formidable task. However, mathematical studies of submodels of this general model are central to the execution of extensive numerical simulations of more detailed models.

We conclude with a not very optimistic view of the predictive value of mathematical models for HIV transmission. The recent literature in HIV modeling reveals a *potentially* very complex picture: multiple endemic equilibria and possibility of oscillations. This dynamic behavior is not observed in less detailed versions of these models (see Castillo-Chavez, 1989a, b) indicating that very aggregated versions of these models may not be adequate. Unfortunately, more detailed models require more data (most of which are not unavailable but not of sufficient quality). As Ludwig (1985, 1989) has shown, these demands put very severe limits on our ability to generate accurate predictions. The theoretical value of these models is nevertheless very important.

Acknowledgements

This research has been partially supported by NSF grant DMS-89006580 and Hatch project grant NYC 151-409, USDA awarded to Carlos Castillo-Chavez. We are grateful to K. L. Cooke and S. Busenberg for their stimulating conversations.

References

Anderson, R.M., S.P. Blythe, S. Gupta, E. Konings (1989): The transmission dynamics of the human immunodeficiency virus type 1 in the male homosexual community in the united kingdom: the influence of changes in sexual behaviour. Phil. Trans. R. Soc. London. B 325, 45-89.

Anderson, R.M., H.C. Jackson, R.M. May, and A.D.M. Smith. (1981). Population dynamics of fox rabies in Europe. Nature 289, 765–771.

Anderson, R.M. and R.M. May. (1987). Transmission dynamics of HIV infection. Nature 326, 137–142.

Anderson, R.M., R.M. May, and G.F. Medley. (1986). A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS. IMA J. Math. Med. Biol. 3, 229–263.

Andreasen, V. (thesis). Dynamical models of epidemics in age-structured populations: Analysis and simplifications. Ph.D. Thesis, Cornell University, 1988.

Andreasen, V. (1989). Multiple time scales in the dynamics of infectious diseases. In Mathematical Approaches to Problems in Resource Management and Epidemiology, C. Castillo-Chavez, S.A. Levin, and C. Shoemaker (eds.). Lecture Notes in Biomathematics 81. Springer-Verlag, Berlin, Heidelberg, New York, Tokyo.

Blythe, S.P. and R.M. Anderson. (1988a). Distributed incubation and infectious periods in models of the transmission dynamics of the human immunodeficiency virus (HIV). IMA J. Math. Med. Bio. 5, 1–19.

Blythe, S.P. and R.M. Anderson. (1988b). Variable infectiousness in HIV transmission models. IMA J. of Mathematics Applied in Med. and Biol. 5, 181–200.

Blythe, S.P. and C. Castillo-Chavez. (1989). Like-with-like preference and sexual mixing models. Math. Biosci. 96, 221–238.

Brauer, F. (1989). Epidemic models in populations of varying size. Mathematical Approaches to Problems in Resource Management and Epidemiology, C. Castillo-Chavez, S.A. Levin, and C. Shoemaker (eds.). Lecture Notes in Biomathematics 81, 109–123. Springer-Verlag

Brauer, F. (to appear a). Models for the spread of universally fatal diseases. J. Math. Biol.

Brauer, F. (to appear b). Models for the spread of universally fatal diseases, II. Differential Equations and Applications to Biology and Population Dynamics. Proceedings of the International Conference in Claremont, Jan. 1990. Lecture Notes in Biomathematics. Springer

Busenberg, S. and C. Castillo-Chavez. (1989). Interaction, pair formation and force of infection terms in sexually transmitted diseases. In (C. Castillo-Chavez, ed.) Mathematical

and Statistical Approaches to AIDS Epidemiology. Lecture Notes in Biomathematics 83, Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong: 289–300.

Busenberg, S. and C. Castillo-Chavez. (in press). On the role of preference in the solution of the mixing problem, and its application to risk- and age-structured epidemic models. IMA J. of Math. Appl. to Med. and Biol.

Busenberg, S., K.L. Cooke, and H.R. Thieme. (to appear). Interaction of population growth and disease dynamics for HIV/AIDS in a heterogeneous population. SIAM J. Appl. Anal.

Castillo-Chavez, C. (1989a). Review of recent models of HIV/AIDS transmission. In (S.A. Levin, T.G. Hallam, and L.J. Gross, eds.) Applied Mathematical Ecology, Biomathematics 18, Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, 253–262.

Castillo-Chavez, C. (ed.) (1989b). Mathematical and Statistical Approaches to AIDS Epidemiology. Lecture Notes in Biomathematics 83, Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong.

Castillo-Chavez, C. and S.P. Blythe. (1989). Mixing framework for social/sexual behavior. In (Castillo-Chavez, ed.) Mathematical and Statistical Approaches to AIDS Epidemiology. Lecture Notes in Biomathematics 83, Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong: 275–288.

Castillo-Chavez, C., K.L. Cooke, W. Huang, and S.A. Levin. (1989a). On the role of long periods of infectiousness in the dynamics of acquired immunodeficiency syndrome (AIDS). Mathematical Approaches to Problems in Resource Management and Epidemiology, C. Castillo-Chavez, S.A. Levin, and C. Shoemaker (eds.). Lecture Notes in Biomathematics 81, Springer-Verlag, 177–189.

Castillo-Chavez, C., K.L. Cooke, W. Huang, and S.A. Levin. (1989b). On the role of long incubation periods in the dynamics of acquired immunodeficiency syndrome (AIDS), Part 1. Single population models. J. Math. Biol. 27, 373–398.

Castillo-Chavez, K.L. Cooke, W. Huang, and S.A. Levin. (1989c). Results on the dynamics for models for the sexual transmission of the human immunodeficiency virus. Applied Mathematics Letters 2, 327-331

Castillo-Chavez, C., K.L. Cooke, W. Huang, and S.A. Levin. (1989d). On the role of long incubation periods in the dynamics of acquired immunodeficiency syndrome (AIDS), Part 2. Multiple group models. In Mathematical and Statistical Approaches to AIDS Epidemiology, C. Castillo-Chavez (ed.). Lecture Notes in Biomathematics 83, Springer-Verlag.

Castillo-Chavez, C., H.W. Hethcote, V. Andreasen, S.A. Levin, and W.M. Liu. (1989e). Epidemiological models with age structure, proportionate mixing, and cross-immunity. J.

Math. Biol. 27, 233–258.

Castillo-Chavez, C., H.W. Hethcote, V. Andreasen, S.A. Levin, and W.M. Liu. (1988). Cross-immunity in the dynamics of homogeneous and heterogeneous populations. In Mathematical Ecology, L. Gross, T.G. Hallam, and S.A. Levin (eds.). Proceedings of the Autumn Course Research Seminars, Trieste 1986 and World Scientific Publ. Co., Singapore.

Castillo-Chavez, C., S. Busenberg, K. Gerow (in press). Pair formation in structured populations. In: Differential Equations with Applications in Biology, Physics, and Engineering (J. Goldstein, F. Kappel, W. Schappacher, eds.). Marcel Dekker, New York

Diekmann, O. and R. Montijn. (1982). Prelude to Hopf bifurcation in an epidemic model: analysis of a characteristic equation associated with a nonlinear Volterra integral equation. J. Math. Biol. 14, 117–127.

Diekmann, O. and S.A. van Gils. (1984). Invariant manifolds for Volterra integral equations of convolution type. J. Diff. Equa. 54, 189–190.

Feller, W. (1941). On the integral equation of renewal theory. Ann. Math. Stat. 12, 243–267

Francis, D.F., P.M. Feorino, J.R. Broderson, H.M. McClure, J.P. Getchell, C.R. McGrath, B. Swenson, J.S. McDougal, E.L. Palmer, A.K. Harrison, F. Barré-Sinoussi, J.C. Chermann, L. Montagnier, J.W. Curran, C.D. Cabradilla, and V.S. Kalyanaraman. (1984). Infection of chimpanzees with lymphadenopathy-associated virus. Lancet 2, 1276–1277.

Gripenberg, G. (1980). Periodic solutions to an epidemic model. J. Math. Biol. 10, 271–280.

Gripenberg, G. (1981). On some epidemic model. Appl. Math. 39, 317–327.

Gripenberg, G., S.O. Londen, O. Staffans. (1990). Volterra Integral and Functional Equations. Cambridge Univ. Press

Hale, J.K. and P. Waltman. (1989). Persistence in infinite-dimensional systems. SIAM J. Math. Anal. 20, 388–395.

Haseltine, W.A. (1989): Silent HIV infections. New Engl. J. of Med. 320, 1487–1489

Hethcote, H.W. and S.A. Levin. (1989). Periodicity in epidemiological models. In Applied Mathematical Ecology, S.A. Levin, T.G. Hallam, and L.J. Gross (eds.). Biomathematics 18, Springer-Verlag, Heidelberg.

Hethcote, H.W., H.W. Stech, and P. van den Driessche. (1981). Periodicity and stability in epidemic models: a survey. In Differential Equations and Applications in Ecology, Epidemics and Population Problems, S. Busenberg and K.L. Cooke (eds.). Academic Press, New York.

Hethcote, H.W. and H.R. Thieme. (1985). Stability of the endemic equilibrium in epidemic models with subpopulations. Math. Biosci. 75, 205–227.

Hethcote, H.W. and J.A. Yorke. (1984). Gonorrhea, Transmission Dynamics, and Control. Lecture Notes in Biomathematics 56. Springer-Verlag, Berlin, Heidelberg, New

York, Tokyo.

Holling, C.S. (1966). The functional response of invertebrate predators to prey density. Mem. Ent. Soc. Canada 48.

Hoppensteadt, F. (1974). An age dependent epidemic model. J. Franklin Inst. 297, 325–333

Hoppensteadt, F. (1975). Mathematical Theories of Populations: Demographics, Genetics and Epidemics. Regional Conference Series in Applied Mathematics 20. SIAM

Huang, W. (thesis). Studies in differential equations and applications. Ph.D. Thesis, The Claremont Graduate School (December 1989), Claremont, CA.

Huang, W., K. Cooke, and C. Castillo-Chavez. (in press). Stability and bifurcation for a multiple group model for the dynamics of HIV/AIDS transmission. SIAM J. Appl. Math..

Hyman, J.M. and E.A. Stanley. (1988). A risk base model for the spread of the AIDS virus. Math. Biosci. 90, 415–473.

Hyman, J.M. and E.A. Stanley. (1989). The effects of social mixing patterns on the spread of AIDS. In Mathematical Approaches to Problems in Resource Management and Epidemiology, C. Castillo-Chavez, S.A. Levin, and C. Shoemaker (eds.). Lecture Notes in Biomathematics 81, Springer-Verlag, Berlin, Heidelberg, New York and Tokyo.

Imagawa, D.T.; H.L. Moon, S.M. Wolinsky, K. Sano, F. Morales, S. Kwok, J.J. Sninsky, P.G. Nishanian, J. Giorgi, J.L. Fahey, J. Dudley, B.R. Visscher, R. Detels (1989). Human immunodeficiency virus type 1 infection in homosexual men who remain seronegative for prolonged periods. New Engl. J. of Med. 320, 1458–1462

Kermack, W.O.; McKendrick, A.G. (1927). A contribution to the mathematical theory of epidemics. Proc. Roy. Soc. A 115, 700–721

Kermack, W.O.; McKendrick, A.G. (1932). Contributions to the mathematical theory of epidemics. II.– The problem of endemicity. Proc. Roy. Soc. A 138, 55–83

Kermack, W.O.; McKendrick, A.G. (1933). Contributions to the mathematical theory of epidemics. III.– Further studies of the problem of endemicity. Proc. Roy. Soc. A 141, 94–122

Lange, J.M.A., D.A. Paul, H.G. Huisman, F. De Wolf, H. Van den Berg, C.A. Roel, S.A. Danner, J. Van der Noordaa, and J. Goudsmit. (1986). Persistent HIV antigenaemia and decline of HIV core antibodies associated with transition to AIDS. Brit. Med. J. 293, 1459–1462.

Liu, W-m. H.W. Hethcote, and S.A. Levin. (1987). Dynamical behavior of epidemiological models with nonlinear incidence rates. J. Math. Biol. 25(4), 359–380.

Liu, W-m., S.A. Levin, and Y. Iwasa. (1986). Influence of nonlinear incidence rates upon the behavior of SIRS epidemiological models. J. Math. Biol. 23, 187–204.

Ludwig, D., Walters, C. (1985). Are age structured models appropriate for catch-effort

data? Can. J. Fish. Aquat. Sci. 40, 559–569.

Ludwig, D. (1989). Small models are beautiful: efficient estimators are even more beautiful. In (C. Castillo-Chavez, S.A. Levin, and C. Shoemaker, eds.) Mathematical Approaches to Problems in Resource Management and Epidemiology. Lecture Notes in Biomathematics 81, Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, 274–283.

May, R.M. and R.M. Anderson. (1989). The transmission dynamics of human immunodeficiency virus (HIV). Phil. Trans. R. Soc. London B 321, 565–607.

May, R.M., R.M. Anderson, and A.R. McLean. (1988). Possible demographic consequences of HIV/AIDS epidemics: I. Assuming HIV infection always leads to AIDS. Math. Biosci. 90, 475–506.

May, R.M., R.M. Anderson, and A.R. McLean. (1989). Possible demographic consequences of HIV/AIDS epidemics: II. Assuming HIV infection does not necessarily lead to AIDS. In Mathematical Approaches to Problems in Resource Management and Epidemiology. C. Castillo-Chavez, S.A. Levin, and C. Shoemaker (eds). Lecture Notes in Biomathematics 81, Springer-Verlag, Berlin, Heidelberg, New York and Tokyo.

McKendrick, A.G. (1926). Applications of mathematics to medical problems. Proc. Edin. Math. Soc. 44, 98–130

Michaelis, L., M.I. Menten. (1913). Die Kinetik der Invertinwirkung. Biochem. Z. 49, 333–369

Miller, R.K. (1971). Nonlinear Volterra Integral Equations. Benjamin, Menlo Park.

Monod, J. (1942). Recherches sur la croissance des cultures bacteriennes. Hermann, Paris

Pedersen, C., C.M. Nielsen, B.F. Vestergaard, J. Gerstoft, K. Krogsgaard, J.O. Nielsen. (1987). Temporal relation of antigenaemia and loss of antibodies to core antigens to development of clinical disease in HIV infection. Brit. Med. J. 295, 567–569

Pugliese, A (preprint): An $S \to E \to I$ epidemic model with varying population size.

Reddingius, J. (1971). Notes on the mathematical theory of epidemics. Acta Biotheor. 20, 125–157

Salahuddin, S.Z., J.E. Groopman, P.D. Markham, M.G. Sarngaharan, R.R. Redfield, M.F. McLane, M. Essex, A. Sliski, and R.C. Gallo. (1984). HTLV-III in symptom-free seronegative persons. Lancet 2, 1418–1420.

Sattenspiel, L. and C. Castillo-Chavez. (1990). Environmental context, social interactions, and the spread of HIV. American Journal of Human Biology 2, 397-417

Sharpe, F.R. and A.J. Lotka. (1911). A problem in age-distribution. Phil. Mag. 21, 435–438

Thieme, H.R. (1990). Semiflows generated by Lipschitz perturbations of non-densely defined operators. Differential and Integral Equations 3, 1035–1066

Thieme, H.R. and C. Castillo-Chavez. (1989). On the role of variable infectivity in the dynamics of the human immunodeficiency virus epidemic. Mathematical and Statistical Approaches to AIDS Epidemiology (C. Castillo-Chavez, ed.). Lecture Notes in Biomathematics 83, 157–176

Waltman, P. (1974) Deterministic Threshold Models in the Theory of Epidemics. Lecture Notes in Biomathematics 1, Springer

Webb, G.F. (1985). Theory of Nonlinear Age-Dependent Population Dynamics. Marcel Dekker, New York.