A model of HIV-1 pathogenesis that includes an intracellular delay

Patrick W. Nelson a,b, James D. Murray c, Alan S. Perelson b,*

a Department of Mathematics, Duke University, Durham, NC 27708, USA
b Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Mail Stop K710, Los Alamos, NM 87545, USA
c Department of Applied Mathematics, University of Washington, Seattle, WA 98195, USA

Received 20 July 1999; received in revised form 14 October 1999; accepted 18 October 1999

Abstract

Mathematical modeling combined with experimental measurements have yielded important insights into HIV-1 pathogenesis. For example, data from experiments in which HIV-infected patients are given potent antiretroviral drugs that perturb the infection process have been used to estimate kinetic parameters underlying HIV infection. Many of the models used to analyze data have assumed drug treatments to be completely efficacious and that upon infection a cell instantly begins producing virus. We consider a model that allows for less than perfect drug effects and which includes a delay in the initiation of virus production. We present detailed analysis of this delay differential equation model and compare the results to a model without delay. Our analysis shows that when drug efficacy is less than 100%, as may be the case in vivo, the predicted rate of decline in plasma virus concentration depends on three factors: the death rate of virus producing cells, the efficacy of therapy, and the length of the delay. Thus, previous estimates of infected cell loss rates can be improved upon by considering more realistic models of viral infection. © 2000 Elsevier Science Inc. All rights reserved.

Keywords: HIV; Delay; Viral life cycle; T-cells

1. Introduction

In recent years, clinical research combined with mathematical modeling has enhanced progress in the understanding of HIV-1 infection. The introduction of a new class of drugs, protease
inhibitors, used in combination with previously approved drugs such as zidovudine (AZT) and lamivudine (3TC) has shown much promise in reducing the detectable levels of viral RNA in the plasma. The introduction of these potent antiviral agents has opened the door to defining kinetic parameters associated with HIV-1 dynamics in infected individuals. Using mathematical models to interpret clinical data from HIV-1 infected patients treated with these potent antiviral agents has led to an estimate of the half-life of productively infected cells of 1.6 days or less and of the clearance rate of free virions from the plasma of six hours or less [1–3]. These initial estimates were claimed to be minimal estimates because the models used to analyze the patient data assumed the drug was completely effective. Refining these estimates will improve our quantitative description of viral dynamics and our understanding of the speed at which drug resistance emerges. Also, the dynamics of the models assumed that cells upon infection instantly begin producing virus. Both the assumption of 100% drug efficacy and of instantaneous production of virus upon infection are relaxed in this paper.

Ho et al. [4] and Wei et al. [3] developed models that were applied to data from an experiment where patients received an antiretroviral drug. Before drug therapy was initiated the patients were determined to have a quasi steady state concentration of virus in their plasma, suggesting that the rate of viral production equaled the rate of virus clearance. After administration of the drug, blood samples were taken for several days and plasma viral RNA levels were measured. The results showed that when virus production was interfered with by drugs the virus declined approximately exponentially. The data, when analyzed in the context of the model, allowed an estimate to be made of the quasi-steady state rate of viral production. The estimated production level was much higher then previously thought and provided an explanation for the ease by which HIV could mutate and escape therapy when only a single drug was used. Extensions of this work showed that after drug is applied there is a delay before virus begins to decline [1]. This delay is due to pharmacological events, the mechanism of protease inhibitor action that allows pre-existing virus particles to continue infecting cells and delays in the process of producing virus particles [1,5,6]. The viral decline is at first very rapid and then after about two weeks slows [2]. Thus there is a rapid first phase followed by a slower second phase of viral decline. Analysis of the data in the context of models showed that the viral decline during the first, rapid, phase could be used to estimate both \( c \), the virus clearance rate constant and \( \delta \), the decay rate of productively infected T-cells [1]. In the model, the asymptotic rate of the exponential decline of virus concentration is \( e^{-\delta t} \). In order to get this result, it was assumed the drug was completely effective and that the virus clearance rate constant, \( c \), was much larger than \( \delta \), and therefore associated with processes that decayed more quickly and consequently could be ignored after a brief initial period. It was later shown [7] that if the drug or combination of drugs was less then perfect, the exponential decline of the first phase could be represented by \( e^{-\delta n_c t} \) where \( n_c \) is the effectiveness of the drug combination with a value of one for complete efficacy and zero for no effect. In this paper we study the changes in these approximations when a constant delay is introduced in the model and the drug, a protease inhibitor, is assumed to have variable efficacy. Four models have been developed that attempt to generalize the early models by Wei et al. [3] and Perelson et al. [1] by including delays [5,6,8,9]. Grossman et al. [8] assume that infected cells die after a delay, given by a gamma distribution, and allow drugs to be partially effective. Tam [9] assumes that infected cells produce virus after a fixed delay and shows that including this delay does not affect the stability properties of infected steady state of the model for reasonable parameter values. The
effects of drug therapy are not studied. Herz et al. [5] examined the effect of including a constant delay in the source term for productively infected T-cells but their analysis was limited to the case where the drug treatment was assumed completely effective. Mittler et al. [6] also assumed that there was a delay between initial infection and the start of virus production, a period that virologists call the *eclipse phase* of the viral life cycle, but rather than using a fixed delay they assumed that the delay varied between cells according to a gamma distribution. They used this model to examine if the parameters that characterize viral dynamics could be identified from biological data and then compared the parameter estimates obtained in a previous model without delay and those obtained with their delay model [10]. These works extended the ability of models to describe the relevant biological processes and addressed the implications of ignoring the intracellular delays that are part of the viral life cycle. In the models of Herz et al. [5] and Mittler et al. [6], the drug was assumed to completely block viral production. This assumption, we will show, cancels out the effect the intracellular delay has on the decay rate of virus producing T-cells. Hence, while noticing a change in viral dynamics during the eclipse phase, it was assumed the delay had no effect during the first phase’s exponential decline. This paper’s main focus is to examine the change in dynamics when a delay of discrete type is considered and the drug, a protease inhibitor, is assumed less than perfect. We have already shown [6,11] that when fitting a model to experimental data there are significant changes in the values of the infected T-cell loss rate and the viral clearance rate constant if the model includes a delay. Here, we derive a new approximation, $e^{-\delta n_p C(\tau, n_p, \delta) \tau}$, for the asymptotic first-phase viral decay when a constant intracellular delay $\tau$ is considered and protease inhibitor treatment is applied. Here the new factor is $C(\tau, n_p, \delta) = [1 + (1 - n_p)\delta \tau]^{-1}$, where $n_p$ is the effectiveness of the protease inhibitor. We provide a detailed analytical study of the model equations and examine the effects of adding a delay on the stability of solutions, something that was neglected in the analysis of previous delay models (see Table 1).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Mean/Initial value</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T(t)$</td>
<td>Uninfected T-cells</td>
<td>180 cells/mm$^3$</td>
<td>[4]</td>
</tr>
<tr>
<td>$T^*(t)$</td>
<td>Productively infected T-cells</td>
<td>$\sim 2%$ T-cells</td>
<td>[4]</td>
</tr>
<tr>
<td>$V_I(t)$</td>
<td>Infectious virus</td>
<td>$134 \times 10^3$ virions/ml</td>
<td>[4]</td>
</tr>
<tr>
<td>$V_N(t)$</td>
<td>Non-infectious virus</td>
<td>0 virions/ml</td>
<td>[4]</td>
</tr>
<tr>
<td>$n_p(t)$</td>
<td>Protease inhibitor efficacy</td>
<td>(0,1)</td>
<td>Variable</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>T-cell source term</td>
<td>$0 - 10$ cells/mm$^3$</td>
<td>[30]</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>Death rate of target T-cells</td>
<td>0.03/day</td>
<td>[31]</td>
</tr>
<tr>
<td>$k$</td>
<td>Viral infectivity rate</td>
<td>$3.43 \times 10^{-5}$ ml/virions/day</td>
<td>[4]</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Death rate of an infected T-cell</td>
<td>0.5/day</td>
<td>[1]</td>
</tr>
<tr>
<td>$N$</td>
<td>Bursting term for viral production after lysis</td>
<td>480 virions/cell</td>
<td>[1]</td>
</tr>
<tr>
<td>$c$</td>
<td>Clearance rate of virus</td>
<td>3/day</td>
<td>[1]</td>
</tr>
</tbody>
</table>

*The parameter values used in generating the figures shown later in the paper are given and the references from which they were derived. The values for $c$ and $\delta$ are the published estimates obtained from applying non-delay models to data from drug perturbation experiments.*
2. General model

The models in this paper only deal with dynamics occurring after drug treatment. The models are not designed to examine the progression from time of infection until drug initiation [12,13]. The study of HIV-1 dynamics has focused on the use of a general set of model equations that keep track of different cell populations and virus. In one case the model has been linear [4], while in most cases the model includes non-linear effects [14–17]. In most cases the models, while trivial mathematically, have provided important results when fitted to experimental data [3,4]. Models have distinguished between infected and non-infected T-cells [1], included other populations of immune cells, such as macrophages [2], looked at variability in certain crucial parameters, such as the infectivity rate, \( k \) [18], viral burst size, \( N \), and included drug treatments [7,19,20]. The general form of the non-delay model includes compartments for the densities of uninfected T-cells, \( T \), infected T-cells that are producing virus, \( T^* \), infectious virus, \( V_I \), and non-infectious virus (produced by the action of a protease inhibitor), \( V_{NI} \), and is written as

\[
\begin{align*}
\frac{dT}{dt} &= \lambda - \delta_I T - kV_I T, \\
\frac{dT^*}{dt} &= kV_I T - \delta T^*, \\
\frac{dV_I}{dt} &= (1 - n_p) N \delta T^* - c V_I, \\
\frac{dV_{NI}}{dt} &= n_p N \delta T^* - c V_{NI},
\end{align*}
\]

where \( T, T^*, V_I \) and \( V_{NI} \) are in \( \mathbb{R}^+ \) and all the parameters are in \( \mathbb{R}^+ \). \( \delta \) is the decay rate of virus producing T-cells, \( c \) is the viral clearance rate constant, and \( n_p \) represents the drug efficacy of a protease inhibitor. The system (1) models the dynamics of virus levels in the plasma during drug treatment. The term \( 1 - n_p \) represents the level of leakiness of a protease inhibitor, a drug that inhibits the cleaving of viral polyproteins and renders newly produced virions non-infectious. Thus, if \( n_p = 1 \), the protease inhibitor is 100% effective and no infectious virus particles are produced. Analysis of (1) has shown that allowing \( T \) to vary dynamically has not yielded significantly different behavior in the total viral output, \( \hat{V} = V_I + V_{NI} \), over a period of a week or so [21,22]. Because of this, we will analyze a model in which \( T \) is held constant at the value \( T_0 \).

2.1. Approximation to dominant eigenvalue

The eigenvalues of (1) determine the rate of viral decline, after therapy is initiated. Assuming that the system is in steady state before therapy is initiated so that \( c = NkT_0 \) and that \( T = T_0 \), then the eigenvalues of the final three equations of (1) can be expressed as [22]

\[
\lambda_{1,2} = -\frac{\delta + c}{2} \pm \frac{1}{2} \sqrt{(\delta + c)^2 - 4n_p \delta c}, \quad \lambda_3 = -c.
\]

If one assumes that the viral clearance rate constant, \( c \), is significantly larger then the decay rate of productively infected T-cells, \( \delta \), then the eigenvalue with the largest real part related can be estimated to first order by [7,22]
\[
\lambda_1 \sim \delta n_p.
\]

This shows when therapy is less than perfect the first phase slope of viral decline is not simply the death rate, \( \delta \), of productively infected T-cells. Previously, the value of \( \delta \) has been estimated from experimental data, assuming a completely effective drug [3,4]. Relaxing this assumption of a perfect drug will change the estimate of \( \delta \). However, as we show below, when the drug efficacy is not 100\%, delays between the time of cell infection and viral production will also influence the estimate of \( \delta \) because \( \lambda_1 \) now becomes \( \lambda_1 \sim -\delta n_p C(\tau, n_p, \delta) \) where \( C(\tau, n_p, \delta) \) is a new factor introduced by the delay.

### 3. Intracellular delay model

In HIV-1 infection, the virus life cycle plays a crucial role in disease progression. The binding of a viral particle to a receptor on a target cell initiates a cascade of events that can ultimately lead to the target cell becoming productively infected, i.e. producing new virus. The previous model (1) assumed this process to occur instantaneously. In other words, it is assumed that as soon as virus contacts a target cell the cell begins producing virus. However, in reality there is a time delay between initial viral entry into a cell and subsequent viral production. Previous work has shown that this delay needs to be taken into account to accurately determine the half life of free virus from drug perturbation experiments, but if the drug is assumed to be completely efficacious, the delay does not affect the estimated rate of decay of viral producing T-cells [5,6,10]. Here we show that the delay does affect the estimated value for the infected T-cell loss rate when one assumes the drug is not completely effective.

We incorporate the intracellular delay in the model given by (1) by assuming that the generation of virus producing cells at time \( t \) is due the infection of target cells at time \( t - \tau \), where \( \tau \) is a constant. As before, we assume uninfected T-cells remain constant, i.e., \( T = T_0 \). The equations describing this model are then

\[
\begin{align*}
\frac{dT^*}{dt} &= k T_0 V_1(t - \tau) e^{-mt} - \delta T^*, \\
\frac{dV_1}{dt} &= (1 - n_p) N \delta T^* - c V_1, \\
\frac{dV_{NI}}{dt} &= n_p N \delta T^* - c V_{NI},
\end{align*}
\]

where the term \( V_1(t - \tau) \) allows for the time delay between contact and virus production and \( e^{-mt} \) allows for death of an infected T-cells before it becomes productively infected. The rate constant for infection, \( k \) is assumed constant because the drug we are modeling, a protease inhibitor, does not affect \( k \). If a reverse transcriptase inhibitor were being used then the appropriate model would have \( k \) in the \( dT^*/dt \) equation replaced by \( k(t - \tau) \). Notice the equation for \( V_{NI}(t) \) can be easily solved once the solution for the first two equations is known. Analysis of a more general form of this delay model, which included uninfected T-cells and non-linearities can be found in [21]. The following sections will examine the model’s characteristic equation. Our main focus is to provide an expression for the dominant eigenvalue, i.e., the eigenvalue with the largest real part.
3.1. Analysis of delay model

The model (4) in vector–matrix notation is
\[
\frac{dy(t)}{dt} = B_0 y(t - \tau) + C_0 y(t),
\]
where \( \tau \) is a constant, \( B_0 \) and \( C_0 \) are \( n \times n \) matrices defined as
\[
B_0 = \begin{pmatrix}
0 & kT_0 e^{-mt} & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
\end{pmatrix}, \quad C_0 = \begin{pmatrix}
-\delta & 0 & 0 \\
(1 - n_p)N\delta & -c & 0 \\
n_pN\delta & 0 & -c \\
\end{pmatrix}
\]
and
\[
y(t) = \begin{pmatrix}
T^*(t) \\
V_i(t) \\
V_{NI}(t) \\
\end{pmatrix}.
\]
The initial conditions, on the interval \( 0 \leq t \leq \tau \) are defined as
\[
y(t) = g(t),
\]
where
\[
g(t) = \begin{pmatrix}
T_{ss} \\
V_{ss} \\
0 \\
\end{pmatrix}.
\]
Using the results in [23] it is easy to show that a system of the form (5), with initial conditions (6), has a unique solution which is continuous for \( t \geq 0 \) and which satisfies (5) for \( t > \tau \).

3.1.1. Characteristic equation

Time delays when used in population dynamic models have been shown to create fluctuations in population size due to the generation of instabilities [24,25]. Here, we will show that the eigenvalues of the delay model remain in the left half plane for all values of the delay. With this being the case the dominant eigenvalue will then be the negative eigenvalue whose real part is closest to zero.

Introducing a delay of discrete type in a system of differential equations changes the structure of the solution as determined by the characteristic equation. In the case when no delay is present a linear system of \( N \) differential equations will have \( N \) solutions and a characteristic polynomial of degree \( N \). When a delay is included the characteristic equation is no longer a polynomial but rather takes the form (see [26])
\[
H(\lambda) = G(\lambda) + K(\lambda)e^{-\lambda\tau} = 0,
\]
where \( G(\lambda) \) and \( K(\lambda) \) are polynomials and
\[
P(\lambda) = G(\lambda) + K(\lambda) = 0,
\]
is the corresponding characteristic equation for the system with no delay. Since the model we consider is of retarded and not neutral type, \( G(\lambda) \) is of higher order than \( K(\lambda) \). This ensures (see
[27]) that as \( \tau \to 0 \), \( n \) of the roots of (7) tend towards the roots of \( P(\lambda) = 0 \), while the remainder of the roots have real parts tending to large negative values. Hence as \( \tau \to 0 \) the stability properties of the equilibrium points in the non-delay model (1) are recovered.

The characteristic function of (4), determined from the Laplace transform [23], is

\[
H_1(\lambda, \tau) = (\lambda^2 + (\delta + c)\lambda + \delta c - \eta e^{-\lambda \tau})(\lambda + c),
\]

where \( \eta = (1 - n_p)\delta N k T_0 e^{-\mu \tau} \). Further, if we assume that the system is in quasi-steady state before drug therapy, i.e., \( c = N k T_0 \), and that the death during transition from infected to productively infected is negligible, i.e., \( m \ll 1 \), we can then simplify \( \eta \) to

\[
\eta = (1 - n_p)\delta c.
\]

From (9) one eigenvalue is equal to \(-c\) and hence always negative. In what follows we determine the other eigenvalues from the solutions of

\[
H(\lambda, \tau) = \lambda^2 + (\delta + c)\lambda + \delta c - \eta e^{-\lambda \tau} = 0.
\]

The following results show that the eigenvalues will remain in the left half plane and show the location of the largest eigenvalue. Set \( \lambda = \mu + iv \) where \( \mu, v \in \mathbb{R} \) and substitute into (11) to get

\[
H(\mu, v, \tau) = [\mu^2 - v^2 + (\delta + c)\mu + \delta c - \eta e^{-\mu \tau} \cos v \tau] + i[2\mu v + v(\delta + c) + \eta e^{-\mu \tau} \sin v \tau],
\]

where \( H(\mu, 0, \tau) \equiv H(\mu) = \mu^2 + (\delta + c)\mu + \delta c - \eta e^{-\mu \tau} \).

**Lemma 1.** If \( H(\mu) > 0 \) then there is no \( v \in \mathbb{R} \) such that \( H(\mu + iv) = 0 \). Moreover, if \( H(\mu) = 0 \) then \( H(\mu + iv) = 0 \) if and only if \( v = 0 \).

**Proof (by contradiction).** Assume that \( H(\mu + iv) = 0 \). Then \( H(\mu + iv) \) can be rewritten in terms of its real and imaginary parts as

\[
\mu^2 - v^2 + \mu(\delta + c) + \delta c = \eta e^{-\mu \tau} \cos v \tau \quad \text{and} \quad 2\mu v + v(\delta + c) = -\eta e^{-\mu \tau} \sin v \tau.
\]

Taking the squares of the absolute values of both equations and adding together gives

\[
[v^2 + (\mu + c)^2][v^2 + (\mu + \delta)^2] = \eta^2 e^{-2\mu \tau}
\]

which implies that

\[
(\mu + c)^2(\mu + \delta)^2 - \eta^2 e^{-2\mu \tau} = -v^4 - v^2[(\mu + c)^2 + (\mu + \delta)^2].
\]

The left-hand side of this equation equals

\[
H(\mu)[(\mu + c)(\mu + \delta) + \eta e^{-\mu \tau}] = H(\mu)[H(\mu) + 2\eta e^{-\mu \tau}],
\]

which is always positive given \( H(\mu) > 0 \). The right-hand side is always negative leading to a contradiction. Therefore, \( H(\mu + iv) \neq 0 \) given \( H(\mu) > 0 \). The second part of the proof is easily seen. \( \square \)

Note that \( H(\mu) \) can be rewritten as \( H(\mu) = \mu^2 + (\delta + c)\mu + \delta c(1 - (1 - n_p)e^{-\mu \tau}) \). Since \( \delta > 0 \), \( c > 0 \), and \( 0 \leq n_p \leq 1 \), \( H(\mu) > 0 \) for \( \mu > 0 \). Therefore by Lemma 1 there are no eigenvalues with non-negative real part. Moreover, a special case of the above lemma shows that if \( H(0) > 0 \), then there is no \( v \in \mathbb{R} \) such that \( H(iv) = 0 \). Hence there are no purely imaginary roots.
In the remainder of this section we will show that there exists one real eigenvalue, called \( l_0 \), on the interval \( [-c, 0) \) (see Fig. 1) and that for certain parameter values all other eigenvalues, either real or complex, have real parts at least an order of magnitude smaller than \( l_0 \). (This is similar to showing the approximation of \( k_1 \) was valid given \( c_d \).)

**Theorem 1.** For each \( \tau \geq 0 \) there is exactly one \( l_0 \) such that \( H(l_0) = 0 \). Moreover, when \( H(\mu) > 0 \) for \( \mu > l_0 \) then \( l_0 \) is the unique eigenvalue with the largest real part.

**Proof.**

\[
H(\mu) = \mu^2 + (\delta + c)\mu + dc - \eta e^{-\mu \tau} \quad \text{and} \quad H'(\mu) = 2\mu + (\delta + c) + \eta \tau e^{-\mu \tau}.
\]

For \( \mu \in [-c, -\delta] \), \( H(\mu) < 0 \) and for \( \mu \geq -(c + \delta)/2 \), \( H'(\mu) > 0 \). Hence \( H(\mu) \) is a monotonic increasing function on the interval \( \left[ -\frac{(c + \delta)}{2}, \infty \right) \). Therefore, there exists exactly one unique \( \mu \) such that \( H(\mu) = 0 \) for \( \mu \in \left[ -c, -\delta \right) \). Now for \( \mu > \mu \), \( H(\mu) > 0 \) and by Lemma 1 cannot be the real part of any eigenvalue. Thus, \( \mu \) is the largest real eigenvalue. \( \square \)

Moreover, by direct computation, \( H(0) = n_p dc > 0 \). Thus, by Lemma 1, \( \mu < 0 \). In other words, \( \mu(\tau) \in \left[ \min(-c, -\delta), 0 \right) \), and further assuming \( c \gg \delta \), \( \mu(\tau) \in [-c, 0) \). Thus, there is only one real root in the interval, \( [-c, 0) \), To complete this analysis we must show that this interval contains no complex eigenvalues. If there exists complex eigenvalues in \( [-c, 0) \) then we would be unable to claim that the real eigenvalue, \( \mu \), is the slowest varying, or dominant, eigenvalue since the complex eigenvalues would have real parts of the same order of magnitude as \( \mu \). Hence we would be unable to isolate a single dominant term.
3.2. Numerical analysis of location of complex roots

By definition if there exists a pair of complex eigenvalues in \([-c, 0]\) then given \(\mu \in [-c, 0]\), \(H(\mu + iv) = 0\) for some \(v \in \mathbb{R}\). We can write the eigenvalue equation \(H(\lambda)\) in terms of two real equations defined as

\[
\mu^2 + (\delta + c)\mu + \delta c - v^2 = \eta e^{-\mu t} \cos vt, \tag{13a}
\]

\[
2\mu v + v(\delta + c) = -\eta e^{-\mu t} \sin vt. \tag{13b}
\]

We can rewrite (13a) as

\[
H(\mu) - v^2 = \eta e^{-\mu t} \cos vt, \tag{14}
\]

where \(H = \mu^2 + (c + \delta)\mu + \delta c\). It is easily seen that on the interval \(\mu \in [-c, -\delta]\), \(H(\mu) \leq 0\). We can also define some constraints to validate numerical results.

3.2.1. Constraint I

Given \(\mu \in [-c, -\delta]\), the left-hand side of (14) is always less than or equal to zero. For \(0 < \nu \tau < \pi/2\) the right-hand side of (14) is positive and hence there can be no solution to the equation. Subsequently, for \(0 < \nu \tau < \pi/2\) and \(\mu \in [-\delta, 0]\), (13b) cannot be satisfied. In general, for (13a) and (13b) to have complex roots, the following has to hold:

\[
\max(H - v^2) \geq \min(\eta e^{-\mu t} \cos vt). \tag{15}
\]

For \(\mu \in [-c, -\delta]\), \(\min(\eta e^{-\mu t} \cos vt) = -\eta e^{ct}\) and \(\max(H - v^2) = -v^2\). Thus, (15) implies that for a root to exist, \(v^2 \leq \eta e^{ct}\). Since there is no solution if \(0 < \nu \tau < \pi/2\), \(\pi/2 \tau < \nu < \sqrt{\pi/\eta} e^{ct}\) is a necessary condition for a non-zero root. Hence, if \(\sqrt{\pi/\eta} e^{ct} < \pi/2\tau\) there can be no non-zero solution. Rearranging, we find that if

\[
0 < \tau \sqrt{\pi/\eta} < \frac{\pi}{2\sqrt{\pi}}
\]

then (13a) and (13b) has no solution for \(v \neq 0\).

3.2.2. Constraint II

As pointed out to us by D. Calloway, Cornell University, the derivative of (13a) with respect to \(\mu\) is \(1/\nu\) times the left-hand side of (13b). Therefore,

\[
\frac{\partial}{\partial \mu}(\eta e^{-\mu t} \cos vt) = -\frac{1}{\nu} \eta e^{-\mu t} \sin vt,
\]

and hence a condition for the system (13a) and (13b) to have a solution is

\[
\nu \tau = \tan \nu \tau. \tag{16}
\]

This result gives an infinite number of solutions due to the periodicity of the tangent function. Solving this equation numerically, we find \(\nu \tau = 0, \sim 4.4934, \sim 7.7235, \ldots\). But, from (14), we can write

\[
v^2 < \eta e^{ct} + H(\mu) \leq \eta e^{ct}, \tag{17}
\]
which gives a bound on \( v \) necessary for (13a) and (13b) to have a solution. For example, with \( \delta = 0.5, c = 3 \) and a drug efficacy \( n_p = 0.7, \eta = 0.45 \). Hence for \( \tau < 1.5, v < 6.36 \) and hence the solutions with \( \nu \tau \geq 7.7235 \) can be excluded. Thus the constraints provide a finite range of \( v \) in which to search for solutions, simplifying the following numerical calculations.

3.2.3. Numerical calculations

Using the above results, we examine the possible existence of complex eigenvalues numerically. We restrict our attention to the case \( c = 3 \). Discretizing \( \mu, v, \tau, \) and \( n_p \) with steps sizes of 0.01 and using Matlab we find for \( \mu \in [-3, 0) \), and \( n_p \in (0.7, 0.9) \) there exists no solution of (13a) and (13b) with \( v \neq 0 \) for determined ranges of \( \tau \). As an example, with \( \delta = 0.5, c = 3 \) and \( n_p = 0.8 \), numerically we find no solutions of (13a) and (13b) when \( \tau < 0.825 \). This agrees surprisingly well with constraint I, which implies no solutions given \( \tau < 0.828 \). If we use a higher efficacy of drug treatment, say \( n_p = 0.95 \) then numerically we can show there are no solutions with \( \tau < 1.18 \) and using constraint I, we find no solutions for \( \tau < 1.105 \). Varying \( \delta \) between (0.25, 0.75) adjusts the upper bound on \( \tau \). For example, with \( \delta = 0.35, c = 3 \) and \( n_p = 0.95 \), (13a) and (13b) is not satisfied with \( \mu > -3 \) given \( \tau < 1.15 \). These numerical results compared with the above constraints for \( \tau \) and \( v \) provide sufficient evidence to support the conclusion that there exists only one unique root which is real and dominant while also providing sufficient reasoning for neglecting, to first approximation, all eigenvalues of the delay model except the unique real eigenvalue.

The value of \( c \) may be higher than \( c = 3 \) [10,28,29]. If this is so, then a high drug efficacy is required to guarantee that there will not be any complex eigenvalues. For example, if \( c = 5 \) and \( \tau = 1 \) then only with a drug efficacy of 0.92 or higher, will there be one unique real eigenvalue. If \( c \) is much higher, then our results show the existence of complex eigenvalues but the order of magnitude of their real parts is still dominated by the single negative and real eigenvalue. Therefore, it should still be possible to describe the viral decline seen in patients after drug treatment to a good approximation by a single exponential but the shoulder region no longer depends on \( c \) alone but a combination of \( c \) and the complex eigenvalues.

4. Delay effects on dominant eigenvalue

With the results of the previous section we can now present the main finding of this work: an asymptotic representation for the dominant eigenvalue and how the delay effects its value (see Fig. 1). Since we have shown that for \( \mu \in (-3, 0) \) there are no complex roots for a defined \( \tau \), we can let \( \hat{\mu} \) be the unique root of \( H(\mu, 0, \tau) \equiv H(\hat{\mu}, \tau) = \hat{\mu}^2 + (\delta + c)\hat{\mu} + \delta c - \eta e^{-\hat{\mu} \tau} = 0 \) with the largest real part. Expanding \( e^{-\hat{\mu} \tau} \) in a Taylor series gives

\[
e^{-\hat{\mu} \tau} \approx 1 - \hat{\mu} \tau + \frac{\hat{\mu}^2 \tau^2}{2} + 0(\hat{\mu}^3 \tau^3),
\]

provided that \( |\hat{\mu} \tau| \) is sufficiently small. From our previous results, we have shown that \( \hat{\mu} \) varies between \([-0.5, 0) \) depending on the parameters used. Also, it has been suggested experimentally that a realistic value of the time delay is on the order of 1 day. Hence, the expansion for the exponential term is reasonably accurate and the accuracy improves as the value of \( \hat{\mu} \to 0 \). Also,
notice from (22), that as $\tau$ increases, $\hat{\mu}$ decreases, hence $|\hat{\mu}\tau|$ remains sufficiently small. Substituting
this expansion into (11) yields

$$\hat{\mu}^2 + \hat{\mu}(\delta + c) + \delta c - \eta (1 - \hat{\mu}\tau + \frac{(\hat{\mu}\tau)^2}{2} + O(\hat{\mu}^3\tau^3)) = 0$$

and solving for $\hat{\mu}$ gives

$$\hat{\mu} = \frac{1}{2(1 - \eta \tau^2/2)} \left[ - (\delta + c + \eta \tau) \pm (\delta + c + \eta \tau) \sqrt{1 - \frac{4(\delta c - \eta)(1 - c\eta \tau^2/2)}{(\delta + c + \eta \tau)^2}} \right].$$

The square root term can be approximated by the first two terms of its binomial expansion, since for the parameter values of interest

$$\left| \frac{4(\delta c - \eta)(1 - \frac{\eta \tau^2}{2})}{(\delta + c + \eta \tau)^2} \right| < 1.$$}

Using this and taking the positive term we get

$$\hat{\mu} \sim -\frac{\delta c - \eta}{\delta + c + \eta \tau}.$$

Assuming $c \gg \delta$ [2,10,29], and using the definition of $\eta$ (10), we find

$$\hat{\mu} \sim -\frac{\delta n_p}{1 + (1 - n_p)\delta \tau} = -\delta n_p C(\tau, n_p, \delta),$$

where

$$C(\tau, n_p, \delta) = \frac{1}{1 + (1 - n_p)\delta \tau} < 1$$

is the correction introduced by incorporating a delay. Notice the rate of decay, $\hat{\mu}$, is proportional to the product of the rate constant for death of virus producing T-cells, $\delta$, and the efficacy of the protease inhibitor, $n_p$, but now this product is reduced by the delay since as $s$ gets bigger the value of $\hat{\mu}$ gets smaller.

We illustrate the effects of delay in Figs. 2 and 3. When $n_p \neq 1$, the delay reduces the long term rate of decline of the viral load and hence any estimate of $\delta$ that is made which does not take the delay into account will be flawed. The effect, however, is not large. For example, if $\tau = 1$ day, a typical value, and $n_p = 0.8$ then the correction due to the delay

$$C(\tau, n_p, \delta) = \frac{1}{1 + 0.2\delta}.$$ For $\delta = 0.5 \text{ day}^{-1}$ [1], $C = 0.91$. If the drug is only 50% efficacious, which would be remarkably low, then $C = 0.8$. Hence if the intracellular delay is approximately constant, as assumed here, and if the effects of drug efficacy are properly taken into account, then including the delay will change the estimate of $\delta$ from slope$/n_p$ to slope$/n_p C$, which for the cases considered above yields at most a 25% increase in $\delta$. 
Fig. 2. The top plot shows the numerical solution for viral decay when a perfect drug is used. Notice the slopes of the decay with and without a delay are parallel. The bottom two graphs compare the viral decay with and without a delay assuming a less than perfect drug; with $n_p = 0.8$ in the left panel and $n_p = 0.5$ in the right panel. Both cases show the change in the slope is affected by the delay. The model parameters, besides the ones noted, are given in Table 1.

Fig. 3. The viral decay for $\tau = 0, 0.5, 1$, and 1.5 days when the drug efficacy $n_p = 0.8$. The slopes of the curves generated with different delays differ. Also, the shoulder region on the curve increases as $\tau$ increases, an effect noticed previously [5,6,10].
5. Conclusions

The main goal of this paper was to show that the viral decay seen in HIV-1 infected patients put under antiretroviral therapy depends not only on the lifespan of viral producing cells, \( \delta \), but also the efficacy of the therapy, \( n_p \) and the length of the intracellular delay, \( \tau \), that measures the time between viral infection of a cell and the time the cell begins releasing virus. Here we have focused on the effects of protease inhibitor therapy because detailed data have been published about the kinetics of viral load decline and used to estimate both \( c \) and \( \delta \) [1]. In [1] a model with no delay and \( n_p = 1 \) was used to estimate \( c \) and \( \delta \), and the claim was made that the estimates of both of these parameters were minimal estimates. Our results support the claim that the estimate of \( \delta \) was minimal since the inclusion of drug efficacy and delay both reduce the slope of the viral decay curve. Hence estimates of \( \delta \) made by examining the slope, as was in essence done in [1], underestimate the true value of \( \delta \). Recent work by Mittler et al. [6,10] using a model with delay argue that the published estimates of \( c \) were also minimal as claimed. Previous work of Herz et al. [5] and Mittler et al. [6] have shown that an intracellular delay can affect the length of the ‘shoulder region’ observed in Fig. 3, i.e., the time between initiation of therapy and the observed decay in plasma virus. But in both of these theoretical models, no changes were predicted during the exponential decline of the viral load. We have shown that this is due to assuming the protease inhibitor to be completely effective. When the efficacy, \( n_p = 1 \), the effects of the delay cancel out. The above results imply that the inclusion of a constant delay and allowing the protease inhibitor to be less than perfect changes the estimates for the productively infected T-cell loss rate, \( \delta \), and the viral decay rate, \( c \).

Including a delay also changes the estimated lifespan of a productively infected cell. In previous work without delays the average lifespan of a productively infected cell was calculated as \( 1/\delta \). Here we have taken into consideration that there are two phases in the life of cell from time of infection to death. First, there is an eclipse phase in which the virus enters the cell, uses reverse transcriptase to copy its RNA genome into DNA, and proceeds through a number of steps in its life cycle before new virus particles are produced. The length of the eclipse phase is represented by the delay \( \tau \). Following the eclipse phase, the cell produces virus. In the current model the average lifespan of a virus producing cell is \( 1/\delta \). Thus the total average lifespan of cell from infection to death is \( \tau + 1/\delta \). The precise value of \( \tau \) is not known but current estimates suggest values of 1–1.5 days [1,6,10]. Thus, to a first approximation T-cells infected with HIV-1 might live an average of 2–3 days rather than the average of 1–2 days previously reported [1,3,4]. Currently, work is in progress examining delays in drug activation and modeling the efficacy of the drug as a dynamically varying parameter. Also, there is a need to explore the effects of different types of delays such as distributed delays or multiple discrete delays; some of which is already in progress. The length of the delay is not directly measurable in vivo. Thus, additional in vitro data would prove valuable in expanding our knowledge of the delay characteristics.

Acknowledgements

The authors thank John Mittler and Hans Weinberger for their help and Duncan Calloway for critically reading and commenting on the manuscript. Portions of this work were performed under
the auspices of the US Department of Energy and supported by NIH grant RR06555 (ASP) and an Institute for Mathematics and Its Applications Postdoctoral Fellowship (PWN) with funds provided by the National Science Foundation.

References