MATHEMATICAL ANALYSIS OF AGE-STRUCTURED HIV-1 DYNAMICS WITH COMBINATION ANTIRETROVIRAL THERAPY*

LIBIN RONG†, ZHILAN FENG†, AND ALAN S. PERELSON‡

Abstract. Various classes of antiretroviral drugs are used to treat HIV infection, and they target different stages of the viral life cycle. Age-structured models can be employed to study the impact of these drugs on viral dynamics. We consider two models with age-of-infection and combination therapies involving reverse transcriptase, protease, and entry/fusion inhibitors. The reproductive number $R$ is obtained, and a detailed stability analysis is provided for each model. Interestingly, we find in the age-structured model a different functional dependence of $R$ on $\epsilon_{RT}$, the efficacy of a reverse transcriptase inhibitor, than that found previously in nonage-structured models, which has significant implications in predicting the effects of drug therapy. The influence of drug therapy on the within-host viral fitness and the possible development of drug-resistant strains are also discussed. Numerical simulations are performed to study the dynamical behavior of solutions of the models, and the effects of different combinations of antiretroviral drugs on viral dynamics are compared.

Key words. human immunodeficiency virus type 1, antiretroviral therapy, drug resistance, optimal viral fitness, age-structured model, stability analysis

AMS subject classifications. 35L60, 45D05, 92C37, 92C45, 92C50

DOI. 10.1137/060663945

1. Introduction. Since the discovery of the human immunodeficiency virus type 1 (HIV-1) in the early 1980s, the disease has spread in successive waves to most regions around the globe. It is reported that HIV has infected more than 60 million people, and over a third of them subsequently died [10]. Considerable scientific effort has been devoted to the understanding of viral pathogenesis, host/virus interactions, immune response to infection, and antiretroviral therapy.

Over the last decade, there has been a great effort in the mathematical modeling of HIV infection and treatment strategies. These models mainly investigated the dynamics of the target cells and infected cells, viral production and clearance, and the effects of antiretroviral drugs treatment. Perelson et al. [44] and Ho et al. [22] used a simple mathematical model to analyze a set of viral load data collected from infected patients after the administration of a protease inhibitor, and the virion clearance rate, the rate of loss of productively cells, and the viral production rate were estimated. These estimates were minimal estimates since the effects of antiretroviral drugs were assumed to be 100% effective, and cells were assumed to produce new virus immedi-
ately after they were infected [35, 36]. In order to characterize the time between the infection of target cells and the production of virus particles, an intracellular delay was introduced by Herz et al. [19] in a mathematical model to analyze the clinical data. Subsequently, Culshaw and Ruan [3] investigated the effect of the time delay on the stability of the endemic equilibrium in their model. Criteria were presented to guarantee the asymptotic stability of the infected steady state independent of the time delay. In [35], Nelson, Murray, and Perelson studied a generalized model that included a discrete delay and allowed for less than perfect drug effects. The estimation of kinetic parameters underlying HIV infection was improved by the use of a delay differential equation model. In [31, 32], the authors used a gamma distribution function to describe a continuous delay between infection and viral production and found no change in the estimate of $\delta$, the death rate of productively infected cells. However, Nelson and Perelson [36] extended this model and showed that the constancy of $\delta$ was due to the assumption of 100% drug effectiveness. When drug effectiveness was less than 100%, the estimate of $\delta$ depended on the delay, i.e., the variance and mean of the assumed gamma distribution. Recently, a model including both pharmacokinetics and the intracellular delay has been employed to obtain new estimates of intracellular delay and the antiviral efficacy of ritonavir [7].

Age-structured models have also been developed to study the epidemiology of HIV. Thieme and Castillo-Chavez [52] kept track of an individual’s infection age to study the effect of infection-age-dependent infectivity on the dynamics of HIV transmission in a homogeneously mixing population. Kirschner and Webb [24] proposed a model that incorporated age structure into the infected cells to account for the mechanism of AZT (zidovudine) treatment. Recently, for the within-host dynamics of HIV, age-structured models have received increasing interest due to their greater flexibility in modeling viral production and mortality of infected cells [16, 34]. Nelson et al. [34] considered an age-structured model that allowed for variations in the production rate of virus particles and the death rate of infected T cells. For a specific form of the viral production function and constant death rate of infected cells, the authors performed a local stability analysis of the nontrivial equilibrium point. They used numerical simulations to illustrate that the time to reach the peak viral level depended not only on the initial conditions but also on the speed at which viral production achieves its maximum value. Based on this age-structured model, Gilchrist, Coombs, and Perelson [16] used the various life history trade-offs between viral production and clearance of infected cells to derive the within-host relative viral fitness.

In this article, we develop two age-structured models to study HIV-1 infection dynamics. These models extend the existing age-structured models [16, 24, 34] by incorporating combination therapies to study the influence of antiretroviral therapy on the evolution of HIV-1. The first model includes therapy with a combination of a reverse transcriptase (RT) inhibitor and a protease inhibitor, while the second model includes an entry inhibitor and a protease inhibitor. To account for the fact that reverse transcription takes place in the early stage of infection before an infected T cell produces virus particles, we divide the infected cells into two subclasses. One subclass represents the cells that have been infected by the virus but in which reverse transcription has not been completed. The other subclass contains infected cells that have finished the reverse transcription process and are capable of producing new virions. Our stability analysis is performed for a general form of both the viral production rate and the mortality rate of infected cells. The stability of the infection-free or the infected steady state is shown to depend on the reproductive ratio $R$ being
smaller or greater than 1. The formulation of this reproductive ratio also provides an appropriate measure for the within-host viral fitness, which can be used to explore the optimal viral production rate for which $R$ is maximized.

We also discuss the possible influence of treatment for drug-sensitive strains of HIV-1 on the development of drug-resistant strains of the pathogen. Clinical studies have suggested that prolonged treatment with a single antiretroviral drug may be associated with the emergence of resistant virus [20, 26, 27, 28, 39]. The impact of drug treatment on the dynamics of resistant stains of pathogens has been studied using age-independent mathematical models (see, for example, [11, 26, 55]). We show that if viral production is linked to resistance, then higher treatment efficacy with antiretroviral agents (such as protease inhibitors) may lead to the establishment of multiple viral strains with a wider range of resistance levels.

The organization of the remaining part is as follows. In section 2, we formulate a mathematical model for HIV-1 infection that generalizes the age-structured model proposed in [34] by incorporating an RT inhibitor and a protease inhibitor. Section 3 is devoted to the analysis of our model, including the existence and stability of both the infection-free and the infected steady states. In section 4, another model including therapy with a new class of drugs, fusion/entry inhibitors, is developed. Stability properties of the steady states are also obtained in this section. In section 5, we derive a criterion for invasion by drug-resistant strains and explore how drug treatment may affect the optimal viral fitness of resistant strains. Some numerical simulations are presented in section 6 to illustrate/extend our analytical results. We also compare the treatment effects of these two combination antiretroviral therapies. Section 7 contains concluding remarks.

2. The model with RT and protease inhibitors. HIV infection begins by the attachment of a virus to a CD4$^+$ cell. Inside the cell, the HIV-1 enzyme RT makes a DNA copy of the virus's RNA genome. During this process, if an RT inhibitor is present, then the viral genome will not be copied into DNA, and therefore the host cell will not produce new virus. When the virus replicates, its DNA is read out to produce viral proteins. A large polypeptide is made, and a viral protease is needed to cut the long polypeptide chain into individual components that are needed to produce infectious virus particles. If the HIV-1 protease is inhibited, the newly produced virus will be noninfectious.

From the above description of the HIV life cycle and the roles of various inhibitors, it is clear that the infection age of an infected cell can be important for the study of HIV dynamics under the influence of antiretroviral drug treatment. In [34] the following age-structured model of HIV infection (without drug treatment) was proposed:

\[ \frac{dT}{dt}(t) = s - dT - kVT, \]

\[ \frac{\partial}{\partial t} T^*(a,t) + \frac{\partial}{\partial a} T^*(a,t) + \delta(a)T^*(a,t) = -\delta(a)T^*(a,t), \]

\[ \frac{d}{dt} V(t) = \int_0^\infty p(a)T^*(a,t)da - cV, \]

\[ T^*(0,t) = kVT, \]

where $T(t)$ denotes the concentration of uninfected target T cells at time $t$, $T^*(a,t)$
denotes the concentration of infected T cells of infection age \( a \) (i.e., the time that has elapsed since an HIV virion has penetrated the cell) at time \( t \), and \( V(t) \) denotes the concentration of infectious virus at \( t \). \( s \) is the recruitment rate of healthy T cells, \( d \) is the per capita death rate of uninfected cells, \( \delta(a) \) is the age-dependent per capita death rate of infected cells, \( c \) is the clearance rate of virions, \( k \) is the rate at which an uninfected cell becomes infected by an infectious virus, and \( p(a) \) is the viral production rate of an infected cell with age \( a \).

The functional forms of the viral production kernel, \( p(a) \), and the death rate of infected cells, \( \delta(a) \), need to be determined experimentally [21, 34]. In [34], the authors choose the following function for the production rate:

\[
p(a) = \begin{cases} 
  p^* \left(1 - e^{-\theta(a-a_1)}\right) & \text{if } a \geq a_1, \\
  0 & \text{else,}
\end{cases}
\]

where \( \theta \) determines how quickly \( p(a) \) reaches the saturation level \( p^* \), and \( a_1 \) is the age at which reverse transcription is completed.

To incorporate the two types of treatments mentioned above, we divide the class of infected cells, \( T^*(a,t) \), into two subclasses: \( T^*_{\text{preRT}}(a,t) \) and \( T^*_{\text{postRT}}(a,t) \). \( T^*_{\text{preRT}}(a,t) \) represents the density of cells that have been “infected” by an HIV virion but in which reverse transcription has not been completed at infection age \( a \). An RT inhibitor could allow a preRT cell to revert back to an uninfected cell (because if reverse transcription fails to complete, cellular nucleases will degrade the HIV RNA that entered the cell) or reduce the probability that a preRT cell progresses to the postRT state [9]. \( T^*_{\text{postRT}}(a,t) \) represents the density of infected cells that have progressed to the postRT phase at infection age \( a \). The densities of the preRT and postRT cells are related by a function \( \beta(a) (0 \leq \beta(a) \leq 1) \) that describes the proportion of infected cells that have not completed reverse transcription, i.e.,

\[
T^*_{\text{preRT}}(a,t) = \beta(a) T^*(a,t), \quad T^*_{\text{postRT}}(a,t) = (1 - \beta(a)) T^*(a,t).
\]

We assume that \( \beta(a) \in L^1[0, \infty) \) is a nonincreasing function with the following properties: \( 0 \leq \beta(a) \leq 1; \beta(0) = 1; \beta(a) = 0 \) for \( a \geq a_1; \beta'(a) \leq 0 \) a.e.

Let \( \epsilon_{\text{RT}} \) and \( \epsilon_{\text{PI}} \) denote the efficacy of the therapy with RT inhibitors and protease inhibitors, respectively (\( 0 \leq \epsilon_{\text{RT}}, \epsilon_{\text{PI}} < 1 \)). The efficacy is scaled such that zero represents complete ineffectiveness and unity represents 100% effectiveness. To study the effect of protease inhibitor, we divide the newly produced virus particles into two classes: infectious virions with concentration \( V_I(t) \) and noninfectious virions with concentration \( V_{NI}(t) \). New infectious virus particles are produced at the rate \( \int_0^\infty (1 - \epsilon_{\text{PI}}) p(a) T^*_{\text{preRT}}(a,t) da \).

Let \( \eta(\epsilon_{\text{RT}}) \) denote the rate at which preRT cells revert to the uninfected state due to the failure of reverse transcription. The rate at which preRT cells of all ages become uninfected is then given by \( \int_0^\infty \eta(\epsilon_{\text{RT}}) T^*_{\text{preRT}}(a,t) da \).

The reversion rate \( \eta(\epsilon_{\text{RT}}) \) is an increasing function of drug efficacy \( \epsilon_{\text{RT}} \). In the absence of drug therapy, we assume there are no infected cells going back to the uninfected class, i.e., \( \eta(0) = 0 \). As the limit case, when RT inhibitors are 100% effective (\( \epsilon_{\text{RT}} \to 1 \)), \( \eta(\epsilon_{\text{RT}}) \) should be very large. We shall discuss the functional form of \( \eta(\epsilon_{\text{RT}}) \) more in the simulation section. Our analytical results are obtained for a general reversion rate function.
Incorporating these drugs into the equations for $T$, $T^*$, and $V$ in model (2.1), we have
\begin{equation}
\frac{d}{dt} T(t) = s - dT - kV_I T + \int_0^\infty \eta(\epsilon_{RT}) T_{\text{pre}RT}(a,t) da,
\end{equation}
\begin{equation}
\frac{\partial}{\partial t} T^*(a,t) + \frac{\partial}{\partial a} T^*(a,t) = -\delta(a) T^*(a,t) - \eta(\epsilon_{RT}) T_{\text{pre}RT}(a,t) da,
\end{equation}
\begin{equation}
\frac{d}{dt} V_I(t) = \int_0^\infty (1 - \epsilon_{PI}) p(a) T_{\text{post}RT}(a,t) da - cV_I,
\end{equation}
\begin{equation}
\frac{d}{dt} V_{NI}(t) = \int_0^\infty \epsilon_{PI} p(a) T_{\text{post}RT}(a,t) da - cV_{NI},
\end{equation}
\begin{equation*}
T^*(0, t) = kV_I T.
\end{equation*}
Notice that the variable $V_{NI}$ does not appear in equations for other variables. Thus, we can ignore the $V_{NI}$ equation when studying the dynamics of infection. Using the relation (2.3), we have the following system:
\begin{equation}
\frac{d}{dt} T(t) = s - dT - kV_I T + \int_0^\infty \eta(\epsilon_{RT}) \beta(a) T^*(a,t) da,
\end{equation}
\begin{equation}
\frac{\partial}{\partial t} T^*(a,t) + \frac{\partial}{\partial a} T^*(a,t) = -\delta(a) T^*(a,t) - \eta(\epsilon_{RT}) \beta(a) T^*(a,t),
\end{equation}
\begin{equation}
\frac{d}{dt} V_I(t) = \int_0^\infty (1 - \epsilon_{PI})(1 - \beta(a)) p(a) T^*(a,t) da - cV_I,
\end{equation}
\begin{equation*}
T^*(0, t) = kV_I T.
\end{equation*}
In our analysis, we allow the viral production rate $p(a)$ to be an arbitrary function that is bounded (e.g., it does not have to be a monotone function). $\delta(a)$ is also assumed to be a bounded function.

Since we are interested in the effect of combination therapy on virus dynamics, we assume that the patients are initially at steady state and the combination of drugs is administered at time 0. We choose the initial conditions to be $T(0) = T_0$, $V_I(0) = V_{I0}$, $V_{NI}(0) = 0$, and $T^*(a, 0) = T_0^*(a)$, where $T_0$ and $V_{I0}$ are the steady state levels of target cells and infectious virions, respectively. $T_0^*(a)$ is the age distribution of infected cells at the initial time $t = 0$, and $\int_0^\infty T_0^*(a) da$ represents the steady state level of infected cells before the onset of drug therapy.

System (2.5) can be reformulated as a system of Volterra integral equations. To simplify expressions, we introduce the following notations:
\begin{equation}
K_0(a) = e^{-\int_0^a (\delta(s) + \eta(\epsilon_{RT}) \beta(s)) ds}, \quad K_1(a) = \eta(\epsilon_{RT}) \beta(a) K_0(a),
\end{equation}
\begin{equation}
K_2(a) = (1 - \epsilon_{PI})(1 - \beta(a)) p(a) K_0(a), \quad K_i = \int_0^\infty K_i(a) da, \quad i = 1, 2.
\end{equation}
$K_0(a)$ is the probability of an infected cell remaining infected at age $a$, hereafter the age-specific survival probability of an infected cell. $K_2(a)$ is the product of the age-specific survival probability of an infected cell and the rate at which infectious virus particles are produced by an infected cell of age $a$. Thus, the integral of $K_2(a)$ over all ages, i.e., $K_2 = \int_0^\infty (1 - \epsilon_{PI})(1 - \beta(a)) p(a) K_0(a) da$, gives the total number of infectious virus particles produced by one infected cell over its lifespan. For convenience, we call $K_2$ the infectious virus burst size.
For mathematical convenience, we introduce a new variable, $B(t)$, to describe the rate at which an uninfected T cell becomes infected at time $t$,

$$B(t) = kV_I(t)T(t).$$

Integrating the $T^*$ equation in system (2.5) along the characteristic lines, $t - a = \text{constant}$, we get the following formula:

$$T^*(a, t) = \begin{cases} B(t - a)K_0(a) \over T_0^*(a - t)K_0(a - t) & \text{for } a < t, \\ T_0^*(a - t)K_0(a - t) & \text{for } a \geq t. \end{cases}$$

Substituting (2.8) into the $T$ and $V_I$ equations in (2.5),

$$\frac{d}{dt}T(t) = s - dT - B(t) + \int_0^t K_1(a)B(t - a)da + \hat{F}_1(t),$$

$$\frac{d}{dt}V_I(t) = \int_0^t K_2(a)B(t - a)da - cV_I + \hat{F}_2(t),$$

where

$$\hat{F}_1(t) = \int_t^\infty \eta(\epsilon_{RT})\beta(a)T_0^*(a - t)K_0(a - t)da,$$

$$\hat{F}_2(t) = \int_t^\infty (1 - \epsilon_{PT})(1 - \beta(a))p(a)T_0^*(a - t)K_0(a - t)da.$$

Clearly, $\hat{F}_i(t) \to 0$ as $t \to \infty$, $i = 1, 2$. Integrating the $T$ equation in (2.9) and changing the order of integration, we have

$$T(t) = T_0e^{-dt} + \int_0^t e^{-d(t-u)} \left[ s - B(u) + \int_0^u B(u - \tau)K_1(\tau)d\tau + \hat{F}_1(u) \right] du$$

$$= \int_0^t \left[ e^{-d(t-u)}(s - B(u)) + B(u)H_1(t - u) \right] du + F_1(t),$$

where

$$H_1(t) = e^{-dt} \int_0^t e^{d\tau}K_1(\tau)d\tau, \quad F_1(t) = T_0e^{-dt} + \int_0^t e^{-d(t-u)}\hat{F}_1(u)du.$$  

Similarly, by integrating the $V_I$ equation in (2.9), we get

$$V_I(t) = V_{10}e^{-ct} + \int_0^t e^{-c(t-u)} \left[ \int_0^u B(u - \tau)K_2(\tau)d\tau + \hat{F}_2(u) \right] du$$

$$= \int_0^t B(u)H_2(t - u)du + F_2(t),$$

where

$$H_2(t) = e^{-ct} \int_0^t e^{c\tau}K_2(\tau)d\tau, \quad F_2(t) = T_0e^{-ct} + \int_0^t e^{-c(t-u)}\hat{F}_2(u)du.$$  

Equations (2.11) and (2.13), with $B(t)$ replaced by $kV_I(t)T(t)$, form a system of Volterra integral equations that are equivalent to the original system (2.5). Hence,
for determining the existence and uniqueness of the solutions we need only consider the following system:

\[
\begin{align*}
T(t) &= \int_0^t [e^{-d(t-u)}(s - kV_1(u)T(u)) + kV_1(u)T(u)H_1(t - u)]du + F_1(t), \\
V_1(t) &= \int_0^t kV_1(u)T(u)H_2(t - u)du + F_2(t),
\end{align*}
\]

(2.15)

where \(H_i\) and \(F_i\) \((i = 1, 2)\) are given in (2.12) and (2.14).

3. Analysis of the system (2.5). In this section, we provide analytic results on the existence of positive solutions as well as possible steady states and their stability for the system (2.5) or the equivalent system (2.15).

3.1. Existence of positive solutions. Let \(x(t) = (T(t), V_1(t))^\top\), where \(^\top\) denotes the transpose of the vector. System (2.15) can be written in the form

\[
x(t) = \int_0^t \kappa(t-u)g(x(u))du + f(t),\]

where \(f(t) = (F_1(t), F_2(t))^\top\) is a continuous function from \([0, \infty)\) to \([0, \infty)^2\), \(\kappa\) is the \(2 \times 2\) matrix with entries being locally integrable functions on \([0, \infty)\),

\[
\kappa(t) = \begin{pmatrix}
se^{-dt} & H_1(t) - e^{-dt} \\
0 & H_2(t)
\end{pmatrix},
\]

and \(g\) is defined by \(g(x) = (1,kV_1T)^\top\). Obviously, \(f \in C([0, \infty); \mathbb{R}^2)\), \(g \in C(\mathbb{R}^2, \mathbb{R}^2)\), and \(\kappa \in L^1_{loc}([0, \infty); \mathbb{R}^{2 \times 2})\). Theorem 1.1 in Gripenberg, Londen, and Staffans [17, section 12.1], shows that a continuous solution exists on a maximal interval such that the solution goes to infinity if this maximal interval is finite.

To see that all solutions will remain nonnegative for positive initial data, we use the following system (see (2.7) and (2.9)) that is also equivalent to system (2.5):

\[
\begin{align*}
\frac{d}{dt}T(t) &= s - dT(t) - B(t) + \int_0^t K_1(a)B(t-a)da + \tilde{F}_1(t), \\
\frac{d}{dt}V_1(t) &= \int_0^t K_2(a)B(t-a)da - cV_1 + \tilde{F}_2(t),
\end{align*}
\]

(3.1)

where \(\tilde{F}_i\) is given in (2.10) and \(\tilde{F}_i(t) > 0, \lim_{t \to \infty} \tilde{F}_i(t) = 0\) for \(i = 1, 2\).

Suppose that there exists a \(\bar{t} > 0\) such that \(T(\bar{t}) = 0\) and \(T(t), V_1(t) > 0\) for \(0 \leq t < \bar{t}\). Then \(B(\bar{t}) = kV_1(\bar{t})T(\bar{t}) = 0\), \(B(t) = kV_1(t)T(t) > 0\) for \(0 \leq t < \bar{t}\), and thus from the \(T\) equation in (3.1) we have \(\frac{d}{dt}T(t) = s + \int_0^t K_1(a)B(t-a)da + \tilde{F}_1(t) > 0\). Hence, \(T(t) \geq 0\) for all \(t \geq 0\). Similarly, we can show that \(V_1(t) \geq 0\) and \(B(t) \geq 0\) for all \(t \geq 0\) and for all positive initial data.

3.2. Steady states and their stability. We use the system (3.1) for our stability analysis. According to [30], any equilibrium of system (3.1), if it exists, must be a constant solution of the following limiting system:

\[
\begin{align*}
\frac{d}{dt}T(t) &= s - dT(t) - B(t) + \int_0^\infty K_1(a)B(t-a)da, \\
\frac{d}{dt}V_1(t) &= \int_0^\infty K_2(a)B(t-a)da - cV_1, \\
B(t) &= kV_1(t)T(t).
\end{align*}
\]

(3.2)
We mention that the introduction of the variable $B(t)$ is just for mathematical convenience. If we substitute $kV_I(t)T(t)$ for $B(t)$ in the first two equations of (3.2), then we will obtain the same stability results.

System (3.2) has two constant solutions, the infection-free steady state $\bar{E} = (\bar{T}, \bar{V}_I, \bar{B}) = (s/d, 0, 0)$, and the infected steady state $E^\diamond = (T^\diamond, V_I^\diamond, B^\diamond)$, where

$$
\begin{align*}
T^\diamond &= \frac{c}{k\hat{K}_2}, \\
V_I^\diamond &= \frac{sk\hat{K}_2}{kc(1 - K_1)}, \\
B^\diamond &= kT^\diamond V_I^\diamond,
\end{align*}
$$

with $K_1$ and $K_2$ given in (2.6). Notice that $K_1$ is less than 1. Thus, $V^\diamond > 0$ if and only if $sk\hat{K}_2/kc > dc$, or $R_1 > 1$, where

$$
R_1 = \frac{sk\hat{K}_2}{dc}.
$$

Clearly, the infected steady state (3.3) is feasible if and only if $R_1 > 1$. Notice that $s/d$ is the cell density in the absence of infection, and $k$ and $c$ are the cell infection and viral clearance rate, respectively. Recall that $\hat{K}_2$, the infectious virus burst size, gives the number of infectious virus particles produced by one infected cell over its lifespan. Therefore, $R_1$ gives the reproductive ratio of the virus under the impact of drugs.

We now consider the stability of steady states. Let us first consider the infection-free steady state $\bar{E}$. The following result suggests that the population sizes of virus and infected cells will go to zero if the reproductive ratio is less than 1.

**Theorem 1.** The noninfected steady state $\bar{E}$ is locally asymptotically stable (l.a.s) if $R_1 < 1$, and it is unstable if $R_1 > 1$.

**Proof.** The Jacobian matrix of (3.2) at the steady state $\bar{E}$ is

$$
J = \begin{bmatrix}
-d - \lambda & -ks/d & \hat{K}_1(\lambda) \\
0 & -c - \lambda & \hat{K}_2(\lambda) \\
0 & ks/d & -1
\end{bmatrix},
$$

where $\lambda$ is an eigenvalue and $\hat{K}_i(\lambda)$ denotes the Laplace transform of $K_i(a)$, i.e., $\hat{K}_i(\lambda) = \int_0^{\infty} K_i(a)e^{-\lambda a}da$, $i = 1, 2$. The corresponding characteristic equation is

$$
(\lambda + d) \left( \lambda + c - \frac{sk}{d} \hat{K}_2(\lambda) \right) = 0.
$$

One negative root of equation (3.5) is $\lambda = -d$, and all other roots are given by the equation

$$
\lambda + c = \frac{sk}{d} \hat{K}_2(\lambda),
$$

which can be rewritten as

$$
\frac{\lambda}{c} + 1 = \frac{R_1 \hat{K}_2(\lambda)}{\hat{K}_2}.
$$

Notice that $|\hat{K}_2(\lambda)| \leq K_2$ for all complex roots $\lambda$ with nonnegative real parts (i.e., $\Re\lambda \geq 0$). Hence, the modulus of the right-hand side of (3.7) is less than 1, provided that $R_1 < 1$. Since the modulus of the left-hand side of (3.7) is always greater than
or equal to 1 if \( \Re \lambda \geq 0 \), we conclude that all roots of (3.6) have negative real parts if \( \Re_1 < 1 \). It follows that \( \bar{E} \) is l.a.s. when \( \Re_1 < 1 \).

In the case of \( \Re_1 > 1 \), let \( \psi(\lambda) = \frac{\lambda}{c} + 1 - \Re_1 \frac{K_2(\lambda)}{K_2} \). Thus, any real roots of \( \psi(\lambda) = 0 \) are also roots of (3.6). Recognizing that \( \psi(0) = 1 - \Re_1 < 0 \) and \( \lim_{\lambda \to \infty} \psi(\lambda) = \infty \), we know that \( \psi(\lambda) = 0 \) has at least one positive root \( \lambda^* > 0 \), which is a positive eigenvalue of the characteristic equation (3.5). This shows that the infection-free steady state is unstable when \( \Re_1 > 1 \). \( \square \)

The following theorem deals with the global stability of the noninfected steady state \( \bar{E} \).

**Theorem 2.** For \( \Re_1 < 1 \), the noninfected steady state \( \bar{E} \) is a global attractor, i.e., \( \lim_{t \to \infty} (T(t), V_I(t), B(t)) = (s/d, 0, 0) \).

In order to prove Theorem 2, we need the following lemma, in which the following notations are used: \( \varphi_* = \lim \inf_{t \to \infty} \varphi(t) \), \( \varphi^* = \lim \sup_{t \to \infty} \varphi(t) \), where \( \varphi \) is a real-valued function on \([0, \infty)\).

**Lemma 1** (see [51]). Let \( \varphi : [0, \infty) \to \mathbb{R} \) be bounded and continuously differentiable. Then there exist sequences \( s_n, t_n \to \infty \) as \( n \to \infty \) such that \( \varphi(s_n) \to \varphi_* \), \( \varphi'(s_n) \to 0 \) and \( \varphi(t_n) \to \varphi^* \), \( \varphi'(t_n) \to 0 \).

**Proof of Theorem 2.** It is difficult to apply Lemma 1 to the \( T \) equation of (2.5) directly. We introduce a new variable, \( W(t) = T(t) + T^*(t) \), where \( T^*(t) \) denotes the total number of infected cells at \( t \). Notice from the \( T^* \) equation in (2.5) that \( T^* \) satisfies the equation \( \frac{dW}{dt} = kV_T - \int_0^\infty \left( \delta(a) + \eta(cE_T)\beta(a) \right) T^*(a,t) da \). Then we get \( \frac{dW}{dt} = s - d(W - T^*) - \int_0^\infty \delta(a)T^*(a,t) da = s - dW - \int_0^\infty (\delta(a) - d)T^*(a,t) da \leq s - dW. \)

The last inequality holds because of the fact that \( \delta(a) \geq d \) (i.e., the death rate of infected cells \( \delta(a) \) is equal to the natural death rate \( d \) plus an extra death rate due to the infection). By Lemma 1, we can choose a sequence \( t_n \to \infty \) such that \( W(t_n) \to W^\infty \), \( W'(t_n) \to 0 \). From \( \frac{dW}{dt} \leq s - dW \), we have \( W^\infty \leq s/d \).

Rewrite the \( V_I \) equation in (2.15) as \( V_I(t) = \int_0^t kV_I(t-u)H_2(u)du + F_2(t) \).

We use Lemma 1 to choose a sequence \( s_n \to \infty \) such that \( V_I(s_n) \to V_I^\infty \) as \( n \to \infty \), we have \( V_I^\infty \leq kV_T^\infty \int_0^\infty H_2(u)du \). Noticing that \( T^\infty \leq W^\infty \leq s/d \) and that \( \int_0^\infty H_2(u)du = K_2/c \), we get \( V_I^\infty \leq ksK_2V_T^\infty / (cd) = \Re_1 V_I^\infty \). Since \( \Re_1 < 1 \), we see that \( V_I^\infty = 0 \). Thus, \( V_I(t) \to 0 \) as \( t \to \infty \). It also follows that \( B(t) \to 0 \) since \( B(t) = kV_I(t)T(t) \) and \( T \leq W \leq s/d \). We use Lemma 1 again to choose a sequence \( s_n \to \infty \) such that \( T(s_n) \to T^\infty \) and \( T'(s_n) \to 0 \). Using the \( T \) equation in (3.2) we get \( T^\infty \geq s/d \). But \( T^\infty \leq W^\infty \leq s/d \). This shows that \( T(t) \to s/d \) as \( t \to \infty \), which finishes the proof of Theorem 2. \( \square \)

Next, we consider the stability of the infected steady state \( E^\circ \). As noted earlier, this steady state exists if and only if \( \Re_1 > 1 \). The following result suggests that the virus population will be established if the reproductive ratio is greater than 1.

**Theorem 3.** The infected steady state \( E^\circ \) is l.a.s if \( \Re_1 > 1 \).

**Proof.** The Jacobian at the steady state \( E^\circ \) is

\[
J = \begin{bmatrix}
-d - kV_I^\circ & -\lambda & -kT^\circ & K_1(\lambda) \\
0 & -c - \lambda & K_2(\lambda) & kT^\circ \\
-kV_I^\circ & kT^\circ & -1 & 0 \\
\end{bmatrix}
\]

Using the notation \( \Re_1 = skK_2/dc \), the corresponding characteristic equation can be
written as
\[
(1 - K_1)\lambda + d(R_1 - K_1) \left( \lambda + c - c\frac{K_2(\lambda)}{K_2} \right) \\
= d(R_1 - 1) \left( (\lambda + c)K_1(\lambda) - c\frac{K_2(\lambda)}{K_2} \right),
\]
(3.8)

or
\[
\left( 1 + \frac{\lambda}{c} \right) \left( A(\lambda + d) + 1 - K_1(\lambda) \right) = \frac{K_2(\lambda)}{K_2} A(\lambda + d),
\]
(3.9)

where \( A = (1 - K_1)/(d(R_1 - 1)) \).

We can exclude the possibility of a nonnegative real root of (3.9) as follows. Suppose \( \lambda \geq 0 \). Then \( K_1(\lambda) \leq K_1(0) = K_1 < 1 \). It follows that \( A > 0 \) and \( (1 + \frac{1}{c}) (A(\lambda + d) + 1 - K_1(\lambda)) > A(\lambda + d) \). Hence, (3.9) yields \( K_2(\lambda)/K_2 > 1 \). However, since \( \lambda \geq 0 \), we have \( K_2(\lambda) \leq K_2(0) = K_2 \), which leads to a contradiction. Thus, (3.9) has no nonnegative real roots.

In the next step, we will exclude the possibility that (3.9) has a complex root \( \lambda \) with a nonnegative real part. We prove this by contradiction. Suppose that \( \lambda = x_0 + iy_0 \) is a root with \( x_0 \geq 0 \) and \( y_0 > 0 \). From (3.8), we have
\[
(\lambda + d) \left( \lambda + c - \frac{K_2(\lambda)}{K_2} \right) \to 0 \quad \text{as} \quad R_1 \to 1.
\]
(3.10)

It follows from a similar argument as in Theorem 1 that \( \lambda = x_0 + iy_0 \) cannot be a root if \( x_0 > 0 \). Now we let \( x_0 = 0 \) and \( y_0 > 0 \). In this case, (3.10) has a negative root \(-d\), and all other roots are determined by the equation \( (1 + \frac{1}{c}) = \frac{K_2(\lambda)}{K_2} \) or
\[
1 + \frac{y_0}{c}i = \int_0^\infty \frac{K_2(a) \cos(ya)da}{K_2} - \int_0^\infty \frac{K_2(a) \sin(ya)da}{K_2}i.
\]
(3.11)

Comparison of the real parts of both sides yields \( \cos(ya) = 1 \). Thus, \( \sin(ya) = 0 \), which implies that (3.11) cannot hold. Therefore, (3.8) has no roots with nonnegative real parts when \( R_1 \to 1 \).

By the continuous dependence of roots of the characteristic equation on \( R_1 \), we know that the curve determined by the roots must cross the imaginary axis as \( R_1 \) decreases close to 1. That is, the characteristic equation (3.8) or (3.9) has a pure imaginary root, say, \( iy \), with \( y > 0 \). Replacing \( \lambda \) in (3.9) with \( iy \), we see that the modulus of the left-hand side of (3.9) satisfies
\[
|LHS| > \left| Ad + 1 - \int_0^\infty K_1(a) \sin(ya)da + i \left( Ay + \int_0^\infty K_1(a) \sin(ya)da \right) \right|.
\]
(3.12)

We claim that \( \int_0^\infty K_1(a) \sin(ya)da \geq 0 \). In fact, notice that \( \int_0^\infty K_1(a) \sin(ya)da = \int_0^{a_1} K_1(a) \sin(ya)da \), where \( a_1 \) is the age at which reverse transcription is complete. Notice also that \( K_1(0) = \eta(\epsilon_{RT}) \) and \( K_1'(a) = \eta(\epsilon_{RT})[\beta(a)K_0(a) + \beta(a)K_0'(a)] \leq 0 \) a.e. on \([0, \infty)\). Integrating \( \int_0^{a_1} K_1(a) \sin(ya)da \) by parts, we get
\[
\int_0^{a_1} K_1(a) \sin(ya)da = \frac{\eta(\epsilon_{RT})}{y} - \frac{1}{y} K_1(a_1) \cos(ya_1) + \frac{1}{y} \int_0^{a_1} K_1'(a) \cos(ya)da \\
\geq \frac{\eta(\epsilon_{RT})}{y} - \frac{1}{y} K_1(a_1) \cos(ya_1) + \frac{1}{y} \int_0^{a_1} K_1'(a)da \\
= \frac{1}{y} K_1(a_1)(1 - \cos(ya_1)) \geq 0.
\]
Thus, we have \( \int_0^\infty K_1(a) \sin(\gamma a) da \geq 0 \). We also observe that \( 1- \int_0^\infty K_1(a) \cos(\gamma a) da \geq 1 - K_1 > 0 \). It follows from (3.12) that \( |LHS| > A|d + iy| \). On the other hand, the modulus of the right-hand side of (3.9) satisfies \( |RHS| \leq A|d + iy| \). This leads to a contradiction. We conclude that the characteristic equation (3.9) has no roots with nonnegative real parts. Therefore, Theorem 3 is proved. \( \Box \)

4. The model with entry and protease inhibitors. Since the discovery of RT inhibitors and protease inhibitors, significant progress in drug development has been made. Recently, a new class of drugs, entry/fusion inhibitors, has been introduced \([10, 18]\). These compounds can block the fusion of the viral envelope to the target cell membrane and interfere with continued infection. They became available with the FDA approval of enfuvirtide (Fuzeon) in 2003.

In this section, we develop an age-structured model that takes into account the effects of both entry inhibitors and protease inhibitors. The model can be described by the following equations:

\[
\begin{align*}
\frac{d}{dt}T(t) &= s - dT - (1 - \epsilon_{EI})kV_I T, \\
\frac{d}{dt}T^*(a, t) + \frac{\partial}{\partial a} T^*(a, t) &= -\delta(a) T^*(a, t), \\
\frac{d}{dt}V_I(t) &= \int_0^\infty (1 - \epsilon_{PI})(1 - \beta(a))p(a) T^*(a, t) da - cV_I, \\
\frac{d}{dt}V_{NI}(t) &= \int_0^\infty \epsilon_{PI}(1 - \beta(a))p(a) T^*(a, t) da - cV_{NI}, \\
T^*(0, t) &= (1 - \epsilon_{EI})kV_I T,
\end{align*}
\]

where \( \epsilon_{EI} \) represents the efficacy of the entry inhibitor. The other parameters and variables have the same meaning as in the model (2.4). We remark that the model in \([34]\) is a special case of our model (4.1) when \( \epsilon_{EI} = \epsilon_{PI} = \beta(a) = 0 \). Our result applies to a general form of the viral production rate \( p(a) \) and the death rate \( \delta(a) \).

The existence and uniqueness of (nonnegative) solutions for the system (4.1) can be proved in a similar way as for the system (2.4). Here we present only the stability analysis. The following notations are used throughout the rest of this section:

\[
K_3(a) = e^{-\int_0^a \delta(s) ds}, \quad K_4(a) = (1 - \epsilon_{PI})(1 - \beta(a))p(a)K_3(a), \quad K_4 = \int_0^\infty K_4(a) da.
\]

The following limiting system is used to derive stability results:

\[
\begin{align*}
\frac{d}{dt}T(t) &= s - dT(t) - Y(t), \\
\frac{d}{dt}V_I(t) &= \int_0^\infty K_4(a) Y(t - a) da - cV_I, \\
Y(t) &= (1 - \epsilon_{EI})kV_I T(t),
\end{align*}
\]

where the variable \( Y(t) \) is introduced for mathematical convenience.

System (4.2) has two constant solutions (steady states): the noninfected steady state \( \bar{E} = (\bar{T}, \bar{V}_I, \bar{Y}) = (s/d, 0, 0) \), and the infected steady state \( \bar{E}^\circ = (T^\circ, V_I^\circ, Y^\circ) \), where

\[
T^\circ = \frac{c}{k(1 - \epsilon_{EI})K_4}, \quad V_I^\circ = \frac{sk(1 - \epsilon_{EI})K_4 - dc}{kc(1 - \epsilon_{EI})}, \quad Y^\circ = (1 - \epsilon_{EI})kT^\circ V_I^\circ.
\]
Clearly, \( V^\circ_r > 0 \) if and only if \( R_2 > 1 \), where \( R_2 = sk(1 - \epsilon_{EI})K_4/(dc) \) is the reproductive ratio for model (4.1). Hence, \( E^\circ \) exists if and only if \( R_2 > 1 \). The stability results are given in the following theorem. It can be proved similarly by previous arguments. Here we omit the proof due to the space limit.

**Theorem 4.** (a) The noninfected steady state \( E \) is a global attractor if \( R_2 < 1 \) and it is unstable if \( R_2 > 1 \).

(b) When \( R_2 > 1 \), the infected steady state \( E^\circ \) is l.a.s.

Results obtained in this section and in the previous section will be used in the next section to explore the impact of drug treatment on the evolution of HIV-1.

5. Influence of drug therapy on the invasion of resistant strains. In the previous sections, we have shown that a virus population can establish itself if and only if its reproductive ratio exceeds 1. Consider an environment in which the drug-sensitive strain of HIV-1 infection is at the infected steady state \( E^\circ = (T^\circ, V^\circ_r, B^\circ) \) (see (3.3)), and a small number of drug-resistant virions has been introduced into the virus population. Denote the reproductive ratio of the sensitive strain by \( R_s \), which is the same as \( R_1 \) defined in (3.4). We can rewrite the population size of uninfected cells in terms of \( R_s \), i.e., \( T^\circ = s/(dR_s) \). Assume that \( R_s \) is greater than 1.

Let \( \epsilon_{RT} \) and \( \epsilon_{PT} \) denote the efficacies of the two types of drugs for the resistant strain, respectively, and let \( \tilde{p}(a) \) denote the viral production rate of the resistant strain. We can define the corresponding \( \tilde{R}_s \) as the age-specific survival probability of T cells infected with the resistant strain (an equivalent quantity for the sensitive strain is given in (2.6)). For ease of illustration, we assume that all other parameters are the same for both strains. We derive an invasion criterion for a resistant strain by using a heuristic argument, as is done in [16]. This criterion will be applied to different scenarios of antiretroviral therapy, such as single-drug therapy (e.g., \( \epsilon_{PT} > 0 \) and \( \epsilon_{RT} = 0 \)) or combination therapy (i.e., \( \epsilon_{PT} > 0 \) and \( \epsilon_{RT} > 0 \)).

Notice that \( 1/c \) is the average lifespan of a free virus. Thus a single resistant virus can infect on average \( kT^\circ/c \) cells in its whole life. Each of these infected cells can produce a total of

\[
N_r = \int_0^\infty (1 - \tilde{\epsilon}_{PT})(1 - \beta(a))\tilde{p}(a)\hat{K}_0(a)da
\]

infectious drug-resistant virus particles during its lifespan (burst size). Thus, the reproductive ratio of the resistant strain at the resident equilibrium density \( T^\circ \) is

\[
R^\circ_r = \frac{kT^\circ}{c} \int_0^\infty (1 - \tilde{\epsilon}_{PT})(1 - \beta(a))\tilde{p}(a)\hat{K}_0(a)da,
\]

and the invasion criterion is \( R^\circ_r > 1 \). Substituting \( s/(dR_s) \) for \( T^\circ \), we obtain that the condition for the resistant strain to invade the sensitive strain is \( R_r > R_s \), where the quantity

\[
R_r = \frac{s k}{d c} \int_0^\infty (1 - \tilde{\epsilon}_{PT})(1 - \beta(a))\tilde{p}(a)\hat{K}_0(a)da
\]

represents the reproductive ratio of the resistant strain when the equilibrium density of uninfected cells is \( s/d \) (which is the value of \( T \) at the infection-free steady state).

Viral fitness is often used to describe the relative replication competence of a virus in a given environment. \( R_r \) can be regarded as a good measure of the fitness of a resistant virus. Thus the inequality \( R_r > R_s \) implies that natural selection within a host favors the virus strain that maximizes its reproductive ratio.
In order to calculate the reproductive ratio, we consider the case when the viral production rate for the resistant strain has the form given in (2.2). That is,

\[
\tilde{p}(a) = \begin{cases} 
\tilde{p}^* (1 - e^{-\theta(a-a_1)}) & \text{if } a \geq a_1, \\
0 & \text{else},
\end{cases}
\]

where \(\tilde{p}^*\) is the saturation level for production of the resistant strain. Accordingly, we choose \(\beta(a)\) to be

\[
\beta(a) = \begin{cases} 
1, & 0 \leq a < a_1, \\
0, & a \geq a_1.
\end{cases}
\]

The death rate of cells is assumed to be the same for both strains with the form

\[
\delta(a) = \begin{cases} 
\delta_0, & 0 \leq a < a_1, \\
\delta_0 + \mu, & a \geq a_1,
\end{cases}
\]

where \(\delta_0\) and \(\mu\) are positive constants with \(\delta_0\) representing a background death rate of cells and \(\mu\) representing an extra death rate for productively infected cells due to either viral cytopathicity or cell-mediated immune responses.

Drug resistance is incorporated by assuming that the efficacy of antiretroviral therapy for the resistant strain is lower than that for the drug sensitive strain by a factor between 0 and 1, i.e., \(\tilde{\epsilon}_{RT} = \sigma_{RT} \epsilon_{RT}, \tilde{\epsilon}_{PI} = \sigma_{PI} \epsilon_{PI}\). For ease of demonstration, we assume that \(\sigma_{RT} = \sigma_{PI} = \sigma\). \(\sigma = 0\) corresponds to the completely resistant strain, while \(\sigma = 1\) corresponds to the completely sensitive strain. Other strains have an intermediate value \(0 < \sigma < 1\). Many drug-resistant HIV variants display some extent of resistance-associated loss of fitness as the resistant viral strains propagate at a reduced rate when compared to sensitive strains [2]. Therefore, there is a trade-off between drug resistance and viral production rate \(\tilde{p}(a)\). We choose two types of functional forms for the cost by which the saturation level \(p^*\) is reduced in resistant strains, using the following formulas:

Type I:

\[
\tilde{p}(a) = \sigma p^* (1 - e^{-\theta(a-a_1)}),
\]

Type II:

\[
\tilde{p}(a) = e^{-\phi \frac{1}{1-\sigma} p^* (1 - e^{-\theta(a-a_1)})},
\]

where \(\phi\) is a measure for the level of cost. We provide analytic results for the Type I cost and illustrate that the qualitative properties of the two types of costs are similar. Using (5.2)–(5.5), we have the following relationship between \(R_r\) and \(R_s\) (see [13]):

\[
R_r = \frac{\sigma(1 - \sigma \epsilon_{PI}) e^{-\eta(\epsilon_{RT})(1-\sigma)a_1}}{1 - \epsilon_{PI}} R_s.
\]

We consider \(R_r = R_r(\sigma)\) as a function of \(\sigma\). A drug-resistant strain with resistance \(\sigma\) can invade the sensitive strain if \(R_r(\sigma) > R_s\). Obviously, it is not easy to draw conclusions from this condition. We first derive some analytic understanding for a simpler case in which only single-drug therapy with a protease inhibitor is considered, i.e., \(\epsilon_{PI} > 0\) and \(\epsilon_{RT} = 0\). The case of combined therapy will be explored numerically.

(a) Single-drug therapy. In this case, since \(\epsilon_{PI} > 0\) and \(\epsilon_{RT} = 0\), (5.7) simplifies to \(R_r(\sigma) = \frac{\sigma(1-\sigma \epsilon_{PI})}{1-\epsilon_{PI}} R_s\). It is easy to check that in order to have \(R_r(\sigma) \geq R_s\) for some \(\sigma \in (0,1)\) it is necessary that \(\epsilon_{PI} > \frac{1}{2}\). In fact, there exists a maximum
Fig. 1. Plots of the reproductive ratio $R_r$ of a resistant strain vs. the resistance $\sigma$ for different treatment efficacy $\epsilon_{PI}$ ($\epsilon_{RT}$ is chosen to be 0). In (a) and (b), it is shown that $R_r < R_s$ for all $\sigma < 1$. Therefore, no resistant strains can invade. In (c) and (d), resistant strains with resistance $\sigma$ in $(\sigma_{min}, 1)$ can invade. The optimal resistance is $\sigma_{opt}$ at which $R_r$ reaches its maximum $R_{r_{max}}$.

level of resistance (corresponding to the smallest value of $\sigma$), $\sigma_{min} = \frac{1 - \epsilon_{PI}}{\epsilon_{PI}} < 1$, such that $R_r(\sigma) > R_s$ if and only if $\sigma_{min} < \sigma < 1$ (see Figure 1). Clearly, if $\epsilon_{PI} < \frac{1}{2}$, then $\sigma_{min} > 1$, and hence $R_r < R_s$ for all $\sigma$. This indicates that when the drug efficacy is very low, the sensitive strain is favored. The intuitive reason for this is that if the cost of resistance is high, one would not expect resistance when there is little selection pressure from the drugs. Other nonresistant strains would outcompete it under these conditions. Resistant strains can increase in frequency only when the selection pressure (drug efficacy) is high.

We can also determine an optimal resistance, $\sigma_{opt}$, which maximizes the reproductive ratio. In fact, we can easily check that $R_r(\sigma)$ has only one critical point in the interval $(\sigma_{min}, 1)$, $\sigma = \frac{1}{2 \epsilon_{PI}}$, at which $\frac{dR_r(\sigma)}{d\sigma} = 0$ (see Figure 1). Hence, $\sigma_{opt} = 1/(2\epsilon_{PI})$.

We summarize the following results for the case of single-drug therapy. Recall that a resistant strain with resistance $\sigma$ can invade the sensitive strain if and only if $R_r(\sigma) > R_s$.

(i) There exists a threshold drug efficacy $\epsilon_{PI}^*$ ($\epsilon_{PI}^* = 1/2$ for Type I cost) below which no resistant strains can invade (see Figure 1(a)–(b)). Analytically, this is due to the fact that $\sigma_{min} \geq 1$ when $\epsilon_{PI} < \epsilon_{PI}^*$. Hence, $R_r(\sigma) < R_s$ for all $\sigma < 1$.

(ii) When the drug efficacy is above the threshold $\epsilon_{PI}$, there is a range of resistance levels for which the resistant strains are able to invade. This is because, analytically, $\sigma_{min} < 1$ when $\epsilon_{PI} > \epsilon_{PI}^*$, and $R_r(\sigma) > R_s$ for all $\sigma$ in $(\sigma_{min}, 1)$.

(iii) When $\sigma_{min} < 1$, the range of invasion strains, $(\sigma_{min}, 1)$, increases with the drug efficacy $\epsilon_{PI}$. The optimal resistance, $\sigma_{opt}$, decreases with the drug efficacy $\epsilon_{PI}$ (a more resistant strain corresponds to a smaller $\sigma$ value; see Figure 1(c)–(d)).
Fig. 2. Plots of the reproductive ratio $R_r$ vs. resistance $\sigma$ for $\epsilon_{RT} = 0.1$ (solid), $\epsilon_{RT} = 0.3$ (long dashed), $\epsilon_{RT} = 0.5$ (short dashed). The value of $\epsilon_{PI}$ is fixed at $\epsilon_{PI} = 0.6$ for which invasion is possible in the absence of the an RT drug (i.e., if $\epsilon_{RT} = 0$). For each given $\epsilon_{RT}$, the values of $\sigma$ for which $R_r(\sigma) > R_s$ give the range for resistance invasion, which is the range between the two intersection points of the $R_r$ curve and the $R_s$ horizontal line.

This increasing property is also clear from the formulas $\sigma_{\text{min}} = (1 - \epsilon_{PI})/\epsilon_{PI}$ and $\sigma_{\text{opt}} = 1/(2\epsilon_{PI})$.

(iv) As the drug efficacy increases, the optimal viral fitness, $R_r(\sigma_{\text{opt}})$, decreases (see Figure 1(c)-(d)).

(b) Combination therapy. We now consider the case of combination therapy, i.e., $\epsilon_{PI} > 0$ and $\epsilon_{RT} > 0$. Again, we consider $R_r = R_r(\sigma)$ in (5.7) as a function of $\sigma$. Then $R_r(\sigma) > R_s$ if and only if $\sigma$ satisfies the inequality

\[
\frac{\sigma(1 - \sigma\epsilon_{PI})e^{-\eta(\epsilon_{RT})(1-\sigma)\alpha_1}}{1 - \epsilon_{PI}} > 1.
\]

(5.8)

To explore the role of $\epsilon_{RT}$, we fix $\epsilon_{PI}$ (e.g., $\epsilon_{PI} = 0.6$ in Figure 2). Because the numerical simulations appear qualitatively similar for different increasing reversion rate functions, we choose $\eta(\epsilon_{RT}) = \epsilon_{RT}$ for simplicity here. We will discuss the selection of the function $\eta(\epsilon_{RT})$ in the next section. Equation (5.8) cannot be solved analytically for $\sigma$. However, plots of $R_r(\sigma)$ for different values of $\epsilon_{RT}$ suggest that, as $\epsilon_{RT}$ increases, the range for $R_r(\sigma) > R_s$ also increases (see Figure 2). Figure 3 illustrates the joint effect of $\epsilon_{RT}$ and $\epsilon_{PI}$ on the reproductive ratios $R_s$ and $R_r$. From the contour plot (see Figure 3(c)), we see that when the drug efficacy is low (the region in the lower-left corner in which $R_s > R_r > 1$) the resistant strain cannot invade. Neither strain can survive when the drug efficacy is high (the top-right region in which $R_s < 1$ and $R_r < 1$). In the middle region, the invasion of resistant strains is possible as $R_r > R_s$. 
Figure 3. Plots of the reproductive ratios $R_r$ and $R_s$ as functions of $\epsilon_{RT}$ and $\epsilon_{PI}$. Three surfaces are plotted in (a): $R_r(\epsilon_{RT}, \epsilon_{PI})$ (the top surface near the origin), $R_s(\epsilon_{RT}, \epsilon_{PI})$ (middle surface), and the constant 1 (the bottom surface). The intersection of the top two surfaces is the curve on which $R_r = R_s$. In (b), two surfaces, $R_s(\epsilon_{RT}, \epsilon_{PI})$ and the constant 1, are plotted to show the curve on which $R_s = 1$. (c) is a contour plot of the surfaces $R_r(\epsilon_{RT}, \epsilon_{PI})$ and $R_s(\epsilon_{RT}, \epsilon_{PI})$.

Figure 4 shows that when the Type II cost is used, the qualitative property of the reproductive ratio $R_r$ as a function of $\sigma$ is very similar to that when the Type I cost is used. For example, the function $R_r(\sigma)$ admits a unique $\sigma_{\text{min}}$ and a unique $\sigma_{\text{opt}}$ for sufficiently small values of $\phi$.

6. Numerical results. In this section, we provide numerical simulations to confirm and/or extend our analytical results. Backward Euler and the linearized finite difference method are used to discretize the ODE and PDE, respectively, and the integral is evaluated using Simpson’s rule. For all simulations, we choose the viral production rate $p(a)$ as (2.2) and $\beta(a)$ as (5.3) with $a_1 = 0.25$ days [24]. The death rate of infected cells $\delta(a)$ is assumed to be constant $\delta = 1$ day$^{-1}$ [29], and the virion clearance rate is set to our best estimate $c = 23$ day$^{-1}$ [45]. The other model parameters are chosen as follows [8]: $s = 10^4$ ml$^{-1}$ day$^{-1}$, $d = 0.01$ day$^{-1}$, $k = 2.4 \times 10^{-8}$ ml day$^{-1}$, and the burst size is $N = 2500$.

The reversion rate function, $\eta(\epsilon_{RT})$, remains to be determined. We know $\eta(0) = 0$, and when $\epsilon_{RT} \to 1$, $\eta(\epsilon_{RT})$ should be sufficiently large such that all the preRT cells will revert back to the uninfected class. In our simulation, we assume the reversion rate function takes the following form: $\eta(\epsilon_{RT}) = -\rho \ln(1 - \epsilon_{RT})$, where the constant
**Fig. 4.** Plots of the reproductive ratio $R_r$ vs. resistance $\sigma$ when Type II cost is considered. The value of $\phi$ measures the cost of resistance. Invasion is possible for $\sigma$ in the range between the two intersection points at which $R_r = R_s$. It also shows that invasion is impossible if the cost is too high (e.g., $\phi = 2.5$).

$\rho$ controls the steepness of the function. From the standard model in which there are only short-lived infected cells (see [44]), the viral level will be theoretically suppressed to be below the limit of viral detection ($50$ RNA copies ml$^{-1}$ in the blood) in $10.2$ days if RT inhibitors are assumed to be $100\%$ effective (we assume the same parameters as above and choose the initial viral load to be $6.7038 \times 10^5$ ml$^{-1}$). In our model (2.4), under the same initial conditions and parameters, if we choose $\rho = 2$ day$^{-1}$, then the viral load can reach the same limit in $10.2$ days when the drug efficacy of RT inhibitors is very close to $1$. Therefore, we will use the value $\rho = 2$ day$^{-1}$ in our simulation to study the RT inhibitor’s effects on the dynamics of viral load. The abilities of RT inhibitors with different $\rho$ to suppress the viral load will be discussed later.

Figures 5 and 6 show numerical simulations of the first model (2.4) and the second model (4.1), respectively. For the calculations underlying Figure 5, the maximum age of infected cells $a_{max}$ is chosen to be $10$ days [34]. In (2.2), we choose $p^* = 6.4201 \times 10^3$ and $\theta = 1$ to guarantee the burst size is $2500$ [8]. To see the influence of antiretroviral drug therapy on the viral dynamics, we choose the initial conditions to be the steady states of the standard model [44] in the absence of drug treatment. We use $T(0) = 10^6$ ml$^{-1}$ [42] and $V(0) = 10^{-6}$ ml$^{-1}$ [50] in the standard model to get the following steady state values: $T = 3.8333 \times 10^5$ ml$^{-1}$, $T^* = 6.1675 \times 10^3$ ml$^{-1}$, $V = 6.7038 \times 10^5$ ml$^{-1}$, which are used as the initial values of our models (2.4) and (4.1). The value for the efficacy of the protease inhibitor is fixed at $\epsilon_{PI} = 0.50$. Figures 5(a)–(b) and (c)–(d) are for different values of $\epsilon_{RT}$ that increase from $\epsilon_{RT} = 0.2$ (Figure 5(a)–(b)) to $\epsilon_{RT} = 0.5$ (Figure 5(c)–(d)). We observe that, when $\epsilon_{RT}$ is increased, the infection

---

$^3$In reality, the time to reach this limit is much longer, probably due to the existence of long-lived infected cells and latently infected cells [40, 41].
Fig. 5. Simulation of model (2.4) with $\epsilon_{RT} = 0.50$. The upper panel: $\epsilon_{RT} = 0.20$; the lower panel: $\epsilon_{RT} = 0.50$. The other parameters for each panel are the same: $s = 10^4 \text{ ml}^{-1} \text{ day}^{-1}$, $d = 0.01 \text{ day}^{-1}$, $c = 23 \text{ day}^{-1}$, $k = 2.4 \times 10^{-8} \text{ ml day}^{-1}$, $\theta = 1$, $T_0 = 3.8333 \times 10^5 \text{ ml}^{-1}$, $V_{I0} = 6.7038 \times 10^5 \text{ ml}^{-1}$, $V_{NI0} = 0$, $T_{*0} = 6.1675 \times 10^3 \text{ ml}^{-1}$ (see text for description). The reproductive numbers of the upper and lower panel are 1.1666 and 0.9223, respectively. The upper panel shows that the virus population stabilizes at a steady state and the T cell count remains at 800 $\mu$l$^{-1}$, and the lower panel shows that the virus dies out and the T cell count reaches 1000 $\mu$l$^{-1}$.

level at which the system stabilizes is decreased as expected. When $\epsilon_{RT}$ is greater than a threshold value ($\epsilon_{RT} = 0.41$; see also Figure 8(c)), the virus population will die out. Figure 6 shows a similar qualitative behavior of the viral load, although the efficacy of entry inhibitors has a different threshold value, $\epsilon_{EI} = 0.23$ (Figure 8(c)). The virus population persists when $\epsilon_{EI} < 0.23$ and dies out when $\epsilon_{EI} > 0.23$. This is consistent with our analytic results, as the calculation of the reproductive ratio for this set of parameters shows that $R_2 > 1$ when $\epsilon_{EI} < 0.23$ and $R_2 < 1$ when $\epsilon_{EI} > 0.23$. The different behaviors of the models shown in Figures 5 and 6 indicate that the entry inhibitor appears more effective than the RT inhibitor under given conditions. However, this comparison of effectiveness depends heavily on the choice of parameter $\rho$. If $\rho$ is increased to 5, then the RT inhibitors can suppress viral load more effectively than entry inhibitors (see more discussion in Figure 8).

Figure 7 demonstrates how the viral load can be affected by the virion production rate $p(a)$. Each drug efficacy has a fixed value: $\epsilon_{EI} = 0.20$, $\epsilon_{PI} = 0.40$. We compare
Fig. 6. Simulation of model (4.1) with $\epsilon_{PI} = 0.50$. The upper panel: $\epsilon_{EI} = 0.20$; the lower panel: $\epsilon_{EI} = 0.50$. The other parameters are the same as those in Figure 5. The reproductive numbers of the upper and lower panel are 1.0435 and 0.6522, respectively. The upper panel shows that the virus population stabilizes at a lower steady state than in Figure 5(b) (the graphs do not show this clearly, but the numerical values show the difference) and the uninfected T cell concentration remains more than $900 \mu l^{-1}$. The lower panel shows that the virus dies out and the T cell count reaches $1000 \mu l^{-1}$. This implies that the entry inhibitor appears more effective than the RT inhibitor in the given conditions.

Comparing Figure 6(a)–(b) with Figure 7(a)–(b), we observe that when the drug treatment becomes more effective ($\epsilon_{PI}$ increases from 0.4 to 0.5, $\epsilon_{EI} = 0.20$), the amplitude of the viral peak and the steady state viral load are decreased. However, it takes longer for the viral load to reach its peak level when the drug efficacy is higher. A possible explanation for this phenomenon is the following. Because a more effective
drug treatment (assuming that it is not potent enough to eliminate the virus) can suppress the virus more substantially, the nadir that the viral load can reach is much lower than when the treatment is more effective. Thus the time for the viral load to reach its peak level is prolonged.

In Figure 8, we compare the effects of two combination therapies on reducing the viral load. With the choice of $p(a)$ and $\beta(a)$ given in (2.2) and (5.3), we have the following reproductive numbers:

$$R_1 = e^{-a_1 \eta(\epsilon RT)} M_0, \quad R_2 = (1 - \epsilon EI) M_0,$$

where $M_0 = \frac{s k \theta}{c d (\delta + \theta)} (1 - \epsilon EI) p^x e^{-\delta a_1}$. Let $V^{(1)}_I$ and $V^{(2)}_I$ denote the viral steady states of models 1 and 2, respectively. Then $V^{(1)}_I = \frac{d(R_1 - 1)}{K (1 - K I)}, \quad V^{(2)}_I = \frac{d(R_2 - 1)}{K (1 - \epsilon EI)}$, where $K_1 = \frac{\eta(\epsilon RT)}{s + \eta(\epsilon RT)} (1 - e^{-(\delta + \eta(\epsilon RT)) a_1})$. If we assume the reversion rate takes the...
Fig. 8. Comparison of the two combination therapies with fixed protease inhibitor drug efficacy $\epsilon_{\text{PI}} = 0.50$. The other parameters are the same as those in Figure 5. Left column: reproductive numbers $R_1$ and $R_2$ as the function of $\epsilon_{\text{RT}}$ and $\epsilon_{\text{EI}}$, respectively. If $\epsilon_{\text{EI}} > 0.23$, then $R_2 < 1$, and hence virus will die out. Right column: steady state $V_I$ of models (2.4) and (4.1) as the function of $\epsilon_{\text{RT}}$ and $\epsilon_{\text{EI}}$, respectively. The upper panel: $\rho = 1$, the threshold for $R_1 < 1$ is $\epsilon_{\text{RT}} > 0.65$; the middle panel: $\rho = 2$, the threshold for $\epsilon_{\text{RT}}$ is 0.41; the bottom panel: $\rho = 5$, the threshold for $\epsilon_{\text{RT}}$ is 0.19. For a small $\rho (\rho < 4)$, the entry inhibitors appear more effective than the RT inhibitors; for a large $\rho (\rho > 4)$, we have the contrary result.
form $\eta(\epsilon_{RT}) = -\rho \ln(1 - \epsilon_{RT})$, then $R_1$ can be simplified, and (6.1) reduces to

$$
(6.2) \quad R_1 = (1 - \epsilon_{RT})^{a_1 \rho} M_0, \quad R_2 = (1 - \epsilon_{EI}) M_0.
$$

In Figure 8(c), we let $\epsilon_{EI} = 0.50$ and plot $R_1(\epsilon_{RT})$ and $R_2(\epsilon_{EI})$ as functions of $\epsilon_{RT}$ and $\epsilon_{EI}$, respectively. We observe that there is a threshold value, $\epsilon_{RT} = 0.41$, such that $R_1 > 1$ when $\epsilon_{RT} < 0.41$ and $R_1 < 1$ when $\epsilon_{RT} > 0.41$. By comparison, the threshold value for entry inhibitors is $\epsilon_{EI} = 0.23$, which implies that the virus population will die out when $\epsilon_{EI} > 0.23$ (see also Figures 5 and 6). In Figure 8(d), $V_i^{(1)}(\epsilon_{RT})$ and $V_i^{(2)}(\epsilon_{EI})$ are plotted as functions of $\epsilon_{RT}$ and $\epsilon_{EI}$, respectively. Given the same efficacy, the value of steady state $V_i^{(2)}(\epsilon_{EI})$ is less than $V_i^{(1)}(\epsilon_{RT})$. This indicates that the entry inhibitor appears more effective in reducing the viral load than the RT inhibitor in this scenario ($\rho = 2$). In fact, more information can be obtained by looking at the slopes of $R_1(\epsilon_{RT})$ and $R_2(\epsilon_{EI})$ in Figure 8(c) since the slope characterizes the effectiveness of drug treatment in infection control when drug efficacy is increased. From (6.2), we see that $R_1$ decreases non-linearly as $\epsilon_{RT}$ increases and the decay rate is $(1 - \epsilon_{RT})^{a_1 \rho}$, while $R_2$ decreases linearly with the decay rate $(1 - \epsilon_{EI})$ as $\epsilon_{EI}$ increases. This implies that the effectiveness of RT inhibitors depends heavily upon the reversion constant $\rho$. In our simulation, we choose $\rho = 2 \text{ day}^{-1}$ and find that the entry inhibitor is more likely able to annihilate the virus population than the RT inhibitor when the efficacy is increased by the same percentage (see Figures 5, 6, and 8(c)–(d)). However, we obtain the contrary result when $\rho$ is chosen to be greater than $1/a_1$ (Figure 8(e)–(f)).

7. Concluding remarks. We have formulated two age-structured models for HIV-1 infection with drug treatments to study the influence of antiretroviral therapy on viral dynamics. We considered two types of combination therapies. One is the standard combination of RT inhibitors and protease inhibitors, and the other is a combination of an entry inhibitor with protease inhibitors. For each of these cases, we have calculated the reproductive ratio $R_i$ ($i = 1, 2$), which is shown to determine the asymptotic stability of the infection-free steady state (when $R_i < 1$) and the infected steady state (when $R_i > 1$). In simple nonage-structured models, both RT and entry inhibitors are modeled in the same way, i.e., as a factor that reduces the rate of infection. Here, by considering the details of the viral life cycle, we model these inhibitors differently and explicitly. When an RT inhibitor is administered, some infected cells may have already completed reverse transcription (postRT cells). For these cells, the RT inhibitor will have no effect. For infected cells that have not completed reverse transcription (preRT cells), the RT inhibitor will prevent completion of reverse transcription and allow, under the influence of enzymes that degrade the HIV-1, a reversion of infected cells back into an uninfected state. These features of our model are novel. An entry inhibitor, enfuvirtide, has been used in a combination of RT and protease inhibitors [1, 33]. The addition of enfuvirtide to the regimen was reported to increase the antiretroviral potency in one study [33] but not the other [1]. Thus the impact of entry inhibitor use needs further evaluation, and mathematical modeling may be able to help in this regard.

We studied the impact of combination therapy using RT and protease inhibitors on the emergence of drug-resistant HIV-1 strains. The cost of resistance was assumed to be a reduced viral production rate. We calculated the reproductive ratio for the resistant strain $R_r(\sigma)$ with a resistance level $\sigma$ and provided a criterion for the potential invasion of resistant strains, i.e., $R_r(\sigma) > R_s$, in an environment where the
wild-type strain was already established. We argue that natural selection within a host favors virions that maximize the reproductive ratio, which is consistent with earlier findings (see, for example, [16]). Consequently, we show that natural selection should favor viral strains that have an intermediate level of resistance and that the optimal resistance level ($\sigma_{opt}$) decreases with increasing drug efficacy (see Figures 1 and 2). Mathematically, increasing the values of $\epsilon_{PI}$ and $\epsilon_{RT}$ results in (1) a reduction in the reproductive ratio $R_s$ of the drug sensitive strain (see Figure 4) and hence a reduction in the equilibrium level of infection (see $T^o = s/(dcR_s)$ and $V^o = (d/k)(R_s - 1)/(1 - K_1)$ in (3.3)); and (2) a decrease in the optimal viral fitness $R_{r,max}$ of the resistant strain and a decrease in the optimal resistance $\sigma_{opt}$ (see Figures 1 and 3), and an increase in the range of resistance ($\sigma_{min} < \sigma < 1$) for which $R_r > R_s$ ($\sigma_{min}$ decreases with both $\epsilon_{PI}$ and $\epsilon_{RT}$; see Figures 1 and 2). These are strains that are able to invade a host population. On the other hand, if $\epsilon_{PI}$ and $\epsilon_{RT}$ are small such that $\sigma_{min}$ is greater than 1, then $R_r < R_s$ for all resistance $\sigma$. These strains will not be maintained in a population. It should be noted that the condition under which drug-resistant virus variants are selected in the presence of drug pressure is very complex due to various factors [14], e.g., drug potency [46], adherence to combination antiretroviral medications [14, 15], spatial heterogeneity [23], and the increasing levels of transmitted resistant virus [4]. The management of HIV-infected patients requires a better understanding of the mechanisms underlying the emergence of drug resistance. HIV resistance testing has proved helpful in clinical practice and is rapidly being incorporated into standard HIV care [47, 49].

We have also examined the effect of drug efficacy on viral dynamics by numerical simulations. As the drug efficacy increases, the steady state of viral load as well as the amplitude of the damped oscillations that characterize the approach to equilibrium decrease, which shows that an effective drug treatment will detectably lower the plasma viral load after the administration (see Figures 5, 6, and 8(c)–(d)). Moreover, the age-dependent virion production rate can also have an effect on the viral dynamics (see Figure 7).

We compared the effects of various treatments on reducing the viral population in plasma. The effectiveness of an RT inhibitor was proved to rely heavily on the reversion rate, $\eta(\epsilon_{RT})$, at which preRT cells revert back to an uninfected state because of the inhibitor. In fact, the reproductive ratio in the presence of an RT inhibitor, $R_1$, is proportional to the factor $e^{-a_1\eta(\epsilon_{RT})}$ ($a_1$ is the age at which reverse transcription is complete), while the reproductive ratio in the presence of an entry inhibitor, $R_2$, is proportional to the factor $(1 - \epsilon_{EI})$. Thus the comparison of the effectiveness of RT and entry inhibitors depends on $a_1$ and the functional form of $\eta(\epsilon_{RT})$. We chose a specific function $-\rho \ln(1 - \epsilon_{RT})$ for $\eta(\epsilon_{RT})$ in our simulations. Then the reproductive ratio $R_1$ is proportional to $(1 - \epsilon_{RT})^{a_1\rho}$. Given the same drug efficacy, an entry inhibitor appears to be more effective in reducing viral load than an RT inhibitor (see Figures 5, 6, and 8(c)–(d)) if $\rho$ is chosen to be 2 day$^{-1}$. We get the contrary result if we choose $\rho$ such that $\rho > 1/a_1$ ($\rho = 5$ day$^{-1}$ in Figure 8(e)–(f)).

Another comparison of the effectiveness of RT inhibitors is between that obtained here using our age-structured model and the previous results in the literature based on the nonage-structured “standard model” (see [37, 44]). The reproductive ratio in the presence of an RT inhibitor for the standard model is $\mathcal{R} = (1 - \epsilon_{RT})kps/(dc\delta)$ [5], where $p$ is the constant production rate and $\delta$ is the constant death rate of productively infected cells. Thus $\mathcal{R}$ decreases linearly as the drug efficacy increases, and the decay rate is $(1 - \epsilon_{RT})$. A similar comparison follows for the drug effect of RT inhibitors on suppressing the viral load between our model and the standard model. The functional
form of the reversion rate $\eta(\epsilon_{RT})$ awaits future studies. These findings might be helpful in designing treatment for the control of HIV infections. In the current model, we have not included both the wild-type and drug-resistant strains explicitly [54]. This will be done in the future work.

REFERENCES