

## Estimation and Prediction With HIV-Treatment Interruption Data

B.M. Adams<sup>a</sup>, H.T. Banks<sup>a,\*</sup>, M. Davidian<sup>a</sup>, E.S. Rosenberg<sup>b</sup>

<sup>a</sup>*Center for Research in Scientific Computation, North Carolina State University, Box 8205, Raleigh, NC 27695-8205, USA*

<sup>b</sup>*I.D. Unit—Gray 5, Massachusetts General Hospital and Harvard Medical School, Fruit Street, Boston, MA 02114, USA*

Received: 30 November 2005 / Accepted: 8 May 2006 / Published online: 9 January 2007  
© Society for Mathematical Biology 2006

**Abstract** We consider longitudinal clinical data for HIV patients undergoing treatment interruptions. We use a nonlinear dynamical mathematical model in attempts to fit individual patient data. A statistically-based censored data method is combined with inverse problem techniques to estimate dynamic parameters. The predictive capabilities of this approach are demonstrated by comparing simulations based on estimation of parameters using only half of the longitudinal observations to the full longitudinal data sets.

**Keywords** HIV models · Treatment interruptions · Censored data · Parameter estimation · Prediction

### 1. Introduction

Human Immunodeficiency Virus (HIV) is a retrovirus that infects T-helper cells of the immune system and is the causative agent for Acquired Immune Deficiency Syndrome (AIDS). HIV and AIDS are among the world's most serious public health concerns, affecting people of all demographics worldwide, with some regions impacted disproportionately. As of 2003, an estimated 38 million HIV-infected individuals are living worldwide, with approximately two-thirds in Africa, where 2.2 million people died from opportunistic infections related to AIDS in 2003 (UNAIDS 2004 Report on the Global HIV/AIDS Epidemic ([Armstrong et al., 2004](#))). Despite many successful public health and clinical interventions since the first identification of HIV-positive patients in 1981, there remains no cure and the HIV/AIDS epidemic continues to grow.

Highly Active Antiretroviral Therapy (HAART), most commonly administered in the form of drug cocktails consisting of a protease inhibitor and at least one or

---

\*Corresponding Author.  
E-mail address: htbanks@ncsu.edu (H.T. Banks).

more reverse transcriptase inhibitors, has been highly successful in suppressing HIV in many patients and therefore improving quality of life. However, contrary to dangerous popular myths, these drugs do not constitute a cure. While antiretroviral drugs are widely available in the United States and Western Europe, their cost and side effects may make their use challenging. In developing nations, UNAIDS estimates that only 7% of the infected population has access to HAART. Access to treatment for and education about this disease remain serious human rights issues around the world. Improved strategies are needed for efficient and appropriate use of drug therapy in both developed and underdeveloped countries.

Studies of the epidemiology of HIV and public health issues such as transmission (interhost dynamics) are important. Equally important to investigate are the effective use and improvement of antiretroviral drugs, which depend on understanding viral behavior within each host, including pathways of infection and effects of drugs. Understanding intrahost viral and immune system pathways depends on knowledge from various biological areas including physiology and immunology. Mathematical models, when combined with statistically-based inverse problem techniques, can aid in quantifying dynamic physiologic and immunologic processes, correlating the scientific knowledge of these processes with observed patient behavior, and predicting patient outcomes. An example of such a modeling approach is given in this paper.

It is believed that the acute and early phases of HIV infection provide crucial information about immune responses and viral dynamics. In particular, long-term viral set points and speed of progression to AIDS may possibly be understood by studying these key periods. Motivated by clinical study data from patients observed during the crucial acute infection phase and beyond, we outline here a combined mathematical and statistical inverse problem approach for modeling HIV infection. We apply the methods to clinical data and demonstrate the types of suggestions and conclusions one may draw from such an effort.

A number of patients for whom we have clinical data underwent therapy interruptions. Some of these drug holidays were unprescribed or single interruptions, while others were structured treatment interruptions (STIs) according to a study protocol. STI therapy protocols are currently being explored (not without controversy) as an alternative to continuous therapy with antiretrovirals since in addition to offering the benefit of reduced side effects, they may also serve to boost HIV-specific immune responses. We, therefore, incorporate STI protocols in our mathematical models. A good overview of the concept of STI and its applicability in various phases of HIV infection can be found in [Lori and Lisiewicz \(2001\)](#).

In previous work ([Adams, 2005](#)), we demonstrated that a differential equation model for inhost HIV-infection dynamics can describe censored clinical data obtained from patients undergoing therapy interruptions. This entailed a process of parameter identification (estimation or model fitting) in order to determine values for the dynamic parameters in the model that will best describe the data. Model fitting in this manner yields valuable estimates of dynamic rates and quantities, for example, the rate of growth of virus or infectivity contact rate, which might be used to differentiate between or explain patient behaviors.

In this paper, we explore one of the most powerful features of mathematical models—the ability to assist in making predictions or understanding biological

phenomena. We demonstrate how one could use longitudinal HIV viral load and CD4<sup>+</sup> T-cell data gathered from a particular patient over a limited observation period, in conjunction with a biologically based mathematical model, to make predictions about the patient's long-term behavior. This might include the patient's viral load or T-cell dynamics over time or a prediction of the long-term viral load set point.

In this way, our HIV model can be used to gain insight into potential clinical outcomes. For example, after calibration (i.e., parameter estimation), one could use the model to explore what would happen to a particular patient under various treatment strategies, including allowing the patient to remain completely off treatment.

## 2. Clinical data description

The data for our investigations come from a study of over 100 adults with symptomatic acute or early HIV-1 infection. These subjects were enrolled in a study based at Massachusetts General Hospital and associated regional centers and followed for varying lengths of time between 1996 and 2004. The study cohort is unique in that its members were all identified soon after initial infection, making its data particularly useful for understanding early viral dynamics and related immune responses. A principal goal of the clinical study is to assess the potential immunologic consequences of early treatment initiation, including preservation of HIV-specific CD4<sup>+</sup> T-cells, extent of latent reservoir development, and homogeneity of viral population. Clinical researchers also strive to understand the role of early immune responses in long-term viral suppression.

Clinical and demographic data were collected at the time of study enrollment and blood draw assays of CD4<sup>+</sup> T-lymphocyte count and RNA viral load performed at roughly monthly follow-up visits. Viral load was quantified with Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR) methods using the commercially available HIV-1 Roche Amplicor or Chiron Quantiplex assay, yielding measurements in viral RNA copies per milliliter (ml). The standard assay has a linear range of 400 to 750,000 copies/ml, while the ultrasensitive assay has a range of 50 to 100,000 copies/ml. The latter is typically employed when a measurement is below the 400 copies/ml limit of the standard assay, as is often the case for a patient successfully suppressing virus. Standard flow cytometry methods were employed to obtain total plasma CD4<sup>+</sup> T-lymphocyte counts per microliter ( $\mu$ l) (Kassutto et al., 2006).

Nearly all subjects in the study underwent combination therapy with three or more antiretroviral drugs, although the precise regimen varied from patient to patient as dictated by the treating physician. Fourteen of the subjects underwent structured treatment interruptions according to a study protocol, including patients with identification numbers 2, 4, 5, 6, 10, 13, and 14 for whom immune responses were assessed during interruption (Rosenberg et al., 2000). Several others simply discontinued drugs at various points.

In this paper, we focus on data from a subset of 45 patients wherein each patient has 10 or more each of CD4 and viral load measurements in the first half of their

**Table 1** Summary of data for 45 patients with 10 or more each CD4 and viral load measurements in the first half of their longitudinal data, ordered by clinical identification number

Pat. Num.	Num. Meas.		Total days	Days on/off	Periods on/off	# in 1st half	
	VL	CD4				VL	CD4
1	102	84	1527	1316/211	4/3	49	41
2	107	82	1966	902/1064	2/2	67	48
3	76	61	1943	1589/354	3/2	40	31
4	154	107	1919	1248/671	4/4	81	58
5	158	115	2061	1067/994	4/4	88	66
6	143	111	1839	923/916	4/5	85	64
7	23	22	1932	1924/8	1/1	15	14
8	34	33	1672	1668/4	1/1	24	23
9	32	32	1626	1112/514	2/3	24	24
10	73	63	1711	582/1129	1/1	57	47
12	24	19	1575	1540/35	2/1	18	15
13	64	55	914	537/377	3/3	34	28
14	136	91	1637	659/978	3/3	78	57
15	46	46	1659	932/727	1/1	23	23
18	32	30	1545	1545/0	1/0	26	25
19	21	19	1430	1416/14	1/1	19	17
20	29	27	1581	1469/112	1/2	18	17
21	38	36	1433	1412/21	1/1	26	24
23	37	36	1505	671/834	4/5	21	21
24	36	35	1436	841/595	4/3	27	26
25	83	60	1412	1255/157	4/4	46	33
26	100	72	1434	754/680	3/4	59	37
27	36	35	1379	591/788	2/2	21	21
29	34	34	1024	1017/7	1/1	25	25
30	16	13	841	837/4	1/1	12	10
31	30	30	1256	1228/28	2/2	16	16
32	33	33	1230	1209/21	1/1	21	20
33	75	52	1302	658/644	4/4	54	34
34	24	23	1174	1173/1	1/1	14	13
36	33	31	1167	1161/6	1/1	21	19
37	25	25	1146	1139/7	1/1	18	18
39	29	28	1023	910/113	3/3	20	20
41	22	21	717	688/29	2/1	16	15
42	30	30	1218	1170/48	2/1	17	17
43	28	29	1134	1060/74	1/1	15	16
45	46	28	499	418/81	2/2	14	12
46	100	55	1004	496/508	3/3	39	24
47	23	23	1002	496/506	1/2	10	10
52	20	19	708	674/34	1/1	13	12
54	25	25	878	868/10	1/1	15	14
55	14	14	806	748/58	1/1	10	10
60	19	18	746	720/26	1/1	11	10
65	18	17	755	728/27	1/1	13	12
75	16	15	549	521/28	3/3	11	10
84	16	15	461	461/0	1/0	11	10

*Note.* Includes number of measurements, duration of observation, time on versus off treatment, and the number of measurements in the early half of the time series.

longitudinal data. We denote this set by PS45 and summarize the data in Table 1, which includes the clinical identification number assigned to the patient, number of longitudinal viral load and CD4<sup>+</sup> measurements, the total length of time from presentation to last observation, total number of days on and off treatment, and the number of periods (of any length) the patient was off and on therapy. The last two columns indicate the number of available data points during the first half of each patient's time-series. The number of treatment interruptions varies drastically over the population and some patient records include an initial brief off-treatment phase after presentation before therapy commenced.

The treatment patterns and overall lengths of observation for each of the 45 patients are depicted in Fig. 1. In these schematics, thicker lines denote on-treatment periods and the thinner lines, off-treatment.

We expect that some aspects of the mathematical model later considered are more readily validated in the context of treatment schemes with a balance between time on and time off treatment. Of the 45 patients considered in this paper, 16 (those numbered 2, 4, 5, 6, 9, 10, 13, 14, 15, 23, 24, 26, 27, 33, 46, and 47) spend 30–70% time off treatment. Of these only patients 9, 15, and 47 do not spend appreciable time off treatment during the early half of their observation period.

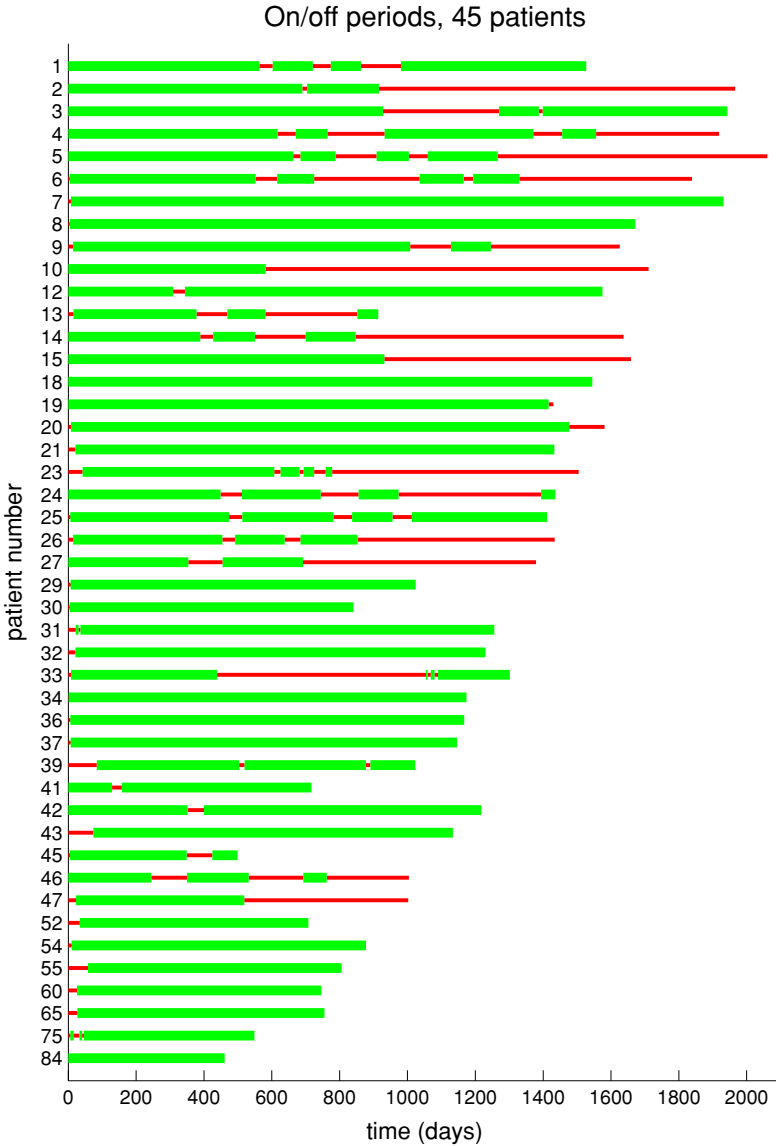
Due to the linear range limits described above, the clinical viral load assays effectively have lower and upper limits of quantification. The upper limit is typically readily handled by repeatedly diluting the sample until the resulting viral load measurement is in range and then scaling. The lower limit, or left censoring point, however, directly influences the observed data. When a data point is left-censored (below the lower limit of quantification), the only available knowledge is that the true measurement is between zero and the limit of quantification  $\bar{L}_*$  for the assay. Those at hand have two limits of quantification,  $\bar{L}_1 = 400$  copies/ml for the standard and  $\bar{L}_2 = 50$  copies/ml for the ultrasensitive assay. These are illustrated in sample data from patient 6 shown in Fig. 2, where censored data points are those appearing identically on the horizontal censoring lines  $\bar{L}_1 = 400$ ,  $\bar{L}_2 = 50$ . A statistical methodology for handling this type of censored data is described in Section 3.2.

The observation times and intervals vary substantially between patients. The sample data in Fig. 2 also reveal that observations of viral load and CD4 may not have been made at the same time points, so in general for patient number  $j$  we have CD4<sup>+</sup> T-cell data pairs  $(t_1^{ij}, y_1^{ij})$ ,  $i = 1, \dots, N_1^j$  and (potentially at different times) viral RNA data pairs  $(t_2^{ij}, y_2^{ij})$ ,  $i = 1, \dots, N_2^j$ .

### 3. HIV model and inverse problem techniques

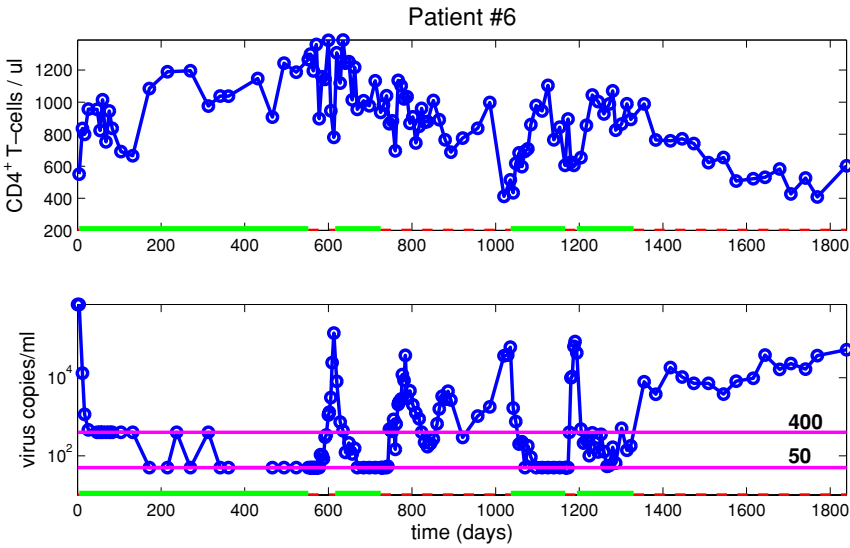
#### 3.1. Model description

Many HIV models have been considered in the literature, including those surveyed in Callaway and Perelson (2002) and Perelson and Nelson (1999). To demonstrate the potential predictive ability of such mathematical models, we employ the model developed in Adams et al. (2004), subsequently modified in Adams (2005), and



**Fig. 1** Treatment protocols and observation periods for patients in PS45. Thick (*green*) lines denote on-treatment periods whereas thin (*red*) lines denote off-treatment.

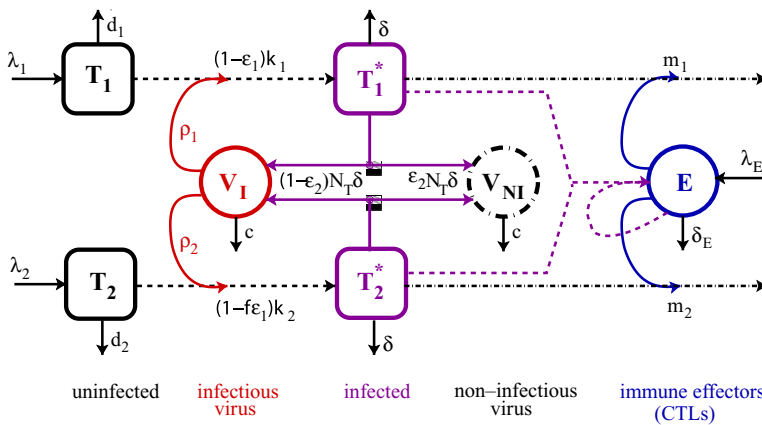
depicted in Fig. 3; other models could be readily treated in our framework. The model compartments are denoted by variables  $T_1$  (type-1 target cells, e.g.,  $CD4^+$  T-cells, cells/ $\mu$ l),  $T_2$  (type-2 target cells, e.g., macrophages, cells/ $\mu$ l),  $V_1$  (infectious free virus, RNA copies/ml),  $V_{NI}$  (noninfectious free virus, RNA copies/ml), and  $E$  (cytotoxic T-lymphocytes, cells/ $\mu$ l). A superscript asterisk (\*) denotes infected



**Fig. 2** Patient 6 CD4<sup>+</sup> T-cell and viral load data, including censor points (lines at  $\bar{L}1 = 400$ ,  $\bar{L}2 = 50$ ) for viral load, and periods of on-therapy (solid lines on axis) and periods of off-therapy (dashed line on axis).

cells. The available clinical data include total CD4<sup>+</sup> T-cell count, represented by the sum  $T_1 + T_1^*$ , and total free virus,  $V_1 + V_{NI}$ .

While the remaining compartments  $T_2$ ,  $T_2^*$ , and  $E$  were not observed in the data used in this paper, they are important for modeling and predicting long-term longitudinal data. The presence of a secondary target cell population  $T_2$  helps to satisfy a modeling requirement suggested by Callaway and Perelson (2002) in their 2002 review paper: a reasonable model of HIV infection predicts a nonzero



**Fig. 3** Schematic of compartmental HIV-infection dynamics model. Only key pathways are indicated in the schematic—for further details, see the system of differential equations (1).

steady-state viral load, even in the presence of effective drug therapy. Patients subjected to drug therapy often successfully suppress virus for a long time, potentially at undetectable levels. However, some reservoir or mechanism exists that almost invariably causes the virus to grow out to detectable levels upon removal of drug therapy. Hence, one does not expect incorporation of drug therapy in the model, at a sensible efficacy, to drive the viral load to zero, but rather reduce it considerably, perhaps below the assay limits of quantification. One way to incorporate this is shown in Fig. 3, where there are two cocirculating populations of target cells, potentially representing  $CD4^+$  T-lymphocytes ( $T_1$ ) and macrophages or other HIV-targeted cells ( $T_2$ ). The two cell populations may have different activation requirements or susceptibility to drug therapy, represented by the different rate constants, thus potentially creating a nonzero, but low viral load steady state. This is crucial for modeling our long time horizon data, where patients may remain on treatment for an extended time. The differential efficacy also enables the model to exhibit reasonable sensitivity of the viral load equilibrium to treatment efficacy. For a survey of models and discussion of which exhibit reasonable sensitivity to drug efficacy, consult [Callaway and Perelson \(2002\)](#).

The documented importance of the immune system in responding to HIV infection (and especially its apparent crucial role during structured treatment interruptions) strongly motivates the inclusion of at least one model compartment representing immune response to the pathogen. We, therefore, include a measure  $E$  of cytotoxic T-lymphocyte (CTL)  $CD8^+$  response to HIV infection. While the presently available data do not directly quantify the presence of HIV-specific CTLs, these immune responders are important for control of infected cells and may eventually be correlated to available epitope-challenge data. It is known that the immune response system is much more complicated than as represented in a single (composite) compartment denoted as CTL effectors  $E$ . Indeed, while present knowledge is incomplete, there are strong indications that a more complex modeling view of immune response involving naive and activated classes of  $CD4^+$  and HIV-specific  $CD8^+$  cells as well as memory and latent reservoir classes will be important in understanding the chronic versus acute response of the immune system to HIV-1 infection ([Lichterfeld et al., 2004](#); [Norris and Rosenberg, 2002](#)).

The corresponding compartmental ordinary differential equation (ODE) model for inhost HIV-infection dynamics is given by (1). This model is essentially one suggested in [Callaway and Perelson \(2002\)](#), but includes an immune response compartment and dynamics as suggested by [Bonhoeffer et al. \(2000\)](#). This compartment, denoted by  $E$ , represents CTLs. The adapted system of ODEs is given by

$$\dot{T}_1 = \lambda_1 - d_1 T_1 - (1 - \bar{\epsilon}_1(t)) k_1 V_I T_1 \quad (1a)$$

$$\dot{T}_2 = \lambda_2 - d_2 T_2 - (1 - f\bar{\epsilon}_1(t)) k_2 V_I T_2 \quad (1b)$$

$$\dot{T}_1^* = (1 - \bar{\epsilon}_1(t)) k_1 V_I T_1 - \delta T_1^* - m_1 E T_1^* \quad (1c)$$

$$\dot{T}_2^* = (1 - f\bar{\epsilon}_1(t)) k_2 V_I T_2 - \delta T_2^* - m_2 E T_2^* \quad (1d)$$

$$\dot{V}_I = (1 - \bar{\epsilon}_2(t)) 10^3 N_T \delta (T_1^* + T_2^*) - c V_I \quad (1e)$$

$$-(1 - \bar{\epsilon}_1(t)) 10^3 k_1 T_1 V_I - (1 - f\bar{\epsilon}_1(t)) 10^3 k_2 T_2 V_I$$

$$\dot{V}_{NI} = \bar{\epsilon}_2(t)10^3 N_T \delta(T_1^* + T_2^*) - cV_{NI} \tag{1f}$$

$$\dot{E} = \lambda_E + \frac{b_E(T_1^* + T_2^*)}{(T_1^* + T_2^*) + K_b} E - \frac{d_E(T_1^* + T_2^*)}{(T_1^* + T_2^*) + K_d} E - \delta_E E, \tag{1g}$$

together with an initial condition vector

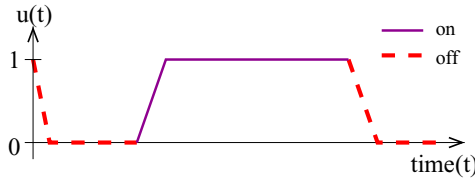
$$(T_1(0), T_1^*(0), T_2(0), T_2^*(0), V_I(0), V_{NI}(0), E(0))^T.$$

Here the factor  $10^3$  is introduced to convert between microliter and milliliter scales, preserving the units from some of the published papers.

As is common in models of HIV infection, infected cells  $T_i^*$  result from encounters between uninfected target cells  $T_i$  and infectious free virus  $V_I$  in a well-mixed environment. As noted above, this model involves two cocirculating populations of target cells, perhaps representing  $CD4^+$  T-lymphocytes ( $T_1$ ) and macrophages ( $T_2$ ). The natural infection rate  $k_i$  may differ between the two populations, which could account for suspected differences in activation rates between lymphocytes and macrophages. The treatment factor  $\bar{\epsilon}_1(t)$ , described further below, represents a reverse transcriptase inhibitor (RTI) that blocks new infections and is potentially more effective in population 1 ( $T_1, T_1^*$ ) than in population 2 ( $T_2, T_2^*$ ), where the efficacy is  $f\bar{\epsilon}_1$ , with  $f \in [0, 1]$ . The differences in infection rates and treatment efficacy help create a low, but nonzero, infected cell steady state for  $T_2^*$ , which is commensurate with the idea that macrophages may be an important source of virus after T-cell depletion. The populations of uninfected target cells  $T_1$  and  $T_2$  may have different source rates  $\lambda_i$  and natural death rates  $d_i$ .

Free virus particles are produced by both types of infected cells, which we assume produce virus at the same rate (again, this could be readily generalized to account for different productivity). In this model, virus may leave the  $V_I$  compartment due to natural death at rate  $c$  or via infecting a target cell (at rate  $k_i T_i$ ). The action of a protease inhibitor (PI), which causes infected cells to produce noninfectious virus  $V_{NI}$  is modeled by  $\bar{\epsilon}_2$ . Tracking noninfectious virus is important because the clinically-measured viral load data for patients includes total free virus (sum of infectious  $V_I$  and noninfectious  $V_{NI}$ ).

Finally, the immune effectors  $E$  (CTLs), are produced in response to the presence of infected cells and existing immune effectors. The immune response assumed here is similar to that suggested by Bonhoeffer, et al., in their 2000 paper (Bonhoeffer et al., 2000), with a Michaelis–Menten type saturation nonlinearity. (Such a saturation-type nonlinearity might be more biologically realistic in place of the product nonlinearities used elsewhere in the model, but to date our computations do not suggest a need for them at the present levels of modeling.) The infected cell-dependent death term in the immune response represents immune system impairment “at high virus load.” In Bonhoeffer et al. (2000), the authors present simulations, which suggest that a model with this immune response structure and a latently infected cell compartment can exhibit transfer between “healthy” and “unhealthy” locally stable steady states via STI, making it a good candidate for our investigation. (Indeed, further investigations (Adams et al., 2004; Banks et al., 2006) with (1) substantiate that active control through optimal or



**Fig. 4** Sample control input (treatment protocol)  $u(t)$  representing structured treatment interruption. This is a schematic in that interruption periods need not be periodic and one might assume more smooth ramp functions for the absorption and dissipation of the drug.

suboptimal STI therapies can readily effect such a transfer.) We add a source term  $\lambda_E$  to create a nonzero off-treatment steady state for  $E$ , rather than explicitly modeling immune memory. While immune effectors are not inherently present in the absence of pathogen, they persist at low levels during infection. We note that other immune response models, such as those considered by [Wodarz and Nowak \(1999\)](#) or [Nowak and Bangham \(1996\)](#) could be substituted, if desired. However, the latter does not appear to admit multiple stable off-treatment steady states.

The immune response we model is that of cytotoxic T-lymphocytes. CTLs act by lysing infected cells, causing them to explode. Thus, they remove infected cells from the system in the equations for  $T_1^*$  and  $T_2^*$ , at rates  $m_1$  and  $m_2$ , respectively. Unlike interferons, they do not directly target free virus, so there is no interaction term with the virus compartment.

In this dynamical system, the treatment factors  $\bar{\epsilon}_1(t) = \epsilon_1 u(t)$  and  $\bar{\epsilon}_2(t) = \epsilon_2 u(t)$  represent the effective treatment impact, consisting of efficacy factors  $\epsilon_1, \epsilon_2$  and a time-dependent treatment function  $0 \leq u(t) \leq 1$  representing HAART drug level, where  $u(t) = 0$  is fully off and  $u(t) = 1$ , fully on. Figure 4 depicts a sample time-varying treatment protocol representing structured therapy interruption. The relative effectiveness of RTIs is modeled by  $\epsilon_1$  and that of PIs by  $\epsilon_2$ . Since HIV treatment is nearly always administered as combination therapy, we do not consider the possibility of monotherapy, even for a limited period of time, though this could be implemented by considering separate treatment functions  $u_1(t), u_2(t)$ . In the case of model fitting, the treatment protocol  $u(t)$  is dictated by the clinical records for each patient. For a more thorough description of this model and its generalizations, the interested reader is referred to ([Adams, 2005](#); [Adams et al., 2004](#)).

In this paper,  $\bar{x}$  will denote the vector of solutions to the ODE system (1); that is,

$$\bar{x}(t) = (T_1(t), T_1^*(t), T_2(t), T_2^*(t), V_I(t), V_{NI}(t), E(t))^T, \tag{2}$$

where components 1–4 of  $\bar{x}$  are on a cells/ $\mu$ l scale, 5 and 6 (corresponding to  $V_I$  and  $V_{NI}$ ) on a copies/ml scale, and 7 on a cells/ $\mu$ l scale. The differential equation model (1) can, therefore, be represented by

$$\frac{d\bar{x}}{dt} = \bar{g}(t, \bar{x}; q),$$

with  $q$  denoting model dynamic parameters and  $\bar{g}$  the vector of derivatives. Model fits will be to the base-10 logarithm scale of these quantities ( $x = \log_{10} \bar{x}$ ) and in

general, variables with an overbar will denote an unscaled quantity and those without denote  $\log_{10}$ -transformed or scaled variables.

In solving the HIV dynamics system numerically, we use a log-transformed system. This is common practice in modeling biological quantities for several reasons. First, this resolves a problem of states (e.g., concentrations or cell counts) becoming (unrealistically) negative due to round-off error during numerical solution. Moreover, it also enables efficient numerical treatment of extremely small values or values that vary over several orders of magnitude during the simulations. From a statistical point of view, log transformation is a standard technique to render the observations more nearly normally distributed, which supports use of the least squares criterion as in Section 3.2. Using the transformation  $x = \log_{10}(\bar{x})$ , with the original system  $\dot{\bar{x}}_i = \bar{g}_i(t, \bar{x}; q)$  we obtain the system

$$\frac{dx_i}{dt} = \frac{10^{-x_i}}{\ln(10)} \bar{g}_i(t, 10^{x_i}; q), \quad i = 1, \dots, 5, 7, \tag{3}$$

which is the log-transformed analog of a reduced system for all states except  $V_{NI}$ . Given a vector of model dynamics parameters  $q$  and specified initial conditions  $\bar{x}(0)$ , we calculate numerical solutions for the reduced model using the variable-order adaptive BDF-based integrator included in Lawrence Livermore’s `LSODE` (Hindmarsh, 1983) with relative error tolerance  $10^{-9}$ .

Having obtained the model solution  $x(t)$  to (3) and, therefore,  $\bar{x}(t)$  without the  $V_{NI}$  component, we use its information on  $T_1^*(t)$  and  $T_2^*(t)$  to integrate for  $V_{NI}$  using composite 16-point Gaussian quadrature. If  $T_F$  denotes the final time (day) at which a solution is desired, then for  $t = 1, \dots, T_F$  we integrate on 1-day subintervals  $[t - 1, t]$ :

$$V_{NI}(t) = V_{NI}(t - 1)e^c + \int_{t-1}^t e^{-c(t-s)} [\bar{\epsilon}_2 10^3 N_T \delta(T_1^*(s) + T_2^*(s))] ds,$$

using 16 Gauss points to evaluate the integral in the second term. We take this approach rather than integrating the full seven-state ODE system since in the absence of PI treatment ( $\bar{\epsilon}_2(t) \equiv 0$ ), exponential decay makes it impractical to solve for  $V_{NI}$  on the logarithmic scale.

### 3.2. Inverse problem and censored data techniques

We wish to use the HIV model (1) to describe clinical data and make predictions, but it must first be “calibrated” to patient data by estimating appropriate parameters. That is, we use individual patient data (either partial or full longitudinal sets) to carry out inverse or parameter estimation problems to obtain patient-specific parameter estimates in the model. In this section, we describe the standard nonlinear least squares method for doing so and then modify it with a method for treating the censored data measurements. As noted in the data description, in performing an inverse problem, we do not have the luxury of observing the full vector of model states at each measurement time. Given an observation operator appropriate for the data, let  $\bar{x}$  denote native model solutions and  $\bar{z} = \mathcal{O}\bar{x}$

be the observed model solutions. Recall that the number of observations might vary from patient to patient, so for each of  $N_p$  patients, we have for  $j = 1, \dots, N_p$ , times  $\{t_1^{ij}, i = 1, \dots, N_1^j\}$  for CD4 measurements  $y_1(t)$  and  $\{t_2^{ij}, i = 1, \dots, N_2^j\}$  for viral RNA measurements  $y_2(t)$ . We fit the model using the base-10 logarithm of these quantities:  $x = \log_{10} \bar{x}$ ,  $z = \log_{10} \bar{z}$ ,  $y = \log_{10} \bar{y}$ . Note that throughout our discussions we use subscripts 1 and 2 for parameters, data, or indices to distinguish CD4 and viral RNA quantities, respectively. Superscripts  $j$  will be used to denote corresponding patient quantities for the  $j^{\text{th}}$  patient.

The inverse problem method will employ data from a single patient  $j$  in order to estimate one or more parameters ( $q$ ). In this case, for each fixed patient  $j$ , the goal is to fit the ODE model to his/her data by minimizing the cost criterion

$$q^{*j} = \arg \min_{q \in Q} J(q) = \sum_{s=1}^2 \frac{1}{N_s^j} \sum_{i=1}^{N_s^j} |z_s(t_s^{ij}; q) - y_s^{ij}|^2 \quad (4)$$

over an admissible parameter set  $Q \subset \mathbb{R}^p$  to obtain optimal estimates for the  $p$ -vector parameters  $q$ . This is the typical nonlinear least squares formulation, where  $J(q)$  depends through  $z$  on the solution to the nonlinear system of differential equations. There is a substantial literature on such problems ([Banks and Kunisch, 1989](#)), and, in particular, many sampling- and gradient-based methods are available to iteratively solve (4) for  $q^{*j}$  (see [Kelley, 1999](#) and the references therein). From a statistical point of view, minimizing (4) corresponds to maximum likelihood estimation of  $q$  assuming that the log-scaled measurements  $y_s^{ij}$  are normally distributed, i.e.,

$$y_s^{ij} \sim \mathcal{N}(z_s(t^i; q^0), \sigma_s^2), \quad s = 1, 2,$$

for some of the true underlying parameter values  $q^0$  and variance  $\sigma_s^2$ , where (i)  $\sigma_1^2 = \sigma_2^2$ , and both log-transformed CD4 and viral RNA measurements are (ii) independent across time (so not serially correlated) and (iii) independent of each other at each time. Assumptions (ii) and (iii) are reasonable approximations if it is assumed that the dominant source of variation in  $y_s$  values about  $z_s$  is assay error, with joint behavior at each time and across time of error-free CD4 and viral RNA values dictated primarily by the model. Assumption (i) is likely violated, as variation in CD4 and viral RNA measurements due to assay error is apt to be different. To take this into account would involve weighting each summand indexed by  $s$  in  $q^{*j}$  by estimates of  $1/\sigma_s^2$  and require some modification to the algorithm described next; for demonstration purposes, we focus on (4), recognizing that failure to weight in this manner may result in estimators for  $q$  that are less precise.

When viral load measurements are below the limit of quantification for the assay used, the observed values  $y_2^{ij}$  do not represent the true data value and come with knowledge of censoring included. We must, therefore, modify the optimization problem to include this information. This may be accomplished by employing standard methods for censored data regression analysis ([Aitkin, 1981](#); [Schneider, 1986](#)) as follows.

Unscaled measurements of viral load (second-observed component,  $\bar{y}_2^i$ ) are censored when below the limit of quantification, at either  $\bar{L}1 = 400$  or  $\bar{L}2 = 50$ . In handling the censored data, we exploit the assumption that the log-scaled measurements are normally distributed. Denote the log-scaled censoring points by  $L1 = \log_{10} \bar{L}1$ ,  $L2 = \log_{10} \bar{L}2$ . For censored data points, the available knowledge is that the observed value  $y_2^i \leq L^i$ , where  $L^i$  denotes the relevant censoring point ( $L^i = L1$  or  $L^i = L2$ ) at time  $t^i$ .

In this context we observe pairs  $(w^i, \chi^i)$ ,  $i = 1, \dots, N_2$ , where

$$w^i = \begin{cases} y_2^i & \text{if } y_2^i > L^i \\ L^i & \text{if } y_2^i \leq L^i \end{cases}$$

$$\chi^i = I_{\{y_2^i > L^i\}},$$

and  $I_A$  is the indicator function for the set  $A$ . Defining the standard normal pdf  $\phi(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\xi^2/2}$  with corresponding cdf  $\Phi(\xi) = \int_{-\infty}^{\xi} \phi(s) ds$ , we have that the viral load portion of the likelihood function for  $(q, \sigma_2)$  given the observations  $w^i$  is (see Sections 2.2.3 and 3.6 of [Kalbfleisch and Prentice, 2002](#) or Chapter 12 of [Klein and Moeschberger, 2003](#))

$$\tilde{\mathcal{L}}(q, \sigma_2) = \prod_{i=1}^{N_2} \left[ \frac{1}{\sigma_2} \phi\left(\frac{w^i - z_2^i}{\sigma_2}\right) \right]^{\chi^i} \left[ \Phi\left(\frac{w^i - z_2^i}{\sigma_2}\right) \right]^{1-\chi^i},$$

where the first term accounts for the probability of observing  $w^i$  given that it is uncensored and the second term the probability that the observation is in the interval  $(-\infty, L^i)$  when censored. This is using a truncated normal distribution for the censored measurements. The log-likelihood is

$$\begin{aligned} \mathcal{L}(q, \sigma_2) &= \sum_{i=1}^{N_2} \left( \chi^i \left[ \log \phi\left(\frac{w^i - z_2^i}{\sigma_2}\right) - \log \sigma_2 \right] + (1 - \chi^i) \left[ \log \Phi\left(\frac{w^i - z_2^i}{\sigma_2}\right) \right] \right) \\ &= \sum_{i=1}^{N_2} \left( \chi^i \left[ \log \phi\left(\frac{y_2^i - z_2^i}{\sigma_2}\right) - \log \sigma_2 \right] + (1 - \chi^i) \left[ \log \Phi\left(\frac{L^i - z_2^i}{\sigma_2}\right) \right] \right), \end{aligned} \tag{5}$$

which we maximize to estimate  $q$  and  $\sigma_2$ . This is analogous to the typical log-likelihood estimator in the absence of a limit of detection, where

$$\mathcal{L}(q, \sigma_2) = -\frac{N_2}{2} \log 2\pi - N_2 \log \sigma_2 - \sum_{i=1}^{N_2} \frac{(y_2^i - z_2^i)^2}{2\sigma_2^2}. \tag{6}$$

However, while maximizing (6) in the parameters  $q$  is equivalent to minimizing the sum of squared residuals  $\sum_{i=1}^{N_2} (y_2^i - z_2^i)^2$  (typical least squares such as (4)) and the estimation of  $q$  and  $\sigma_2$  decouple, maximizing (5) is not as simple, since a joint

estimation of  $q$  and  $\sigma_2$  must be performed. Maximizing (5) is possible with the Expectation Maximization (EM) algorithm (Dempster et al., 1977; McLachlan and Krishnan, 1997), which iteratively updates the estimates of  $q$  and  $\sigma_2$  until the maximum is achieved.

First, with the assumptions about distributions made above, let  $\xi^i = \frac{L^i - z_2^i}{\sigma_2}$  and  $\Lambda(\xi^i) = \frac{\phi(\xi^i)}{\Phi(\xi^i)}$  and use properties of a truncated normal distribution to obtain

$$E[y_2^i | y_2^i \leq L] = z_2^i - \sigma_2 \Lambda(\xi^i), \quad \text{and}$$

$$E[(y_2^i)^2 | y_2^i \leq L] = (z_2^i)^2 - 2\sigma_2 z_2^i \Lambda(\xi^i) - \sigma_2^2 \xi^i \Lambda(\xi^i) + \sigma_2^2.$$

These can be used to obtain updates  $(\tilde{y}^i, \tilde{r}^i)$  of the data points and estimates of squared residuals for the second-observed state given by

$$\begin{aligned} \tilde{y}^i &= \chi^i y_2^i + (1 - \chi^i) E[y_2^i | y_2^i \leq L^i] \\ &= \chi^i y_2^i + (1 - \chi^i) [z_2^i - \sigma_2 \Lambda(\xi^i)] \end{aligned} \tag{7}$$

and

$$\begin{aligned} \tilde{r}^i &= \chi^i E[(y_2^i - z_2^i)^2] + (1 - \chi^i) E[(y_2^i - z_2^i)^2 | y_2^i \leq L] \\ &= \chi^i (y_2^i - z_2^i)^2 + (1 - \chi^i) \left\{ E[(y_2^i)^2 | y_2^i \leq L^i] - 2z_2^i E[y_2^i | y_2^i \leq L^i] + (z_2^i)^2 \right\} \\ &= \chi^i (y_2^i - z_2^i)^2 + (1 - \chi^i) \sigma_2^2 [1 - \xi^i \Lambda(\xi^i)]. \end{aligned} \tag{8}$$

We can thus outline the EM Algorithm as follows.

**Algorithm 3.1.** Expectation Maximization (EM) Algorithm

Step 1 (Initialize) Create adjusted data  $\tilde{y}^i$  by replacing censored  $y_2^i$  values (those for which  $\chi^i = 0$ ) by  $L^i/2$ , and use ordinary least squares to estimate  $\hat{q}^{(0)}$  using both CD4 data  $y_1^i, i = 1, \dots, N_1$ , and viral RNA data  $\tilde{y}^i, i = 1, \dots, N_2$ , (which includes replaced censored values). Obtain an initial estimate for  $\sigma_2^2$  from

$$(\hat{\sigma}_2^{(0)})^2 = \frac{1}{N_2} \sum_{i=1}^{N_2} \left| \tilde{y}^i - z_2(t_2^i; \hat{q}^{(0)}) \right|^2.$$

Set  $k = 0$ .

Step 2 Define  $\hat{z}_2^{i(k)} = z_2(t^i; \hat{q}^{(k)})$  and  $\hat{\xi}^{i(k)} = \frac{L^i - \hat{z}_2^{i(k)}}{\hat{\sigma}_2^{(k)}}$  and update the data and residuals by

$$\tilde{y}^{i(k)} = \chi^i y_2^i + (1 - \chi^i) \left[ \hat{z}_2^{i(k)} - \hat{\sigma}_2^{(k)} \Lambda(\hat{\xi}^{i(k)}) \right] \tag{9}$$

$$\tilde{r}^{i(k)} = \chi^i (y_2^i - \hat{z}_2^{i(k)})^2 + (1 - \chi^i) (\hat{\sigma}_2^{(k)})^2 \left[ 1 - \hat{\xi}^{i(k)} \Lambda(\hat{\xi}^{i(k)}) \right]. \tag{10}$$

Step 3 Update the estimates to  $\hat{q}^{(k+1)}, \hat{\sigma}_2^{(k+1)}$  by performing ordinary least squares minimization in the parameters  $q$

$$\hat{q}^{(k+1)} = \arg \min_q \sum_{i=1}^{N_1} |y_1^i - z_1(t_1^i; q)|^2 + \sum_{i=1}^{N_2} |\bar{y}^{i(k)} - z_2(t_2^i; q)|^2$$

and computing

$$(\hat{\sigma}_2^{(k+1)})^2 = \frac{1}{N_2} \sum_{i=1}^{N_2} \bar{r}^{i(k)}.$$

If relative changes in  $\hat{q}$  and  $\hat{\sigma}$  are small, terminate. Otherwise set  $k = k + 1$  and then go to *Step 2*.

This iterative process yields estimates of the parameters, variance, and expected values of the data at times where censored observations were recorded. This information can then be used to compute standard errors and confidence intervals on parameter estimates. While this is not pursued here, see [Adams \(2005\)](#) for further discussions and results.

### 3.3. Validation process

One of our goals is to develop a methodology to use patient data from early infection periods with models to predict long-term set points (e.g., viral loads, CD4 counts) in individual patients. The current model requires values of 20 model parameters and seven initial conditions to carry out longitudinal simulations. While our current data sets typically contain from 4 to 5 years data for each patient and allow us to estimate all 27 of these values, we would not expect sufficient data in early disease progression in patients to allow us to estimate all model parameters and initial conditions. One approach under these conditions is to substitute for some of the parameters in the model “book” values or previously estimated population average values (in general, book values are not yet available for such models), so that those parameters are held fixed at the same values for all patients. While this approach is unrealistic in that it is certainly the case that there is biological variation in all parameters across patients, fixing parameters to which model set points are the least sensitive may yield a reasonable approximation in the case of limited data. To test this idea, we used our rich data sets to estimate all 27 parameters and initial condition values and then averaged these to obtain population averages. We then fixed 16 (12 parameters and four initial conditions—see [Table 2](#)) of these at the population averages and attempted only to estimate the remaining eight model parameters and three initial conditions. These 11 parameters/initial conditions were those to which the model had greatest semirelative sensitivity as discussed in [Adams \(2005\)](#). This is a reasonable scenario for testing model prediction capabilities in the presence of limited longitudinal data for a given patient in the early stages of therapy.

To validate the concept of using the model to predict long-term behavior, we consider the first half of each patient’s time series data in estimating parameters

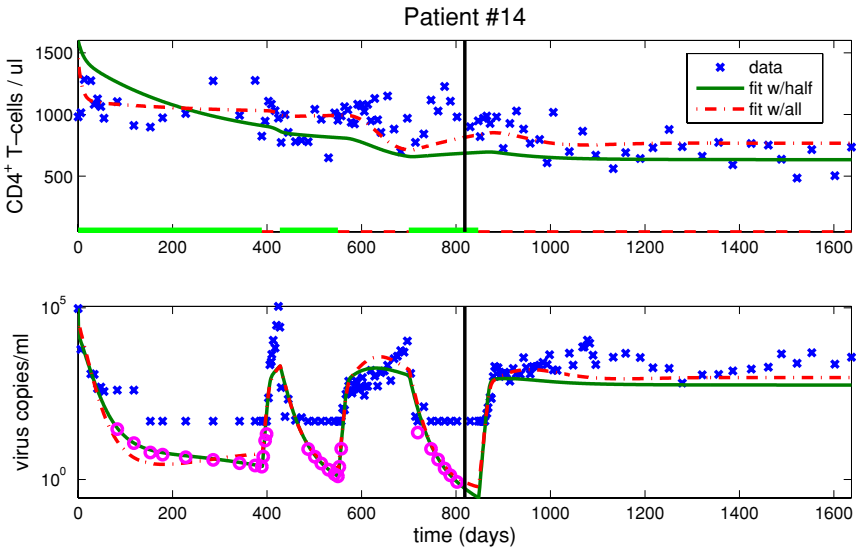
**Table 2** Average parameter values (12) and initial conditions (4) used in model fitting with half and full longitudinal data sets

$\lambda_2$	1.0099e-01	$K_d$	8.3790e-01
$d_2$	2.2109e-02	$\delta_E$	7.0299e-02
$f$	5.3915e-01	$d_E$	1.0213e-01
$k_2$	5.5290e-04	<i>initial conditions</i>	
$\delta$	1.8651e-01	$T_2^0$	1.7545e+01
$m_1$	2.4385e-02	$T_2^{*0}$	6.0955e-01
$m_2$	1.3099e-02	$V_{NI}^0$	4.9909e+03
$\lambda_E$	9.9085e-03	$E^0$	1.8834e-01
$K_b$	3.9087e-01		

and then use the calibrated model to try to predict the remaining half of the data. We employ the following process for each patient in the data set:

1. *Fix a number of model parameters (12) and initial conditions (4) at average values obtained from earlier model fitting processes.* While one could alternately use published literature values or those obtained through another process, we used values resulting from initial model fitting to each of the 45 patients. In this process, we fit the differential equation model to each patient's half time series data by estimating all 20 model dynamics parameters and seven initial conditions. This involves an initial optimization process with the hypercube sampling-based DIRECT algorithm (Finkel, 2005), followed by application of the censored data algorithm within which we employ a gradient-based optimizer. (In the results reported here, we used Matlab's LSQNONLIN which includes a finite difference approximation to residual derivatives.) We then average the estimated values across the 45 patients to obtain "typical" parameter values. Values obtained this way are shown in Table 2 and are comparable to those obtained from literature (see Adams (2005) for a summary), from using the full time series data, or from using a larger patient population.
2. *Use the first half of the patient's longitudinal data to estimate the remaining dynamic parameters and initial conditions.* We estimate eight dynamic parameters ( $\lambda_1, d_1, \epsilon_1, k_1, \epsilon_2, N_T, c, b_E$ ) and three initial conditions ( $T_1^0, T_1^{*0}, V_1^0$ ). We again apply the DIRECT algorithm, followed by the gradient-based censored data algorithm to obtain optimal parameter estimates.
3. *Evaluate how well the model, given these estimated parameters, describes the full longitudinal data set for the patient.* We simulate the trajectory over the full time span of the patient's observations, using the model parameters obtained from the first half of the data.

In addition, we also compare these model fitting results to those obtained by applying the DIRECT/censored sequence of algorithms to the full longitudinal data set. In many cases, the results obtained are similar.



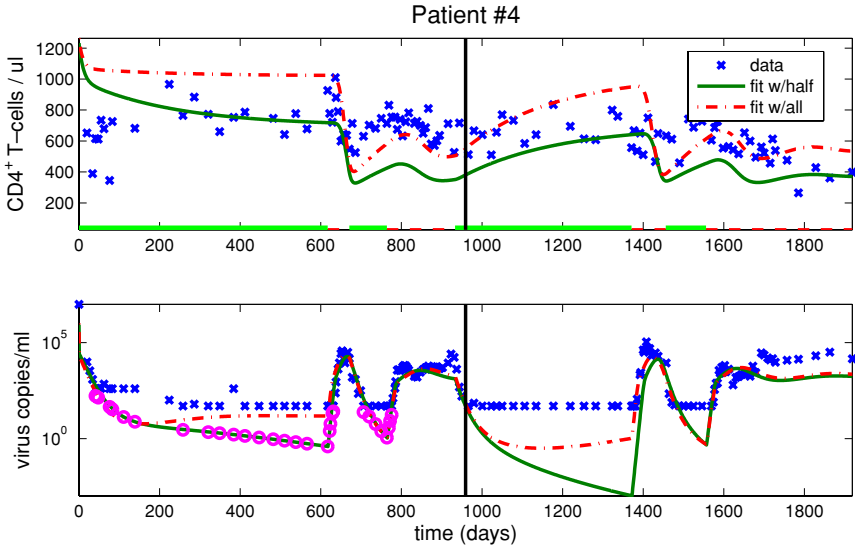
**Fig. 5** Model fit to data (“x”) for patient 14 with parameters estimated from half longitudinal data (solid line) or full dataset (dash-dot line). Circles denote predictions of censored data measurements and the vertical line delineates between the two halves of the longitudinal data.

#### 4. Model fits and sample predictions

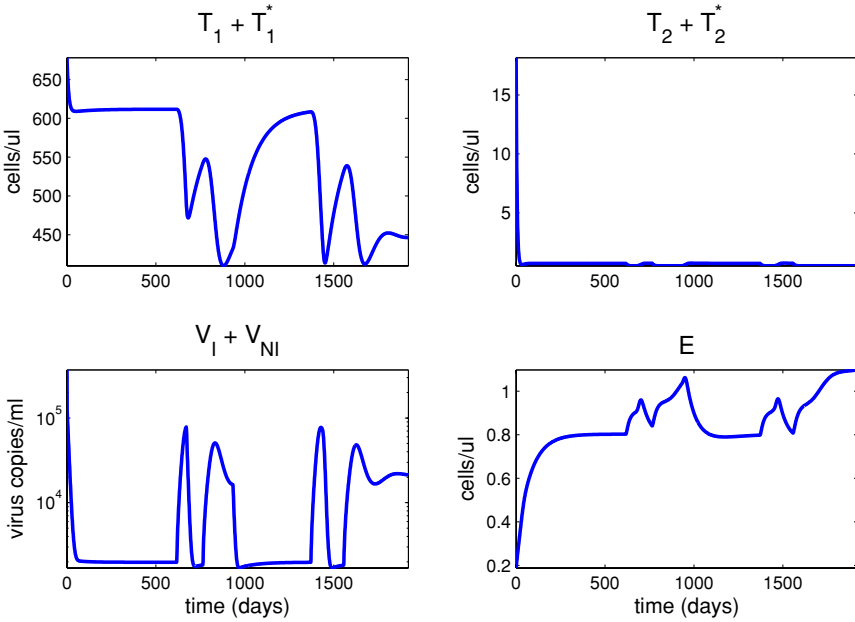
Fitting patient data using half the longitudinal data and then extrapolating over the whole time horizon often yields results similar to fitting the entire data set, supporting the model’s predictive ability. In Figs. 5 and 6 corresponding to patients 14 and 4, respectively, we see model fits obtained using half and full time series data. Both these patients undergo two treatment interruptions during the early half of their data. The fit to viral load data is nearly the same regardless of whether half or all of the data are used. The fits to T-cell data are qualitatively different, but using half or all the data both yield plausible fits to data. For patient 14, the predicted viral load during the final off-treatment phase is within 1 log of the observed data. Similar results are obtained for other patients undergoing multiple interruptions, e.g., see figures for patients 2 and 6 in the appendix of the technical report CRSC-TR05-40 (Adams et al., 2005) where fits for each of the patients in our data set are given.

While calibrated solely with total virus and total T-cell count data, the model also suggests dynamics for the other (unobserved compartments). For parameters estimated using half of the data from patient 4, Fig. 7 presents the model dynamics for target cell population 1, target cell population 2, total virus, and immune response  $E$ .

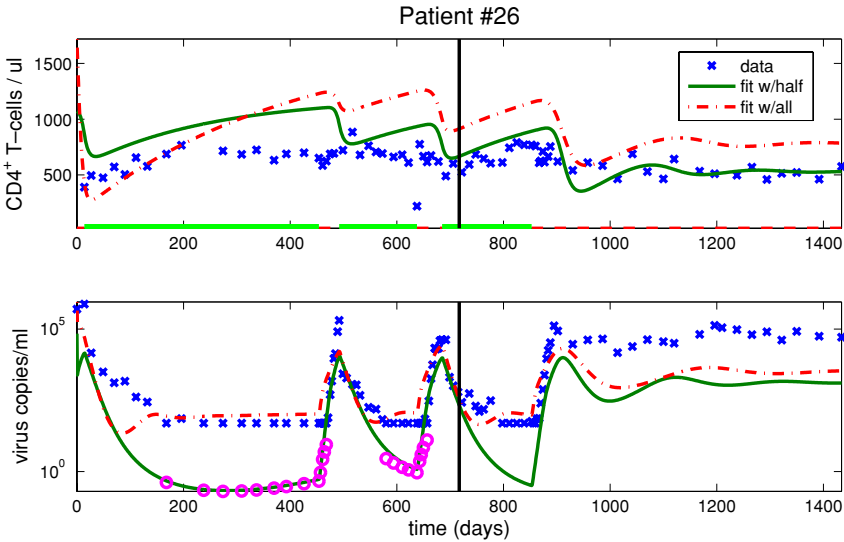
From Fig. 8, we observe that even with two interruptions wherein viral peaks are well represented by the model, the method may not yield an accurate prediction. As calibrated with only half the data in time, the model does not accurately predict the long-term off-treatment steady state exhibited by patient 26. The model fit



**Fig. 6** Model fit to data (“x”) for patient 4 with parameters estimated from half longitudinal data (solid line) or full dataset (dash-dot line). Circles denote predictions of censored data measurements and the vertical line delineates between the two halves of the longitudinal data.



**Fig. 7** Model dynamics using parameters estimated from half of the time series data from patient 4.

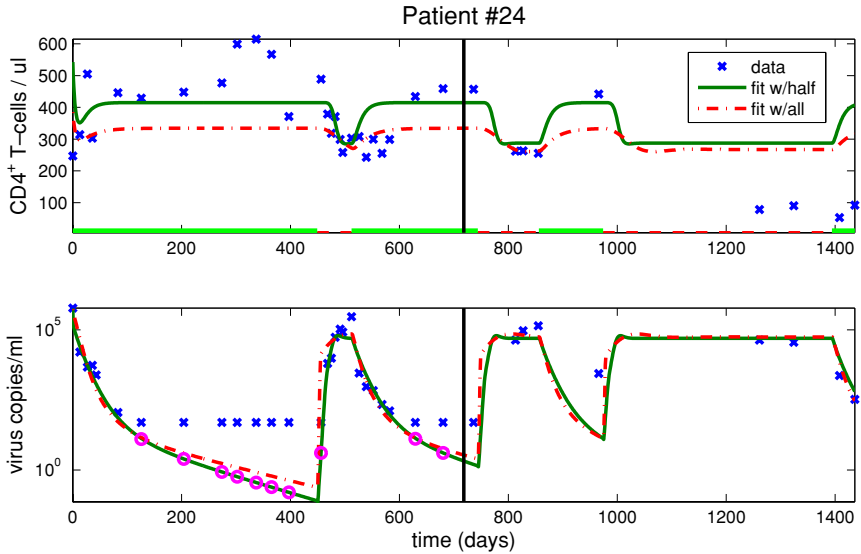


**Fig. 8** Model fit to data (“x”) for patient 26 with parameters estimated from half longitudinal data (solid line) or full dataset (dash-dot line). Circles denote predictions of censored data measurements and the vertical line delineates between the two halves of the longitudinal data.

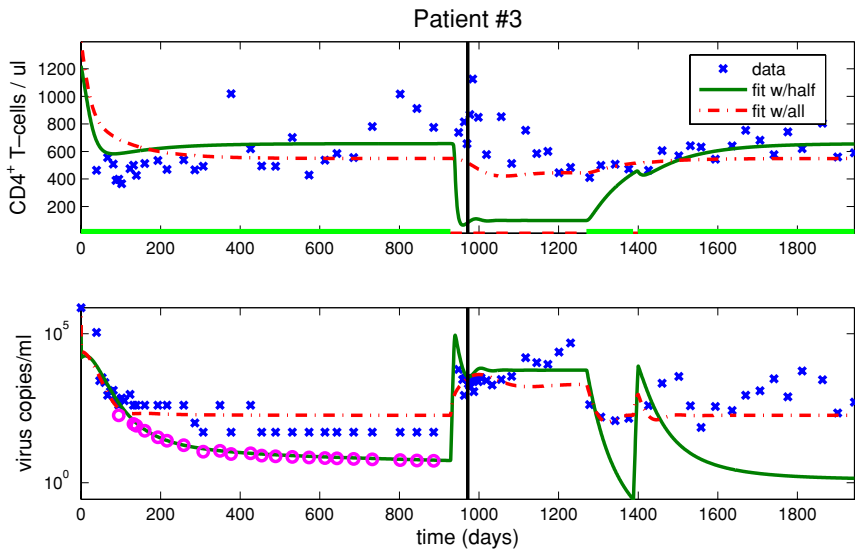
resulting from using half the data underpredicts the viral load when off treatment, which is likely related to either an underestimation of viral productivity or viral infectivity, overestimation of viral death rate, or poor modeling/estimation of immune responsiveness. In this period of treatment discontinuation (beginning about 850 days into observation), the T-cells, virus, and immune responders all interact without the intervention of drugs, i.e., naturally. The prediction may, therefore, also be due to the overly simplistic and limited immune response model considered here. A conservative estimate of the off-treatment viral load setpoint results even when using the full time series data, suggests that for this patient, fixing the other parameters at the prescribed average values made it difficult to fit the model. This method should probably be accompanied by a measure of certainty of the prediction, perhaps based on how well the early time series data has been fit. (Although it would be difficult to argue which of the two fits to the early longitudinal data is “better.”)

As shown in Fig. 9 for patient 24, it is possible to gain valuable information about a patient from even a single treatment interruption. Patients 10, 12, 13, and 25 yield similar results. Overall, in most cases, having one or two treatment interruptions yields a good prediction of long-term viral dynamics. For example, for the 11 patients with a single treatment interruption during the first half of their data, the method only severely mispredicts the remaining data for patient 27. For the remainder of patients, we see similar results when using half or all of the data.

As one might expect, if a patient does not undergo a therapy interruption during the observation period used to fit the model, it is difficult to predict a later treatment interruption. We examine this for patient 3 in Fig. 10 and similar results hold for patients 23 and 47.



**Fig. 9** Model fit to data (“x”) for patient 24 with parameters estimated from half longitudinal data (*solid line*) or full dataset (*dash-dot line*). Circles denote predictions of censored data measurements and the vertical line delineates between the two halves of the longitudinal data.



**Fig. 10** Model fit to data (“x”) for patient 3 with parameters estimated from half longitudinal data (*solid line*) or full dataset (*dash-dot line*). Circles denote predictions of censored data measurements and the vertical line delineates between the two halves of the longitudinal data.

The difference in model dynamics is reflected in the estimated parameters. For example, for patient 3, the estimates for  $N_T$  (average virus released per burst T-cell) are  $1.829e+01$  (half data) versus  $3.677e+01$  (full data), so the underprediction of viral load in this case may well be simply the result of underestimation of the rate  $N_T$  when only using half the data.

## 5. Conclusions

A number of goals have been achieved in this paper. First, we have demonstrated that we can fit a complex mathematical model of HIV infection to long-term time series clinical data for individual patients. The data includes patients who experienced treatment interruptions. The novel inverse problem method we employ incorporates a censored data algorithm. After fitting the model to data, we investigated capabilities of the model in prediction. In particular, our findings suggest that: (i) one may fit all time series and subsequently predict possible multiple stable steady states and (ii) one could use a subset of data to fit the model and then extrapolate over longer time horizons to predict viral load set points that might be most valuable in therapy decisions. We thus demonstrate that the treatment interruption data can provide crucial information for model fitting in terms of determining CD4 and viral load steady states.

While the model provides reasonable fits to most patient data, there are rather obvious areas for model improvement including additional compartments to better represent overall immune response to infection. We are currently pursuing such modeling efforts among others.

## Acknowledgements

This work was supported in part by the Joint DMS/NIGMS Initiative to Support Research in the Area of Mathematical Biology under grant 1R01GM67299-01, in part by the NCI under grant R01CA085848, in part by the NIAID under grant R37AI031789, and benefited from facilities at the Statistical and Applied Mathematical Sciences Institute, which is funded by NSF under grant DMS-0112069.

## References

- Adams, B.M., 2005. Non-parametric parameter estimation and clinical data fitting with a model of HIV Infection. PhD Thesis, NC State University, Raleigh.
- Adams, B.M., Banks, H.T., Davidian, M., Kwon, H.D., Tran, H.T., Wynne, S.N., Rosenberg, E.S., 2005. HIV dynamics: Modeling, data analysis, and optimal treatment protocols. *J. Comput. Appl. Math.* 184(1), 10–49.
- Adams, B.M., Banks, H.T., Davidian, M., Rosenberg, E.S., 2005. Model fitting and prediction with HIV treatment interruption data, Center for Research in Scientific Computation Technical Report CRSC-TR05-40, NC State University, Raleigh, October. Online: <http://www.ncsu.edu/crsc/reports>.
- Adams, B.M., Banks, H.T., Tran, H.T., Kwon, H., 2004. Dynamic multidrug therapies for HIV: Optimal and STI control approaches. *Math. Biosci. Eng.* 1(2), 223–241.

- Aitkin, M., 1981. A note on the regression analysis of censored data. *Technometrics* 23, 161–163.
- Armstrong, S., Fontaine, C., Wilson, A., 2004. 2004 Report on the Global AIDS Epidemic. UNAIDS/Joint United Nations Programme on HIV/AIDS, Geneva, Switzerland. Online: <http://www.unaids.org>.
- Banks, H.T., Kunisch, K., 1989. Estimation Techniques for Distributed Parameter Systems. Birkhauser, Boston.
- Banks, H.T., Kwon, H., Toivanen, J.A., Tran, H.T., 2006. An SDRE-based estimator approach for HIV feedback control [Technical Report CRSC-TR05-20, NC State University, Raleigh, April]. *Optim. Control Appl. Methods* 27, 93–121.
- Bonhoeffer, S., Remiszewski, M., Ortiz, G.M., Nixon, D.F., 2000. Risks and benefits of structured antiretroviral drug therapy interruptions in HIV-1 infection. *AIDS* 14, 2313–2322.
- Callaway, D.S., Perelson, A.S., 2002. HIV-1 infection and low steady state viral loads. *Bull. Math. Biol.* 64(1), 29–64.
- Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc., Ser. B* 39(1), 1–38.
- Finkel, D.E., 2005. Global optimization with the DIRECT algorithm. PhD Thesis, NC State University, Raleigh. Online: <http://www4.ncsu.edu/~definkel/research/Direct.m>.
- Hindmarsh, A.C., 1983. Scientific Computing. Chapter ODEPACK, A Systematized Collection of ODE Solvers, North-Holland, Amsterdam, pp. 55–64. Online: <http://www.llnl.gov/CASC/odepack/>.
- Kalbfleisch, J.P., Prentice, R.L., 2002. The Statistical Analysis of Failure Time Data. Wiley, New York.
- Kassutto, S., Maghsoudi, K., Johnston, M.N., Robbins, G.K., Burgett, N.C., Sax, P.E., Cohen, D., Pae, E., Davis, B., Zachary, K., Basgoz, N., D'agata, E.M.C., DeGruttola, V., Walker, B.D., Rosenberg, E.S., 2006. Longitudinal analysis of clinical markers following antiretroviral therapy initiated during acute or early HIV-1 infection. *Clin. Infect. Dis.* 42, 1024–1031.
- Kelley, C.T., 1999. Iterative methods for optimization. In: *Frontiers in Applied Mathematics FR18*. SIAM, Philadelphia.
- Klein, J.P., Moeschberger, M.L., 2003. Survival Analysis: Techniques for Censored and Truncated Data. Springer, New York.
- Lichterfeld, M., Kaufman, D.E., Yu, G., Mui, S.K., Addo, M.M., Johnston, M.N., Cohen, D., Robbins, G.K., Pae, E., Alter, G., Wurcel, A., Stone, D., Rosenberg, E.S., Walker, B.D., Altfield, M., 2004. Loss of HIV-1-specific CD8+ T-cell proliferation after acute HIV-1 infection and restoration by vaccine-induced HIV-1-specific CD4+ T-cells. *J. Exp. Med.* 200(6), 701–712.
- Lori, F., Lisziewicz, J., 2001. Structured treatment interruptions for the management of HIV infection. *J. Am. Med. Assoc.* 286(23), 2981–2987.
- McLachlan, G.J., Krishnan, T., 1997. The EM Algorithm and Extensions. Wiley, New York.
- Norris, P.J., Rosenberg, E.S., 2002. CD4+ T-helper cells and the role they play in viral control. *J. Mol. Med.* 80, 397–405.
- Nowak, M.A., Bangham, C.R.M., 1996. Population dynamics of immune responses to persistent viruses. *Science* 272, 74–79.
- Perelson, A.S., Nelson, P.W., 1999. Mathematical analysis of HIV-1 dynamics in vivo. *SIAM Rev.* 41(1), 3–44.
- Rosenberg, E.S., Altfield, M., Poon, S.H., Phillips, M.N., Wilkes, B., Eldridge, R.L., Robbins, G.K., D'Aquila, R.D., Goulder, P.J.R., Walker, B.D., 2000. Immune control of HIV-1 after early treatment of acute infection. *Nature* 407, 523–526.
- Schneider, H., 1986. Truncated and Censored Samples from Normal Populations. Marcel Dekker, New York.
- Wodarz, D., Nowak, M.A., 1999. Specific therapy regimes could lead to long-term immunological control of HIV. *Proc. Natl. Acad. Sci.* 96(25), 14464–14469.