A Dynamical Modeling Approach for Analysis of Longitudinal Clinical Trials in the Presence of Missing Endpoints

H.T. Banks*, Shuhua Hu+* and Eric Rosenberg@*
*Center for Research in Scientific Computation
North Carolina State University
Raleigh, NC 27695-8212
and
+Certara, Inc.
Cary, NC 27518
and
@Departments of Pathology and Medicine
Massachusetts General Hospital and Harvard Medical School
Boston, MA 02114

July 2, 2016

Abstract

Randomized longitudinal clinical trials are the gold standard to evaluate the effectiveness of interventions among different patient treatment groups. However, analysis of such clinical trials becomes difficult in the presence of missing data, especially in the case where the study endpoints become difficult to measure because of subject dropout rates or/and the time to discontinue the assigned interventions are different among the patient groups. Here we report on using a validated mathematical model combined with an inverse problem approach to predict the values for the missing endpoints. A small randomized HIV clinical trial where endpoints for most of patients are missing is used to demonstrate this approach.

Key words: HIV, hypothesis testing, ordinary differential equation, inverse problems

Mathematics Subject Classification: 62G10, 65L09, 92C50

1 Introduction

Even though most randomized clinical trials are carefully designed, it often becomes inevitable for participants to go off the study or discontinue the assigned intervention before study completion. This is especially true for longitudinal clinical trials conducted over a substantial period of time in which new or alternative strategies are introduced into clinical

practice which may effect continued particpation of study sbjects, resulting in a incomplete or dificult to analyze data set. The impact of conducting of an incomplete or altered clinical trial is large, especially given the expense and time involved for study subjects and investigators. However, it is often the case that information contained in the data collected for an early termination time is useful. Consequently, the question of how to efficiently use the collected data and appropriately handle missing data points has become one of the important problems in analysis of clinical trials. This is especially important in cases where the clinical trials involve a small number of participants.

To minimize the number of subjects who are eliminated from the analysis, imputation methods are often used to handle missing data. A number of ad-hoc imputation approaches have been proposed in the literature including the commonly used last-observation-carried-forward, baseline observation carried forward, and intent to treat (ITT) methods. However, these methods do not work well for those longitudinal clinical trials where dropout or discontinuation of the assigned interventions occurs early and the outcomes of interest are changing over time. They may also provide biased treatment comparisons if dropout rates or/and times to dropout are different among the intervention groups. For more information on these methods, we refer interested readers to [5, 7] and the references therein.

To partially alleviate the difficulty encountered by the traditional methods, we propose to use mathematical/dynamical modeling combined with an inverse problem approach to analyze longitudinal clinical trials in the presence of missing endpoints. A randomized trial of treatment versus no treatment in subjects with acute HIV infection is used to illustrate the proposed approach, where the model used to predict the missing endpoints was carefully validated by multiple data sets collected previously. The remainder of this paper is organized as follows: we first give a brief introduction of this clinical trial, and then we talk about how to use the proposed method to analyze this trial. Finally we conclude the paper by some remarks.

2 An HIV Randomized Clinical Trial

The randomized clinical trial used to illustrate our approach was conducted in Massachusetts General Hospital from 2009 to 2014. This trial was approved by the Massachusetts General Hospital human subject protection committee ([IRB]) In this trial, subjects identified with acute HIV infection were randomized to receive either no therapy, 12 weeks therapy, or 32 weeks therapy. The goal of this trial is to determine whether treatment initiated during acute HIV infection followed by terminal interruption results in a lower HIV viral load level and higher CD4+ T cell count than no treatment and to determine whether the length of time in treatment before discontinuation plays an important role. Below we will give detailed information on the objectives and the associated endpoints for this study as well as the data collected.

2.1 Objectives and Endpoints

As we stated earlier, there were two objectives in our study. The primary objectives were to determine whether treatment initiated during acute HIV infection is beneficial as measured by the following primary (P) and secondary (S) study endpoints: (P1) To determine whether or not there is a difference in the *CD4+ T cell count* between the group without treatment at 46-48 weeks after randomization and the group with treatment at 46-48 weeks after discontinuation of treatment; (P2) To determine whether or not there is a difference in the *viral load level* between the group without treatment at 46-48 weeks after randomization and the group with treatment at 46-48 weeks after discontinuation of treatment. The primary endpoints were determined for each study subject by taking the average of two HIV $\log_{10}$ RNA viral load measurements and the average of two log CD4+ T cell measurements determined 46-48 weeks after discontinuation of treatment in subjects randomized to receive therapy and 46-48 weeks after randomization for patients assigned to no treatment. Specifically, for patients randomized to receive 12 weeks (32 weeks) therapy, these endpoints are the average of two HIV $\log_{10}$ RNA viral load measurements and the average of two log CD4+ T cell measurements taken at 58 and 60 weeks (78 and 80 weeks) after randomization. While for subjects randomized to receive no treatment, they are the averages of two observations taken at 46 and 48 weeks after randomization.

The secondary objectives are to determine whether the duration of treatment before interruption is important, and these are detailed as follows: (S1) To determine whether or not there is a difference at 46-48 weeks after discontinuation of treatment in the *CD4+ T cell count* between the group assigned to 12 weeks therapy and the group assigned to 32 weeks therapy; (S2) To determine whether or not there is a difference at 46-48 weeks after discontinuation of treatment in the *viral load level* between the group assigned to 12 weeks therapy and the group assigned to 32 weeks therapy. Thus the same primary and secondary end points are used for each arm of the study.

2.2 Clinical Data

There were a total of 18 subjects who were randomized for the clinical trial: 9 subjects were randomized to receive no therapy, 9 were randomized to receive therapy and then 4 were sub-randomized to receive 12 weeks therapy, and 5 were sub-randomized to receive 32 weeks therapy. However, only 8 subjects reached completion of the study: 2 of them are from the no-therapy arm, 4 are from the 12 weeks therapy arm and the remaining 2 subjects are from the 32 weeks therapy arm. In addition, the total observation times and intervals vary among these study subjects.

For the remaining 7 subjects in the no-treatment group, 2 went off study with withdrawn consent, one was lost to follow up, and 4 developed other ineligibility criteria and eventually went off study. One subject randomized to the no treatment arm started antiretroviral therapy on day 65 of the study.

Even though all the subjects in the 12-week treatment group finished the study protocols, two of them started therapy again in less than 30 weeks after discontinuation of the assigned treatment (far less than the designed 48 weeks). This is also true for one of the two subjects in the 32-week treatment group who completed the study. For the remaining 3 patients in the 32-week treatment group, two of them went off study with withdrawn consent. The last one stayed in study for only 103 days (far less than the assigned treatment period), and hence cannot be used for our analysis. For more information on the reulting data, we refer the interested reader to [4].

3 A Dynamic Modeling Approach

The above section reveals that there are 18 patients in this clinical trial and that two of them needed to be excluded from statistical analysis. This leaves 16 patients (8 of them randomized to receive no therapy, 4 of them assigned to 12 weeks therapy, and the remaining 4 subjects assigned to 32 weeks therapy). Half of the subjects did not complete the study (6 of them in the no-treatment group, and 2 of them in the 32-week treatment group). In addition, even for those 8 subjects who finished the trial, there are 3 subjects who started therapy again in less than 30 weeks after discontinuation of the assigned treatment (far less than the designed 48 weeks). Moreover, the observation times and intervals varied among subjects. This means that the endpoints for most of the participants are missing and the traditional imputation methods do not work well. To alleviate the resulting difficulty, we propose to use *mathematical modeling combined with an inverse problem approach* to obtain the predicted values of those endpoints for our analysis.

3.1 Mathematical Model

The model we used is adopted from [1] with descriptions of the state variables given in Table 1 and the schematic in Figure 1. It was found in [1] that this model has impressive predictive

States	Units	Descriptions
T_1	cells/μl-blood	uninfected activated CD4+ T cells
T_1^*	cells/μl-blood	infected activated CD4+ T cells
T_2	cells/μl-blood	uninfected resting CD4+ T cells
T_2^*	cells/μl-blood	infected resting (or latently infected) CD4+ T cells
V_I	RNA copies/ml-plasma	free infectious virus
V_{NI}	RNA copies/ml-plasma	free noninfectious virus
E_1	cells/μl-blood	HIV-specific effector CD8+ T cells
E_2	cells/μl-blood	HIV-specific memory CD8+ T cells

Table 1: Model States and their corresponding units and descriptions.

capability when comparing model simulations (with parameters estimated using only *half* of

the longitudinal observations) to the corresponding full clinical data sets. This model was further validated in [2, 3]. We suggest this provides a sufficient rationale and support for use of this model to predict the missing endpoints and predict the final outcomes for our clinical trials.

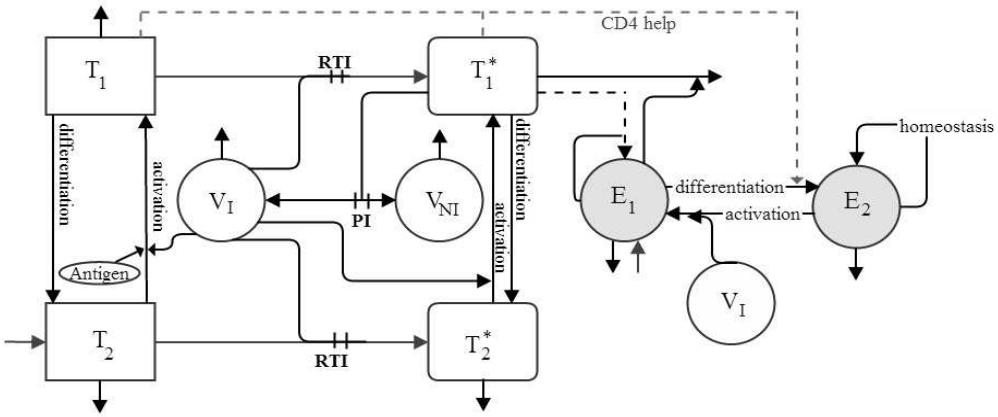

Figure 1: Flow chart of model (3.1)-(3.8) with compartments as described in Table 1. PI and RTI denote protease inhibitor and reverse transcriptase inhibitor, respectively.

The corresponding compartmental ordinary differential equation (ODE) model is given by

$$\dot{T}_1 = -d_1 T_1 - (1-\xi_1(t))k_1 V_I T_1 - \gamma_T T_1 + p_T\left(\frac{a_T V_I}{V_I+K_V} + a_A\right) T_2, \tag{3.1}$$

$$\dot{T}_1^* = (1-\xi_1(t))k_1 V_I T_1 - \delta T_1^* - mE_1 T_1^* - \gamma_T T_1^* + p_T\left(\frac{a_T V_I}{V_I+K_V} + a_A\right) T_2^*, \tag{3.2}$$

$$\dot{T}_2 = \lambda_T \frac{K_s}{V_I+K_s} + \gamma_T T_1 - d_2 T_2 - (1-f\xi_1(t))k_2 V_I T_2 - \left(\frac{a_T V_I}{V_I+K_V} + a_A\right) T_2, \tag{3.3}$$

$$\dot{T}_2^* = \gamma_T T_1^* + (1-f\xi_1(t))k_2 V_I T_2 - d_2 T_2^* - \left(\frac{a_T V_I}{V_I+K_V} + a_A\right) T_2^*, \tag{3.4}$$

$$\dot{V}_I = (1-\xi_2(t))10^3 N_T \delta T_1^* - cV_I - 10^3[(1-\xi_1(t))k_1 T_1 + (1-f\xi_1(t))k_2 T_2]V_I, \tag{3.5}$$

$$\dot{V}_{NI} = \xi_2(t)10^3 N_T \delta T_1^* - cV_{NI}, \tag{3.6}$$

$$\dot{E}_1 = \lambda_E + \frac{b_{E1}T_1^*}{T_1^*+K_{b1}}E_1 - \frac{d_E T_1^*}{T_1^*+K_d}E_1 - \delta_{E1}E_1 - \gamma_E \frac{T_1+T_1^*}{T_1+T_1^*+K_\gamma}E_1 + \frac{p_E a_E V_I}{V_I+K_V}E_2, \tag{3.7}$$

$$\dot{E}_2 = \gamma_E \frac{T_1+T_1^*}{T_1+T_1^*+K_\gamma}E_1 + \frac{b_{E2}K_{b2}}{E_2+K_{b2}}E_2 - \delta_{E2}E_2 - \frac{a_E V_I}{V_I+K_V}E_2, \tag{3.8}$$

with an initial condition

$$(T_1(0), T_1^*(0), T_2(0), T_2^*(0), V_I(0), V_{NI}(0), E_1(0), E_2(0))^T = (T_1^0, T_1^{*0}, T_2^0, T_2^{*0}, V_I^0, V_{NI}^0, E_1^0, E_2^0)^T.$$

The treatment factors $\xi_1(t) = \epsilon_1 u(t)$ in (3.1)-(3.4) and $\xi_2(t) = \epsilon_2 u(t)$ in (3.5)-(3.6) represent the effective treatment impact, consisting of efficacy factors ϵ_1 modeling the relative effectiveness of reverse transcriptase inhibitor (RTI), ϵ_2 describing the relative effectiveness of protease inhibitor (PI), and a time-dependent treatment function $u(t)$ $(0 \leq u(t) \leq 1)$ representing drug level, where $u(t) = 0$ is fully off and $u(t) = 1$ is fully on. 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.

In (3.1), $d_1 T_1$ denotes the natural death of T_1, and $(1 - \xi_1(t))k_1 V_I T_1$ is used to represent the infection process that results from encounters between the uninfected activated CD4+ T cells T_1 and free virus V_I. The term $\gamma_T T_1$ is used to account for the phenomenon of differentiation of uninfected activated CD4+ T cells into uninfected resting CD4+ T cells T_2. In (3.2), δT_1^* denotes the loss of infected activated CD4+ T cells due to the cytopathic effect of HIV, and the corresponding gain term for V_I include a multiplicative factor N_T to account for the number of RNA copies produced during this process. The term $mE_1T_1^*$ is used to account for the elimination of the infected activated CD4+ T cells by the HIV-specific effector CD8+ T cells, and $\gamma_T T_1^*$ is used to account for the phenomenon of differentiation of infected activated CD4+ T cells into latently infected CD4+ T cells T_2^* at rate γ_T.

In (3.3), $\lambda_T \frac{K_s}{V_I + K_s}$ is used to account for the source rate of naive CD4+ T cells, and $d_2 T_2$ denotes the natural death of T_2. The infection process that results from encounters between the uninfected resting CD4+ T cells T_2 and free virus V_I is represented by $(1 - f\xi_1)k_2 V_I T_2$, where the parameter f $(0 \leq f \leq 1)$ is used to account for the fact that treatment is potentially less effective in T_2 than in T_1. The term $\left(\frac{a_T V_I}{V_I + K_V} + a_A\right) T_2$ denotes the activation of the uninfected resting CD4+ T cells, and the corresponding gain term for T_1 include a multiplicative factor p_T to account for the net proliferation due to clonal expansion and programmed contraction. Similarly, $\left(\frac{a_T V_I}{V_I + K_V} + a_A\right) T_2^*$ in (3.4) is used to account for the activation of latently infected CD4+ T cells, and the corresponding gain term for T_1^* also includes a multiplicative factor p_T. The natural death of T_2^* is represented by $d_2 T_2^*$.

In (3.5) and (3.6), cV_I and cV_{NI} respectively denote the clearance of free infectious virus V_I and free noninfectious virus V_{NI}, and the factor 10^3 is introduced to convert between microliter and milliliter scales. The term $10^3[(1 - \xi_1(t))k_1 T_1 + (1 - f\xi_1(t))k_2 T_2]V_I$ in (3.5) is used to account for the removal of free virus that takes place when free virus infects T_1 and T_2, where one free virus particle is assumed to be responsible for each new infection.

The first four terms in the right hand side of (3.7) denotes the source, nonlinear infected cell-dependent birth, nonlinear infected cell-dependent death, and constant death, respectively. The term $\gamma_E \frac{T_1 + T_1^*}{T_1 + T_1^* + K_\gamma} E_1$ is used to include the essential role that activated CD4+ T cells play in the generation of memory CD8+ T cells, where the parameter K_γ is a half-saturation constant and γ_E is the maximum rate at which E_1 differentiates into E_2. In (3.8), $\frac{b_{E2} K_{b2}}{E_2 + K_{b2}} E_2 - \delta_{E2} E_2$ is used to denote the homeostatic regulation of E_2 with b_{E2} being the maximum proliferation rate and δ_{E2} being the death rate. The term $\frac{a_E V_I}{V_I + K_V} E_2$ denotes reactivation

of HIV-specific memory CD8+ T cells, and the corresponding gain terms for E_1 include a multiplicative factor p_E to account for the net proliferation due to clonal expansion and programmed contraction.

3.2 Inverse Problem

An explained in Section 2, we know that the observables are the total number of CD4+ T cells and the viral load level. For model (3.1)-(3.8), these two observables are respectively represented by

$$\begin{aligned} \bar{z}_1(t;\bar{q}) &= T_1^*(t;\bar{q}) + T_2^*(t;\bar{q}) + T_1(t;\bar{q}) + T_2(t;\bar{q}), \\ \bar{z}_2(t;\bar{q}) &= V_I(t;\bar{q}) + V_{NI}(t;\bar{q}), \end{aligned} \tag{3.9}$$

where $\bar{q}$ is a column vector for those model parameters and initial conditions that need to be estimated. With regard to the viral load measurements, it is worth noting that if the measurements of RNA copies are below the limit of quantification for the assay used (48 copies/ml-plasma or 20 copies/ml-plasma), then the observed viral load value is censored to be at its detection limit; that is, in these cases the "observed" values do not represent the true data values anymore. Furthermore, observations of viral load and CD4+ may not be at the same time points and the observation times and intervals vary among patients. So, in general, for the jth patient we have CD4+ T-cell data pairs $(t_1^{ij}, \bar{y}_1^{ij})$, $i = 1, \cdots, N_1^j$, and potentially different time point viral RNA data pairs $(t_2^{ij}, \bar{y}_2^{ij})$, $i = 1, \cdots, N_2^j$, $j = 1, 2, 3, \ldots, 16$ (patients are ordered as follows: the first 8 patients are those randomized to receive no therapy, the next 4 patients are those assigned to 12 weeks therapy and the last 4 patients are the ones assigned to 32 weeks therapy).

To obtain those individual-specific parameter estimates for the model, we carry out an inverse problem for each patient using his/her corresponding clinical data (that is, individuals are fitted individually). The algorithm we use in this paper is the same as the one in [1], where a statistically-based censored data method, an *expectation maximization* algorithm, was used.

For the simulation results shown below, we first estimated all the model parameters (31) and initial conditions (8) for each patient. We then fixed 14 parameters at the population averages across these 16 patients and re-estimated the remaining 25 parameters for each patient, where these fixed 14 parameters are chosen based on sensitivity (or lack thereof) analysis (see [4, Appendix A] for details) and are given in Table 2 along with their corresponding values.

After we reestimated the reduced set of 25 parameters for each patient, we simulated the individual CD4+ T-cell and viral load trajectories over the full time span of the individual patient's observations by using the parameter values obtained and then compared the resulting curves to the experimental data. Figure 2 depicts model fitting results for one of these patients, and reveals that the fits are reasonable. Model fitting results for the remainder of the patients are also either very good or as reasonable as those depicted in Figure 2 (see [3] for details).

Parameters	Values	Parameters	Values	Parameters	Values
ϵ_1	6.389e-01	f	5.303e-01	k_2	3.254e-09
λ_E	3.869e-04	K_{b1}	3.505e-02	K_{b2}	1.814e+02
γ_E	1.570e-03	K_γ	1.155e+00	δ_{E2}	2.120e-03
K_s	2.620e+06	T_2^{*0}	4.262e-01	V_I^0	4.486e+05
V_{NI}^0	8.469e+02	E_1^0	1.502e-02		

Table 2: Results for those parameters whose values are fixed at the population averages across the 16 patients investigated.

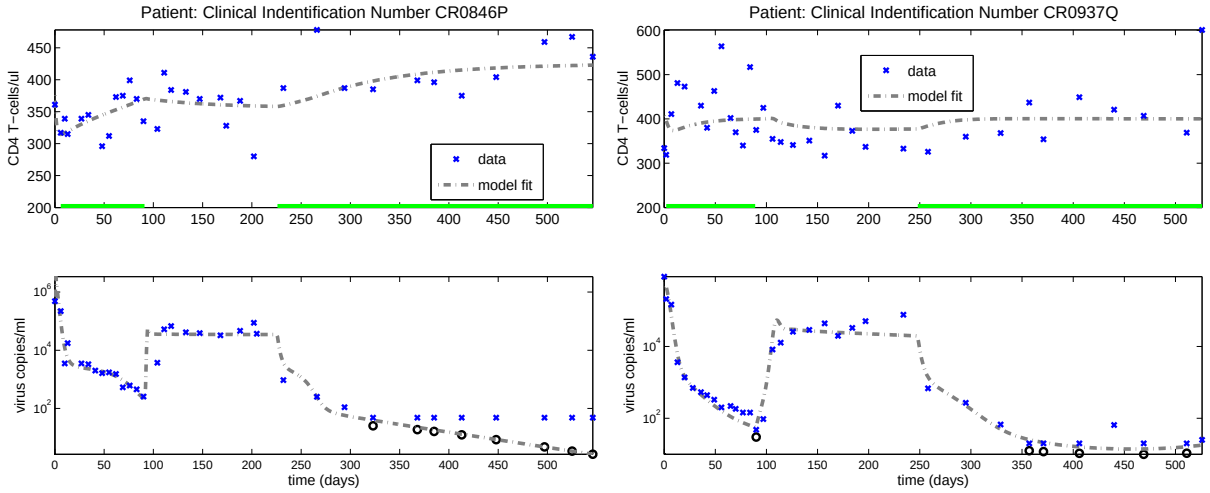

Figure 2: Model fitting results for one of the example patients in the 12-week treatment group, where time zero denotes the time point where we have the first observation, the solid (green) lines along the x-axis indicate periods when the patient is on antiviral therapy treatment, and the gray circles denote the predicted censored data values.

3.3 Statistical Analysis

In this section, we use the obtained predicated values of those endpoints for statistical analysis, and the Mann-Whitney U test (also called the Wilcoxon rank-sum test), is used to determine whether or not there is a difference in the viral load level and CD4+ T cell count between the group without treatment (8 patients) and the group with treatment (8 patients). This test is also used to determine whether or not there is a difference in the viral load level and CD4+ T cell count between the group assigned to 12 weeks therapy (4 patients) and the group assigned to 32 weeks therapy (4 patients). The Mann-Whitney U test is a nonparametric test that can be used to determine whether or not two random variables X_1 and X_2 have the same mean, and the associated statistic is based on the sum of ranks of assigned to the realizations of one random variable (i.e., either X_1 or X_2) when one orders the combined samples of X_1-realizations and X_2-realizations from least to greatest. For more information on this test, we refer the interested reader to [6, Chapter4].

Let $t_{p,1}^{\text{no}}$ and $t_{p,2}^{\text{no}}$ represent the two time points at which the measurements are taken to

determine the primary endpoints for the patients randomized to receive no therapy, $t_{p,1}^{12\mathrm{w}}$ and $t_{p,2}^{12\mathrm{w}}$ for the patients assigned to 12 weeks therapy, and $t_{p,1}^{32\mathrm{w}}$ and $t_{p,2}^{32\mathrm{w}}$ for the patients assigned to 32 weeks therapy. In addition, we let $\hat{\bar{q}}^j$ denote the estimated parameter values for the jth patient, $j = 1, 2, 3, \ldots, 16$.

For the primary objective (P1), the null hypothesis H_0 is that there is no difference in the CD4+ T cell count between the group without treatment and the group with treatment (that is, treatment has no effect), while the alternative hypothesis H_1 is that there is difference in the CD4+ T cell count between these two groups. The primary endpoint PEP_1^j for the jth patient is calculated as follows:

$$\mathrm{PEP}_1^j = \begin{cases} \dfrac{1}{2}(\log(\bar{z}_1(t_{p,1}^{\mathrm{no}}, \hat{\bar{q}}^j)) + \log(\bar{z}_1(t_{p,2}^{\mathrm{no}}, \hat{\bar{q}}^j))), & \text{if } j \in \{1, 2, 3, \ldots, 8\}, \\ \dfrac{1}{2}(\log(\bar{z}_1(t_{p,1}^{12\mathrm{w}}, \hat{\bar{q}}^j)) + \log(\bar{z}_1(t_{p,2}^{12\mathrm{w}}, \hat{\bar{q}}^j))), & \text{if } j \in \{9, 10, 11, 12\}, \\ \dfrac{1}{2}(\log(\bar{z}_1(t_{p,1}^{32\mathrm{w}}, \hat{\bar{q}}^j)) + \log(\bar{z}_1(t_{p,2}^{32\mathrm{w}}, \hat{\bar{q}}^j))), & \text{if } j \in \{13, 14, 15, 16\}. \end{cases} \tag{3.10}$$

(Recall that the first 8 patients are those patients randomized to receive no therapy, the next 4 patients are those assigned to 12 weeks therapy and the last 4 patients are the ones assigned to 32 weeks therapy.) Thus, for the primary objective (P1), the observations for the group without treatment and the ones for the group with treatment are respectively given by

$$\begin{aligned} \mathbb{P}_1^{\mathrm{no}} &= \{\mathrm{PEP}_1^1, \mathrm{PEP}_1^2, \mathrm{PEP}_1^3, \ldots, \mathrm{PEP}_1^8\}, \\ \mathbb{P}_1^{\mathrm{treat}} &= \{\mathrm{PEP}_1^9, \mathrm{PEP}_1^{10}, \mathrm{PEP}_1^{11}, \ldots, \mathrm{PEP}_1^{16}\}. \end{aligned} \tag{3.11}$$

Similarly, for the primary objective (P2), the null hypothesis H_0 is that there is no difference in the viral load level between the group without treatment and the group with treatment, while the alternative hypothesis H_1 is that there is difference in the viral load level between these two groups. The primary endpoint PEP_2^j for the jth patient is calculated as follows:

$$\mathrm{PEP}_2^j = \begin{cases} \dfrac{1}{2}(\log(\bar{z}_2(t_{p,1}^{\mathrm{no}}, \hat{\bar{q}}^j)) + \log(\bar{z}_2(t_{p,2}^{\mathrm{no}}, \hat{\bar{q}}^j))), & \text{if } j \in \{1, 2, 3, \ldots, 8\}, \\ \dfrac{1}{2}(\log(\bar{z}_2(t_{p,1}^{12\mathrm{w}}, \hat{\bar{q}}^j)) + \log(\bar{z}_2(t_{p,2}^{12\mathrm{w}}, \hat{\bar{q}}^j))), & \text{if } j \in \{9, 10, 11, 12\}, \\ \dfrac{1}{2}(\log(\bar{z}_2(t_{p,1}^{32\mathrm{w}}, \hat{\bar{q}}^j)) + \log(\bar{z}_2(t_{p,2}^{32\mathrm{w}}, \hat{\bar{q}}^j))), & \text{if } j \in \{13, 14, 15, 16\}. \end{cases} \tag{3.12}$$

Thus, for the primary objective (P2), the observations for the group without treatment and the ones for the group with treatment are respectively given by

$$\begin{aligned} \mathbb{P}_2^{\mathrm{no}} &= \{\mathrm{PEP}_2^1, \mathrm{PEP}_2^2, \mathrm{PEP}_2^3, \ldots, \mathrm{PEP}_2^8\}, \\ \mathbb{P}_2^{\mathrm{treat}} &= \{\mathrm{PEP}_2^9, \mathrm{PEP}_2^{10}, \mathrm{PEP}_2^{11}, \ldots, \mathrm{PEP}_2^{16}\}. \end{aligned} \tag{3.13}$$

Using (3.11) and (3.13), we can then use Matlab command "ranksum" to obtain the p-values for the primary objectives (P1) and (P2). They are respectively given by 0.7984 and 0.6454.

This indicates that there is no evidence that there is difference in the CD4+ T cell count and the viral load level between the group without treatment and the group with treatment. In summary, there is no evidence that treatment has an effect.

For the secondary objectives (S1), the observations for the group with 12 weeks treatment and the ones for the group with 32 weeks treatment are respectively given by

$$\begin{aligned} \mathbb{S}_1^{12w} &= \{\mathrm{PEP}_1^9, \mathrm{PEP}_1^{10}, \mathrm{PEP}_1^{11}, \mathrm{PEP}_1^{12}\}, \\ \mathbb{S}_1^{32w} &= \{\mathrm{PEP}_1^{13}, \mathrm{PEP}_1^{14}, \mathrm{PEP}_1^{15}, \mathrm{PEP}_1^{16}\}. \end{aligned} \tag{3.14}$$

For the secondary objectives (S2), the observations for the group with 12 weeks treatment and the ones for the group with 32 weeks treatment are respectively given by

$$\begin{aligned} \mathbb{S}_2^{12w} &= \{\mathrm{PEP}_2^9, \mathrm{PEP}_2^{10}, \mathrm{PEP}_2^{11}, \mathrm{PEP}_2^{12}\}, \\ \mathbb{S}_2^{32w} &= \{\mathrm{PEP}_2^{13}, \mathrm{PEP}_2^{14}, \mathrm{PEP}_2^{15}, \mathrm{PEP}_2^{16}\}. \end{aligned} \tag{3.15}$$

Using (3.14) and (3.15), we find that the p-values for the secondary objectives (S1) and (S2) are respectively given by 0.0571 and 0.3429. This indicates that there is no evidence that there is difference in the viral load level between the group assigned to 12 weeks therapy and the group assigned to 32 weeks therapy. We also fail to reject the null hypothesis of (S1) with 5% significance level.

4 Concluding Remarks

Compared to traditional imputation methods used for analysis of longitudinal clinical trials in the presence of missing endpoints, our proposed dynamical modeling approach accounts for the fact that the interested outcomes are usually varying over time and the dropout rates and the time to dropout are often different among intervention groups. This is exactly the case for the HIV random clinical trial considered in this paper. This also can provide more accurate/reliable estimates of missing endpoints.

The model used to predict those missing endpoints for the HIV randomized clinical trial was carefully validated by previously collected data sets. Based on the previously validated model, our findings here suggest that we fail to reject the null hypotheses and hence that there is no benefit for treating subjects during acute infection. However, we must point out that because of the limited number of patients completing the study that the statistical analysis described above is far from conclusive.

Acknowledgements

This research was supported in part by grant number NIAID R01AI071915-10 from the National Institute of Allergy and Infectious Diseases, and in part by the Air Force Office of

Scientific Research under grant number AFOSR FA9550-12-1-0188.

References

[1] H.T. Banks, M. Davidian, S. Hu, G.M. Kepler and E.S. Rosenberg, Modeling HIV immune response and validation with clinical data, *Journal of Biological Dynamics*, **2** (2008), 357–385.

[2] H.T. Banks, R. Baraldi, K. Cross, C. McChesney, L. Poag, E. Thorpe, and K.B. Flores, Uncertainty quantification in modeling HIV viral mechanics, *Mathematical Biosciences and Engineering*, **12** (2015), 937–964.

[3] H.T. Banks, K.B. Flores, S. Hu, E. Rosenberg, M. Buzon, X. Yu and M. Lichterfeld, Immuno-modulatory strategies for reduction of HIV reservoir cells , *Journal Theoretical Biology*, **372** (2015), 146-158.

[4] H.T. Banks, S. Hu and E.S. Rosenberg, Summary Report for HIV Random Clinical Trial Conducted in 2014, CRSC-TR14-16, North Carolina State University, Raleigh, NC.

[5] J.D. Dziuraa, L.A. Post, Q. Zhao, Z. Fub, and P. Peduzzi, Strategies for dealing with missing data in clinical trials: from design to analysis, *Yale Journal of Biology And Medicine*, **86** (2013), 343–358.

[6] M. Hollander, D.A. Wolfe and E. Chicken, *Nonparametric Statistical Methods*, 3rd edition, Jon Wiley & Sons, Hoboken, New Jersey, 2014.

[7] W.R. Myers, Handling missing data in clinical trials: an overview, *Drug Information Journal*, **34** (2000), 525–533.