Maximum Likelihood Estimation in Dynamical Models of HIV — Source link ⤴

Jeremie Guedj, Rodolphe Thiébaut, Daniel Commenges

Institutions: French Institute of Health and Medical Research

Published on: 01 Dec 2007 - Biometrics (Blackwell Publishing Inc)

Topics: Ode and Inference

Related papers:

- HIV-1 Dynamics in Vivo: Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time

- Hierarchical Bayesian Methods for Estimation of Parameters in a Longitudinal HIV Dynamic System

- A Bayesian approach to parameter estimation in HIV dynamical models.

- Maximum likelihood estimation in nonlinear mixed effects models

- Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection

Maximum likelihood estimation in dynamical models of HIV.

Jérémie Guedj, Rodolphe Thiébaut, Daniel Commenges

Maximum Likelihood Estimation in Dynamical Models of HIV

J. Guedj,[1,2] **R. Thiébaut**[1,2,*] **and D. Commenges**[1,2]

[1]INSERM, U875 (Biostatistique), Bordeaux, F-33076, FRANCE

[2]Université Victor Segalen Bordeaux 2, Bordeaux, F-33076, FRANCE

SUMMARY.   The study of dynamical models of HIV infection, based on a system of non-linear Ordinary Differential Equations (ODE), has considerably improved the knowledge of its pathogenesis. While the first models used simplified ODE systems and analyzed each patient separately, recent works dealt with inference in non-simplified models borrowing strength from the whole sample. The complexity of these models leads to great difficulties for inference and only the Bayesian approach has been attempted by now. We propose a full likelihood inference, adapting a Newton-like algorithm for these particular models. We consider a relatively complex ODE model for HIV infection and a model for the observations including the issue of detection limits. We apply this approach to the analysis of a clinical trial of antiretroviral therapy (ALBI ANRS 070) and we show that the whole algorithm works well in a simulation study.

KEY WORDS: Ordinary Differential Equations; HIV dynamics; Non-linear mixed effects models; Likelihood inference

*email: rt@isped.u-bordeaux2.fr

1

1.   Introduction

Studies of the human immunodeficiency virus (HIV) dynamics using biomathematical models have considerably improved the knowledge of the pathogenesis of this infection. For instance, such studies have demonstrated the high turnover of infected cells as well as that of free virions (Ho et al., 1995; Perelson et al., 1996). These pioneering works considered models based on a system of non-linear Ordinary Differential Equations (ODE) without closed-form solution, which were simplified and linearized. The estimation of the parameters was then performed for each subject using simple non-linear regression methods (Perelson et al., 1996). Population approaches that estimate parameters of simplified models using the data from all the subjects of a sample have been proposed (Wu and Ding, 1999): such models are in the framework of Non-Linear Mixed-Effect (NLME) models (Pinheiro and Bates, 2000). However, as noted by Huang et al. (2006), these linearized ODE models are not able to describe the course of the infection for a long-term period. Moreover, only the virus dynamics is taken into account, neglecting the evolution of the CD4+ T lymphocytes (CD4) count. Last, these models do not include parameters such as treatment efficacy, and thus do not help much to a better understanding of the mechanisms of the infection or an evaluation of treatment effect: this is why it is important to work with non-simplified models (that is, based on non-linear ODE systems) which are able to represent the complexity of the dynamics. Several authors (Putter et al., 2002; Huang et al., 2006) have attacked the ambitious program of treating these models in a population context, all using a Bayesian approach. The two above-mentioned difficulties are combined: non closed-form of the solution

2

of the ODE system and numerical integrals involved in the likelihood. In the
Bayesian approach the direct numerical integrations are avoided and the approach relies on the MCMC algorithms to give the *a posteriori* distributions
of the parameters (Gilks et al., 1996). Of course *a priori* distributions have
to be given, and this is both an advantage (*a priori* knowledge can be put in
the model) and a drawback (it is not always clear what quantity of *a priori*
knowledge has been introduced) of the approach.

Although the Bayesian approach may be very attractive in such a complex
problem, the maximum likelihood approach retains some advantages: there
is no need to specify *a priori* distributions, and there is a very well established theory of inference, one can easily compare parametric models using
the Akaike criterion. Moreover, Newton-like methods may be relatively fast
compared to the heavy computation involved in Bayesian approach, convergence criteria are well defined and last, at convergence, the algorithm
gives the maximum likelihood and an approximation of its second derivative,
which is an estimate of the information matrix. To avoid numerical difficulties, mainly due to the need to compute numerically multiple integrals
involved in the likelihood, several approximations of the likelihood have been
proposed (Pinheiro and Bates, 2000) but they may lead to inconsistent estimations (Ding and Wu, 2001). The problem is even more important when no
closed-form solution is available, making numerical derivatives of the likelihood unstable and classical softwares inadequate (Putter et al., 2002; Huang
et al., 2006).

The aim of our work was to develop a maximum likelihood approach to
this problem based on an adaptation of a Newton-like method. We propose

3

a method for calculating with a good precision the likelihood and its score in ODE models; we show that a Newton-like method using only the first derivatives is adequate. We illustrate this approach with data from a clinical trial of antiviral therapy using a rather complex non-linear ODE model with five components. We were able to take into account left-censored data coming from the detection limits in the assays used for quantifying plasma HIV RNA level (viral load). In fact, ignoring this censoring leads to important biases in the estimation (Thiébaut et al., 2006).

The paper is organized as follows. In section 2, we describe the general population ODE model together with the observations we get from this model; in section 3, we present our inference approach, that is the likelihood and our maximization algorithm. In section 4, we describe an HIV dynamics model. In section 5, we analyse the data of the ALBI ANRS 070 trial. We show in section 6 that the algorithm works well through a simulation study.

2. Statistical Model

2.1   Model for the system

Let us consider an ODE model for a population of subjects. For subject i with $i = 1, ...n$, this can be written:

$$\begin{cases} \frac{d\boldsymbol{X}^{(i)}(t)}{dt} = f(\boldsymbol{X}^{(i)}(t), \boldsymbol{\xi}^{(i)}) \\ \boldsymbol{x}^{(i)}(0) = h(\boldsymbol{\xi}^{(i)}) \end{cases} \tag{1}$$

where $\boldsymbol{X}^{(i)}(t) = (X_1^{(i)}(t), ..., X_K^{(i)}(t))'$ is the vector of the K state variables (or components). We write $\boldsymbol{X}(t, \boldsymbol{\xi}^{(i)}) = \boldsymbol{X}^{(i)}(t)$ to underline that $\boldsymbol{\xi}^{(i)}$ completely determines the trajectories $\boldsymbol{X}^{(i)}(t)$. We assume that f and h are twice differentiable with respect to $\boldsymbol{\xi}^{(i)}$; $\boldsymbol{\xi}^{(i)} = (\xi_1^{(i)}, ..., \xi_p^{(i)})'$ ($'$ for transpose) is a vector of p individual parameters which appear naturally in the ODE

4

system and have generally a biological interpretation. We introduce a parsimonious model for $\boldsymbol{\xi}^{(i)}$ to allow inter-individual variability: the variability may be explained, through explanatory variables, or unexplained, and this is accounted for by random effects. Similarly to generalized (mixed) linear models, we introduce a link function which relates $\boldsymbol{\xi}^{(i)}$ to a linear model involving explanatory variables and random effects. For sake of simplicity we restrict to component-wise transforms:

$$\begin{cases} \tilde{\xi}_l^{(i)} = \Psi_l(\xi_l^{(i)}), \\ \tilde{\xi}_l^{(i)} = \phi_l + \boldsymbol{z_l^{(i)}}'\boldsymbol{\beta_l} + \boldsymbol{w_l^{(i)}}'\boldsymbol{b^{(i)}}, \quad l \le p \end{cases} \tag{2}$$

where ϕ_l is the intercept, $\boldsymbol{z_l^{(i)}}$ and $\boldsymbol{w_l^{(i)}}$ are the vectors of explanatory variables associated respectively to the fixed and to the random effects of the lth biological parameter. The $\boldsymbol{\beta_l}$'s are vectors of regression coefficients associated to the fixed effects. We assume $\boldsymbol{b^{(i)}} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$, where $\boldsymbol{b^{(i)}}$ is the individual vector of random effects of dimension q. Let $\boldsymbol{A} = (a_{l''l'})_{l' \le l'' \le q}$ the lower triangular matrix with positive diagonal elements such that $\boldsymbol{AA'} = \boldsymbol{\Sigma}$ (Cholesky decomposition). We can write $\boldsymbol{b^{(i)}} = \boldsymbol{Au^{(i)}}$ with $\boldsymbol{u^{(i)}} \sim \mathcal{N}(0, I_q)$.

2.2 Model for the observations

It often happens that not all the components of the system can be observed. Functions $g_m(.), m = 1, \ldots, M$ of $\mathbb{R}^K$ to $\mathbb{R}$ are introduced to link the potential observations to the original system; they are assumed to be twice differentiable. These functions allow to observe only some of the components of the original system, or observation of combinations of several components: for instance the model may distinguish between non-infected and infected CD4, but only the total number of CD4 is observed. These functions may also include known transformations such as the logarithm; the $g_m(.)$ are thus

assumed completely known and will be called the *observable components*. Let Y_{ijm} denote the jth measurements of the mth observable component for subject i at time t_{ijm}; we assume that:

$$Y_{ijm} = g_m(\boldsymbol{X}(t_{ijm}, \tilde{\boldsymbol{\xi}}^{(i)})) + \epsilon_{ijm} \quad j = 1, ...n_{im}, \quad m = 1, ..., M \qquad (3)$$

where the ϵ_{ijm} are independent Gaussian with zero mean and variances σ_m^2. The ϵ_{ijm}'s are supposed independent because they represent measurement errors: this assumption can be roughly checked by looking at the correlations between residuals (see section 5). Thus, both the observed dependencies for the within-patient observations of a given biomarker and more generally the correlations between the biomarkers $\boldsymbol{Y}$'s are completely determined by the mechanistic relationships between the $\boldsymbol{X}$'s produced by $f(.)$ in the model (1). If necessary, the model could be extended by adding correlated random-effects in model (3) or other covariance structures to reflect any between-variables dependency not captured by $f(.)$

The model for the observation may be complicated by the detection limits of assays leading to left-censored observations Y_{ijm}. This is the case for HIV RNA concentration defined as the first observed component ($m = 1$) in the following. We observe Y_{ij1} or the event $\{Y_{ij1} < \zeta\}$, where ζ is the detection limit. The model can be easily generalized to upper detection limits or detection limits depending on time.

3. Inference

3.1 *Log-Likelihood*

Denoting $\delta_{ij} = I_{\{Y_{ij1} > \varsigma\}}$, the full individual likelihood given the random effects $\mathcal{L}_{\mathcal{F}_i | \boldsymbol{u}^{(i)}}$ is given by:

$$
\mathcal{L}_{\mathcal{F}_i | \boldsymbol{u}^{(i)}} = \prod_{j \leq n_{i1}} \left\{ \frac{1}{\sigma_1 \sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{Y_{ij1} - g_1(\boldsymbol{X}(t_{ij1}, \tilde{\boldsymbol{\xi}}^{(i)}))}{\sigma_1} \right)^2 \right] \right\}^{\delta_{ij}} \left\{ \Phi\left(\frac{\varsigma - g_1(\boldsymbol{X}(t_{ij1}, \tilde{\boldsymbol{\xi}}^{(i)}))}{\sigma_1} \right) \right\}^{1-\delta_{ij}}
$$

$$
\prod_{m>1,\ j \leq n_{im}} \left\{ \frac{1}{\sigma_m \sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{Y_{ijm} - g_m(\boldsymbol{X}(t_{ijm}, \tilde{\boldsymbol{\xi}}^{(i)}))}{\sigma_m} \right)^2 \right] \right\}
$$

where Φ is the cumulative distribution function of the standard univariate normal distribution. The observed individual likelihood $\mathcal{L}_{\mathcal{O}_i}$ is obtained from $\mathcal{L}_{\mathcal{F}_i | \boldsymbol{u}^{(i)}}$ as:

$$
\mathcal{L}_{\mathcal{O}_i} = \int_{\mathbb{R}^q} \mathcal{L}_{\mathcal{F}_i | \boldsymbol{u}^{(i)}}(\mathbf{u}) \phi(\mathbf{u}) d\mathbf{u} \tag{4}
$$

where ϕ is the multivariate normal density of $N(0, I_q)$. We will denote $L_{\mathcal{F}_i | \boldsymbol{u}^{(i)}} = \log \mathcal{L}_{\mathcal{F}_i | \boldsymbol{u}^{(i)}}$ and $L_{\mathcal{O}_i} = \log \mathcal{L}_{\mathcal{O}_i}$ the full (given random effects) and observed individual log-likelihoods, respectively. The global observed log-likelihood is $L_{\mathcal{O}} = \sum_{i \leq n} L_{\mathcal{O}_i}$. The integrand in (4) is centered and scaled as suggested for the adaptive Gaussian quadrature in Pinheiro and Bates (2000) and the integral is then computed with an efficient algorithm developed by Genz and Keister (1996) (see Web Appendix A & C for more details).

3.2 *Algorithm of Likelihood Maximization*

For likelihood inference, we propose a Newton-like method which uses only the first derivatives of the log-likelihood (the score).

Computation of the score

The computation of the score proceeds in two stages: first compute the score of the full likelihood given random effects; second compute the score of the observed likelihood by integration using the relationships given by Louis

(1982) and generalized by Commenges and Rondeau (2006); this approach was used by Hedeker and Gibbons (1994) in another context. For simplicity, we assume in this section that there is no censored data. For subject i at the current point $\boldsymbol{\theta} = ((\phi_l)_{l \leq p}, (\boldsymbol{\beta_l})_{l=1,p}, \boldsymbol{A} = (a_{ll'})_{l' \leq l \leq q}, \boldsymbol{\sigma} = (\sigma_l)_{l \leq M})$ the components of the full score can be written as follows:

$$U^{(\phi_l)}_{\mathcal{F}_i | \boldsymbol{u^{(i)}}}(\boldsymbol{\theta}) = \frac{\partial L_{\mathcal{F}_i | \boldsymbol{u^{(i)}}}}{\partial \tilde{\xi}_l^{(i)}} = \sum_{m \leq M, \ j \leq n_{im}} \frac{1}{\sigma_m^2} \frac{\partial g_m(\boldsymbol{X}(t_{ijm}, \tilde{\boldsymbol{\xi}}^{(i)}))}{\partial \tilde{\xi}_l^{(i)}} \left[Y_{ijm} - g_m(\boldsymbol{X}(t_{ijm}, \tilde{\boldsymbol{\xi}}^{(i)})) \right]$$

$$\boldsymbol{U}^{(\boldsymbol{\beta_l})}_{\mathcal{F}_i | \boldsymbol{u^{(i)}}}(\boldsymbol{\theta}) = \frac{\partial L_{\mathcal{F}_i | \boldsymbol{u^{(i)}}}}{\partial \boldsymbol{\beta_l}} = \boldsymbol{z_l^{(i)}} U^{(\phi_l)}_{\mathcal{F}_i | \boldsymbol{u^{(i)}}}(\boldsymbol{\theta})$$

$$U^{(a_{ll'})}_{\mathcal{F}_i | \boldsymbol{u^{(i)}}}(\boldsymbol{\theta}) = \frac{\partial L_{\mathcal{F}_i | \boldsymbol{u^{(i)}}}}{\partial a_{ll'}} = \sum_{m \leq M, \ j \leq n_{im}} \frac{1}{\sigma_m^2} \left(Y_{ijm} - g_m(\boldsymbol{X}(t_{ijm}, \tilde{\boldsymbol{\xi}}^{(i)})) \right) \left(u_{l'}^{(i)} \sum_{l'' \leq p} w_{l''l}^{(i)} \frac{\partial g_m(\boldsymbol{X}(t_{ijm}, \tilde{\boldsymbol{\xi}}^{(i)}))}{\partial \tilde{\xi}_{l''}^{(i)}} \right.$$

$$U^{(\sigma_l)}_{\mathcal{F}_i | \boldsymbol{u^{(i)}}}(\boldsymbol{\theta}) = \frac{\partial L_{\mathcal{F}_i | \boldsymbol{u^{(i)}}}}{\partial \sigma_l} = \sum_{j \leq n_{il}} \frac{(Y_{ijl} - g_l(\boldsymbol{X}(t_{ijl}, \tilde{\boldsymbol{\xi}}^{(i)})))^2}{\sigma_l^3} - \frac{n_{il}}{\sigma_l}$$

Using the fact that:

$$\frac{\partial g_m(\boldsymbol{X}(t, \tilde{\boldsymbol{\xi}}^{(i)}))}{\partial \tilde{\xi}_l^{(i)}} = \sum_{k \leq K} \frac{\partial g_m(\boldsymbol{X}(t, \tilde{\boldsymbol{\xi}}^{(i)}))}{\partial X^{(k)}} \frac{\partial X^{(k)}(t, \tilde{\boldsymbol{\xi}}^{(i)})}{\partial \tilde{\boldsymbol{\xi}}^{(i)}},$$

the computation of the full score requires to solve numerically the p systems of sensitivity equations $\frac{\partial X^{(k)}(t, \tilde{\boldsymbol{\xi}}^{(i)})}{\partial \tilde{\xi}_l^{(i)}}$ (see example in the Web Appendix B). Then the observed scores can be deduced by Louis' formula:

$$\boldsymbol{U}_{\mathcal{O}_i} = \frac{\partial L_{\mathcal{O}_i}}{\partial \boldsymbol{\theta}} = (\mathcal{L}_{\mathcal{O}_i})^{-1} \int_{\mathbb{R}^q} \mathcal{L}_{\mathcal{F}_i | \boldsymbol{u^{(i)}}}(\mathbf{u}) \boldsymbol{U}_{\mathcal{F}_i | \boldsymbol{u^{(i)}}}(\mathbf{u}) \phi(\mathbf{u}) d\mathbf{u}.$$

The integrals can be computed by adaptive Gaussian quadrature using the same transformation as for the computation of $L_{\mathcal{O}_i}$. Then, the global observed score is $\boldsymbol{U} = \boldsymbol{U}_{\mathcal{O}} = \sum_{i \leq n} \boldsymbol{U}_{\mathcal{O}_i}$.

8

The maximization algorithm

The Newton-Raphson method, or the more robust Marquardt algorithm (Marquardt, 1963), is the most efficient algorithm when the log-likelihood is not too far from a quadratic function. This approach requires to compute the score $\boldsymbol{U}$ and the Hessian $\boldsymbol{H}$ at each current point θ_k of the maximization procedure. Although a semi-analytical expression for the Hessian could be obtained with the same two-stages approach as for the score, the computational burden would become unbearable and we propose to use an iterative method in which $H(\theta_k)$ is replaced by $G(\theta_k) = \sum_{i \leq n} \boldsymbol{U}_{\mathcal{O}_i}(\theta_k) \boldsymbol{U}'_{\mathcal{O}_i}(\theta_k) + \frac{\nu}{n} \boldsymbol{U}(\theta_k) \boldsymbol{U}'(\theta_k)$ (where ν is a weighting coefficient). We have that $n^{-1}G(\hat{\theta})$ converges toward $n^{-1}I(\theta^*)$ where $I(\theta^*)$ is the information matrix under the true probability, θ^* being the true parameter value. Thus $G(\theta_k)$ should be a good approximation of $H(\theta_k)$ near the maximum since $n^{-1}H(\hat{\theta})$ itself converges toward $n^{-1}I(\theta^*)$. As for convergence criterion we use $C(\theta_k) = \boldsymbol{U}(\theta_k)'G^{-1}(\theta_k)\boldsymbol{U}(\theta_k)$. $C(\theta^*)$ has asymptotically a χ^2_p distribution; this gives an idea of which value should be considered as "small". Once the convergence is obtained one may use $G(\hat{\theta})$ as an estimator of $I(\theta^*)$ to build confidence intervals and Wald tests. However since $\hat{\theta}$ is the value for which $\boldsymbol{U}(\hat{\theta}) = 0$, we may expect that the variance of $\boldsymbol{U}$ computed in $\hat{\theta}$ is a negatively biased estimate of its variance at θ^* (that is $I(\theta^*)$). This bias is difficult to estimate in general. In the linear model with known error variance, it can be shown that $E[\boldsymbol{U}(\hat{\beta})\boldsymbol{U}'(\hat{\beta})] = \frac{n-dim(\beta)}{n}I(\beta^*)$, where β is the vector of regression coefficients. By analogy we propose to estimate $I(\theta^*)$ by $\frac{n}{n-dim(\theta)}G(\hat{\theta})$. The whole algorithm (iteration and convergence criterion) has the property that it is invariant under any affine transformation of the parameters. Details about the implementation may be found in

the Web Appendix C.

3.3 *Expectations & Predictions*

The expected trajectory can be obtained by simulating a sample of subjects and averaging, for each time and each marker, over their values.

Also, individual predicted trajectories can be computed as $\hat{\boldsymbol{X}}^{(i)}(t) = \boldsymbol{X}(t, \hat{\tilde{\boldsymbol{\xi}}}^{(i)})$ where $\hat{\tilde{\xi}}_l^{(i)} = \hat{\phi}_l + \boldsymbol{z_l^{(i)}}' \hat{\boldsymbol{\beta_l}} + \boldsymbol{w_l^{(i)}}' \hat{\boldsymbol{A}} \hat{\mathbf{u}}^{(i)}$ and $\hat{\mathbf{u}}^{(i)}$ is the posterior mode (given the data) of $\boldsymbol{u^{(i)}}$. From this we can deduce individual predicted trajectories of observed components. Then, the fit can be checked by comparing the predicted values of the components $\hat{Y}_{ijm} = g_m(\hat{\boldsymbol{X}}^{(i)}(t_{ijm}))$ with the observations Y_{ijm}.

4. Biological Model

4.1 *Motivating application*

As an application of the proposed method, we aimed at estimating the difference of treatment effects in a randomized clinical trial (Molina et al., 1999). The ALBI ANRS 070 trial compared over 24 weeks the combination of zidovudine plus lamivudine (AZT+3TC) with that of stavudine plus didanosine (ddI+d4T)(a third arm alternating from one regimen to another was not considered in this paper). The inclusion criteria were CD4 $\geq$ 200 cells/μL and HIV RNA level between 4 and 5 log_{10} copies/mL within 15 days before entry into the study. Measurements were taken once a month up to six months. Spaghetti plots of the data are shown on Figure 1 .

[Figure 1 about here.]

The primary outcome measure defined in the study protocol was the antiretroviral effect as measured by the mean change in HIV RNA level between baseline and 24 weeks by use of the ultra-sensitive PCR assay with

10

lower limit of quantification of 50 copies/mL (1.7 log_{10}). In the main analysis of Molina et al. (1999), HIV RNA values reported as < 50 copies/mL were considered equivalent to 50 copies/mL; 51 patients were included in each treatment group. Over the 24-week period, HIV RNA level declined in the two groups, with mean (SE) decreases at the end of the study of 1.26 (0.09) log_{10} copies/mL in the AZT+3TC group and 2.26 (0.11) log_{10} copies/mL in the ddI+d4T group. The mean increase in CD4 count was larger in ddI+d4T group than in AZT+3TC group (124 cells/μL vs. 62 cells/μL, p=0.012).

4.2   A mathematical model for HIV dynamics

Conventional models for HIV dynamics have involved target cells (mainly uninfected CD4), infected cells producing viruses, and circulating viruses (Perelson et al., 1996). Because the activation of CD4 has been recognized as a central role in HIV pathogenesis (Grossman et al., 2000), the activated state is worth distinguishing. Actually, activated cells make a better target than quiescent cells and viral replication is rapid and efficient in activated cells (Ribeiro et al., 2002). Also, it is useful to distinguish infectious and non-infectious virions because non-infectious virions are predominant (Chun et al., 1997). In the ALBI ANRS 070 trial, antiretroviral therapy included reverse transcriptase inhibitors only. This type of antiretroviral drugs limits cell infection by inhibiting reverse transcription of HIV RNA and thus was modeled by limiting the new production of T^* through the parameter η. The

11

model can be written as:

$$\begin{cases} \frac{dQ}{dt} = \lambda + \rho T - \alpha Q - \mu_Q Q \\ \frac{dT}{dt} = \alpha Q - (1-\eta)\gamma T V_I - \rho T - \mu_T T \\ \frac{dT^*}{dt} = (1-\eta)\gamma T V_I - \mu_{T^*} T^* \\ \frac{dV_I}{dt} = \omega \mu_{T^*} \pi T^* - \mu_v V_I \\ \frac{dV_{NI}}{dt} = (1-\omega)\mu_{T^*} \pi T^* - \mu_v V_{NI} \end{cases} \qquad (5)$$

where Q, T, T^* are quiescent non-infected, activated non-infected, and activated infected CD4 and V_I and V_{NI} are infectious and non-infectious free virions, respectively. Figure 2 displays a graphical representation of the system.

[Figure 2 about here.]

The definition of each parameter of this system of non-linear differential equations is reported in Table 1. We make the assumption that before initiation of antiretroviral treatment the values of the state variables are that of steady state of the ODE system with $\eta = 0$. This assumption implies that the treatment is initiated far from initial infection. The steady state assumption leads to the following initial conditions (where $t = 0$ refers to treatment initiation):

$$\begin{cases} Q(0) = \frac{1}{\alpha+\mu_Q}(\lambda + \frac{\rho\mu_v}{\omega\gamma\pi}) \\ T(0) = \frac{\mu_v}{\omega\gamma\pi} \\ T^*(0) = \frac{1}{\mu_{T^*}}\left(\frac{\alpha}{\alpha+\mu_Q}(\lambda + \frac{\rho\mu_v}{\omega\gamma\pi}) - \frac{(\rho+\mu_T)\mu_v}{\omega\gamma\pi}\right) \\ V_I(0) = \frac{\alpha\omega\pi}{\mu_v(\alpha+\mu_Q)}(\lambda + \frac{\rho\mu_v}{\omega\gamma\pi}) - \frac{\rho+\mu_T}{\gamma} \\ V_{NI}(0) = \frac{(1-\omega)\pi}{\mu_v}\left(\frac{\alpha}{\alpha+\mu_Q}(\lambda + \frac{\rho\mu_v}{\omega\gamma\pi}) - \frac{\mu_v(\rho+\mu_T)}{\omega\gamma\pi}\right) \end{cases}$$

4.3 *The statistical model*

We used the structure defined in section 2 to estimate the parameters of the biological model (5). Because the study was not designed for dynamic

12

modelling, it was not possible to estimate all the parameters. In the present study, the first measurement after therapy was performed four weeks later rather than several hours in some studies of virus dynamics (Ho et al., 1995). We consequently chose to fix the set $(\mu_Q, \mu_v, \rho, \omega)$ whose estimates can be found in the literature (Table 1). The parameter γ was at the limit of non-identifiability: we determined it in a plausible range of values by profile likelihood.

The vector of natural parameters for subject i was then: $\boldsymbol{\xi}^{(i)} = (\lambda^{(i)}, \alpha^{(i)}, \eta^{(i)}, \mu_T^{(i)}, \mu_{T^*}^{(i)}, \pi^{(i)})'$. The link functions were the log transform for all parameters (because they must be positive) except for $\eta^{(i)}$ for which we took the inverse logistic function (because $0 < \eta^{(i)} < 1$): $\tilde{\eta}^{(i)} = \log \frac{\eta^{(i)}}{1-\eta^{(i)}}$. We introduced in the model only one explanatory variable $z^{(i)}$, which represented the treatment group: $z^{(i)} = 0$ for the AZT+3TC group and $z^{(i)} = 1$ for the ddI+d4T group. $\tilde{\eta}_0$ represented the treatment effect for AZT+3TC in the logistic scale and β represented the differential effect of ddI+d4T relative to AZT+3TC in the logistic scale: $\tilde{\eta}^{(i)} = \tilde{\eta}_0 + z^{(i)}\beta$

Concerning the selection of random effects, we favored a forward selection strategy because the number of random effects is limited by the amount of information available as well as by the curse of dimensionality (the dimension of the multiple integrals is equal to the number q of random effects). Starting with the model without random effect, we introduced one random effect successively on each parameter and selected the one which most increased the likelihood; then we proceeded to include a second random effect according to the same criterion, and so on until the model with q random effects was not rejected by a likelihood ratio test (the distribution of the likelihood ratio

13

statistic being a mixture of chi-square distributions).

Observable components g_1 and g_2 were transforms of the HIV RNA concentration and the total CD4 count respectively with $g_1 = log_{10}(V_I + V_{NI})$ and $g_2 = (Q + T + T^*)^{0.25}$. These transformations of HIV markers values are commonly used for achieving normality and homoscedasticity of measurement error distributions (Thiébaut et al., 2003).

[Table 1 about here.]

5. Analysis of the ALBI ANRS 070 data

We estimated the parameters of the statistical model described above using repeated measurements of both the virus load and the total CD4 count from the ALBI ANRS 070 data. The model for the observations (3) can be given more explicitly:

$$\begin{cases} Y_{ij1} = log_{10}(V_I(t_{ij1}, \tilde{\boldsymbol{\xi}}^{(i)}) + V_{NI}(t_{ij1}, \tilde{\boldsymbol{\xi}}^{(i)})) + \epsilon_{ij1}, & j \leq n_{i1} \\ Y_{ij2} = (Q(t_{ij2}, \tilde{\boldsymbol{\xi}}^{(i)}) + T(t_{ij2}, \tilde{\boldsymbol{\xi}}^{(i)}) + T^*(t_{ij2}, \tilde{\boldsymbol{\xi}}^{(i)}))^{0.25} + \epsilon_{ij2}, & j \leq n_{i2} \end{cases}$$

We tried different starting values (increasing the variance parameters values by 10 folds, intercept value from 10 % to 100 %) and we obtained the same convergence point up to small variations on the third signifcant digit. The value for $\tilde{\gamma}$ obtained by profile likelihood was $\tilde{\gamma} = -3$. The estimates of the model parameters are shown in Table 2.

[Table 2 about here.]

The empirical correlation between residuals R_{ij1} and R_{ij2} where $R_{ij1} = Y_{ij1} - \hat{Y}_{ij1}$ and $R_{ij2} = Y_{ij2} - \hat{Y}_{ij2}$ was equal to 0.017; this is an indication that the proposed mechanistic modelling captures most of the dependency

14

between the observed biomarkers. With the exception of the estimation of α that was about ten fold higher compared to some values found in the literature (Ribeiro et al., 2002), the estimated values of the other parameters were in the same range as published values. For example, the confidence interval for μ_T ($[0.097; 0.14]$) was close to that reported by Ribeiro et al. (2002) ($[0.040; 0.13]$) as well as the confidence interval for μ_{T*} ($[0.5; 0.8]$) to that reported by Markowitz et al. (2003) ($[0.6; 1.4]$).

The contrast β between the two antiretroviral regimen efficacies was tested by a Wald test. It was found significantly different from zero with a stable p-value over the plausible interval for $\tilde{\gamma}$ ($p < 10^{-6}$). The proportion of cells not infected because of the treatment per each unit of time was $\eta_0 = 72.3\%$ ($IC_{95} = [69.1; 75.5]$) in the AZT+3TC group and the difference between between the two groups was 2.0 % ($IC_{95} = [1.4; 2.6]$ calculated by the Delta-method). Interestingly, patients in the AZT+3TC group experienced a similar first decline in HIV RNA level compared to d4T+ddI group. However, after a period of about one month, the HIV RNA level increased, leading to a rebound in the former group. The fit of the model predictions was good in average (as shown in Figure 3) and at individual level (Figure 4). In Figure 3, from the second month, mean predicted values of virus load are lower than the average of the observations; this is because the predictions take censoring into account while the observations below the level of detection ($1.7 log_{10}$) were fixed at this threshold. Concerning the meaning of retained random effects, σ_λ might represent the individual difference in thymic or hematopoietic function. Variation of activation rates between individuals that led to significant σ_α might reflect differences in susceptibility

of activation between individuals. The individual variation of μ_{T*} might reflect the variation in the intensity of the immune response (e.g. cytotoxic lymphocytes that kill infected cells) according to subjects.

[Figure 3 about here.]

[Figure 4 about here.]

6.   Simulation study

In this section, we aim at analyzing the efficiency of our algorithm in term of success of convergence, precision and validity of estimators. Data were simulated using values from the analysis of the ALBI ANRS 070 trial presented in the former section. $\tilde{\gamma}$ was held fixed at the value $\tilde{\gamma} = -3$. We simulated trials of 100 patients, according to the typical schedule of the ANRS ALBI 070 trial. For each parameter, absolute bias, confidence interval coverage, empirical standard deviations and estimated standard deviations were calculated. Of note, the absolute bias of log transformed parameters can be interpreted as a relative bias of parameters on natural scale. The initial values were selected as follows: $\phi_l = \phi_l^*, \beta = 0, \boldsymbol{A} = \sqrt{10}\boldsymbol{A}^*, \boldsymbol{\sigma} = 2\boldsymbol{\sigma}^*$ where $*$ refers to values in Table 2. The maximum number of iterations was fixed at 25 for restricting the duration of the simulation study. Convergence before 25 iterations was successfully reached in 93% of simulations, and the computation time was between two and three hours per simulation. In the simulations where convergence was not achieved, convergence criteria value indicated that the algorithm might converge but with further iterations. Results from 100 successful convergences are summarized in Table 8. The biases were small in regard of the complexity of the model; the coverage rates of

16

the confidence intervals were good and the variances were well estimated. Although the results of this simulation study show that the algorithm is efficient when the model is well specified, this does not warrant that it will be so for real data. Furthermore, more general conclusions on the robustness and efficiency of the algorithm would need an extension of the simulation study over a larger set of plausible values for $\boldsymbol{\theta}$.

[Table 3 about here.]

7. Discussion

In this paper, we presented a robust method to fit non-linear mixed effects models based on differential equations for which no closed-form is available. This approach may be applied to any NLME models. Nevertheless, the method requires some computational skills whereas simpler methods are often sufficient for usual NLME models (with closed-form). The approach was applied to a clinical trial in HIV infection. The model fitted the HIV RNA and CD4 data quite well, providing an *in vivo* estimation of the treatment efficacy. Because of the non-linear interaction between CD4 and virus, the HIV RNA dynamics was very sensitive to a difference in treatment efficacy. Interestingly, the model predicted a viral rebound in the worst treatment group after an initial steep decline although occurrence of HIV resistance was not included in the model. In the present model, the viral rebound can be explained by the joint dynamics of the virus and the CD4. During the first period, the reduction of the rate of infection led to a decrease of virus and an increase of the target cells. However, because the inhibition of the virus infectivity was not strong enough in the AZT+3TC group and because new target cells were available, this led to a rebound of viral load

17

after about one month. Only biomathematical models based on system of differential equations can fit such complex interaction between virus and immune system. The present approach allows to estimate the treatment effect using all available information (including biological knowledge to construct the model) in a much more flexible way than classic multivariate longitudinal models (Thiébaut et al., 2003). Then, the treatment efficacy can be tested using only one statistical test compared to all potential comparisons according to each marker at each given time.

One of the limitations of the present application is the impossibility of estimating all parameters. In fact, the estimation of all parameters would need more intensive schedules and/or the measurements of more compartments such as activated CD4. However, a sensitivity analysis allowed us to conclude that the estimate of the difference in treatment effects between the two groups was robust.

We conclude that the use of such models to analyze clinical trial data could help in having a better understanding of the treatment effect; however, estimating all model parameters requires richer information than usual and this needs to be planned in the study protocol.

8. Supplementary Materials

Web Appendices referenced in Section 3 are available under the Paper Information link at the Biometrics website http://www.tibs.org/biometrics.

18

References

Chun, T. W., Carruth, L., Finzi, D., Shen, X., DiGiuseppe, J. A., Taylor, H., Hermankova, M., Chadwick, K., Margolick, J., Quinn, T. C., Kuo, Y. H., Brookmeyer, R., Zeiger, M. A., Barditch-Crovo, P. and Siliciano, R. F. (1997). Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* **387**, 183–188.

Commenges, D. and Rondeau, V. (2006). Relationship between derivatives of the observed and full loglikelihoods and application to Newton-Raphson algorithm. *International Journal of Biostatistics* **2**, http://www.bepress.com/ijb/vol2/iss1/4/.

Ding, A. and Wu, H. (2001). Assessing antiviral potency of anti-HIV therapies in vivo by comparing viral decay rates in viral dynamic models . *Biostatistics* **2**, 13–29.

Genz, A. and Keister, B. D. (1996). Fully symmetric Interpolatory Rules for Multiple Integrals over Infinite Regions with Gaussian Weight. *Journal of Computational and Applied Mathematics* **71**, 299–311.

Gilks, W., Richardson, S. and Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.

Grossman, Z., Meier-Schellersheim, M., Sousa, A. E., Victorino, R. M. M. and Paul, W. E. (2000). CD4+ T-cell depletion in HIV infection: Are we closer to understanding the cause? *Nature Medicine* **8**, 319–323.

Hedeker, D. and Gibbons, R. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics* **50**, 933–944.

Ho, D. D., Neumann, A. U., Perelson, A. S., Chen, W., Leonard, J. M. and Markowitz, M. (1995). Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **373**, 123–126.

Huang, Y., Liu, D. and Wu, H. (2006). Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. *Biometrics* **63**, 413–423.

Louis, T. (1982). Finding the observed Information matrix when using the *EM* algorithm. *Journal of the Royal Statistical Society. Series B* **44**, 226–233.

Markowitz, M., Louie, M., Hurley, A., Sun, E., Mascio, M. D., Perelson, A. S. and Ho, D. D. (2003). A novel antiviral intervention results in more accurate assessment of human immunodeficiency virus type 1 replication dynamics and T-cell decay in vivo. *Journal of Virology* **77**, 5037–5038.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics* **11**, 431–441.

Mclean, A. R. and Michie, C. A. (1995). *in vivo* estimates of division and death rates of human t lymphocytes. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 3707–3711.

Molina, J., Chene, G., Ferchal, F., Journot, V., Pellegrin, I., Sombardier, M. N., Rancinan, C., Cotte, L., Madelaine, I., Debord, T. and Decazes, J. M. (1999). The ALBI trial: A randomized controlled trial comparing stavudine plus didanosine with zidovudine plus lamivudine and a regimen

alternating both combinations in previously untreated patients infected with human immunodeficiency virus. *The Journal of Infectious Diseases* **180**, 351–358.

Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M. and Ho, D. (1996). Viral dynamics in human immunodeficiency virus type 1 infection. *Science* **271**, 1582–1586.

Piatak, M., Saag, M., Yang, L. C., Clark, S. J., Kappes, J. C., Luk, K. C., Hahn, B. H., Shaw, G. M. and Lifson, J. D. (1993). High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. *Science* **259**, 1749–1754.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS.* Springer, London.

Putter, H., Heisterkamp, S. H., Lange, J. M. A. and deWolf, F. (2002). A Bayesian approach to parameter estimation in HIV dynamic models. *Statistics in Medicine* **21**, 2199–2214.

Ramratnam, B., Bonhoeffer, S., Binley, J., Hurley, A., Zhang, L., Mittler, J. E., Markowitz, M., Moore, J. P., Perelson, A. S. and Ho, D. D. (1999). Rapid production and clearance of HIV-1 and hepatitis C virus assessed by large volume plasma apheresis. *The Lancet* **354**, 1782–1786.

Ribeiro, R. M., Mohri, H., Ho, D. D. and Perelson, A. S. (2002). In vivo dynamics of T cell activation, proliferation, and death in HIV-1 infection: Why are CD4 but not CD8 T cells depleted? *Proceedings of the National Academy of Sciences* **24**, 15572–15577.

Thiébaut, R., Guedj, J., Jacqmin-Gadda, H., Chêne, G., Trimoulet, P., Neau, D. and Commenges, D. (2006). Estimation of dynamical model parame-

ters taking into account undetectable marker values. *BMC Medical Research Methodology* **6**, 1–10.

Thiébaut, R., Jacqmin-Gadda, H., Leport, C., Katlama, C., D., C., Le Moing, V., Morlat, P., Chene, G. and the APROCO Study Group (2003). Bivariate longitudinal model for the analysis of the evolution of HIV RNA and CD4 cell count in HIV infection taking into account left censoring of HIV RNA measures. *Journal of Biopharmaceutical Statistics* **13**, 271–282.

Wu, H. and Ding, A. A. (1999). Population HIV-1 dynamics in vivo: applicable models and inferential tools for virological data from AIDS clinical trials. *Biometrics* **55**, 410–418.
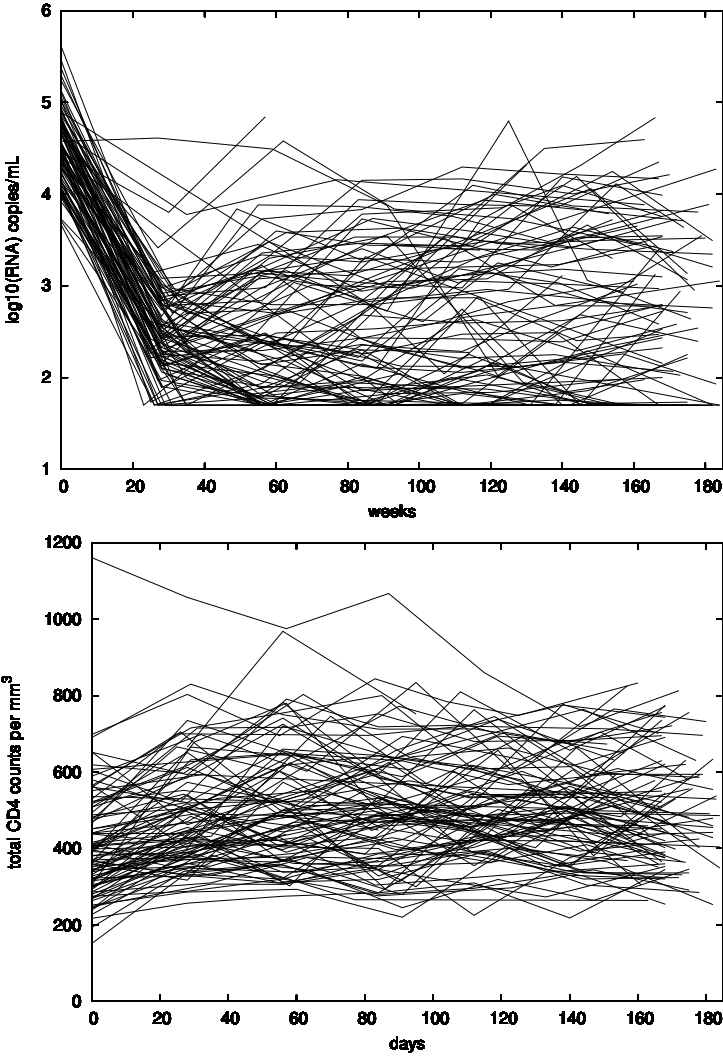
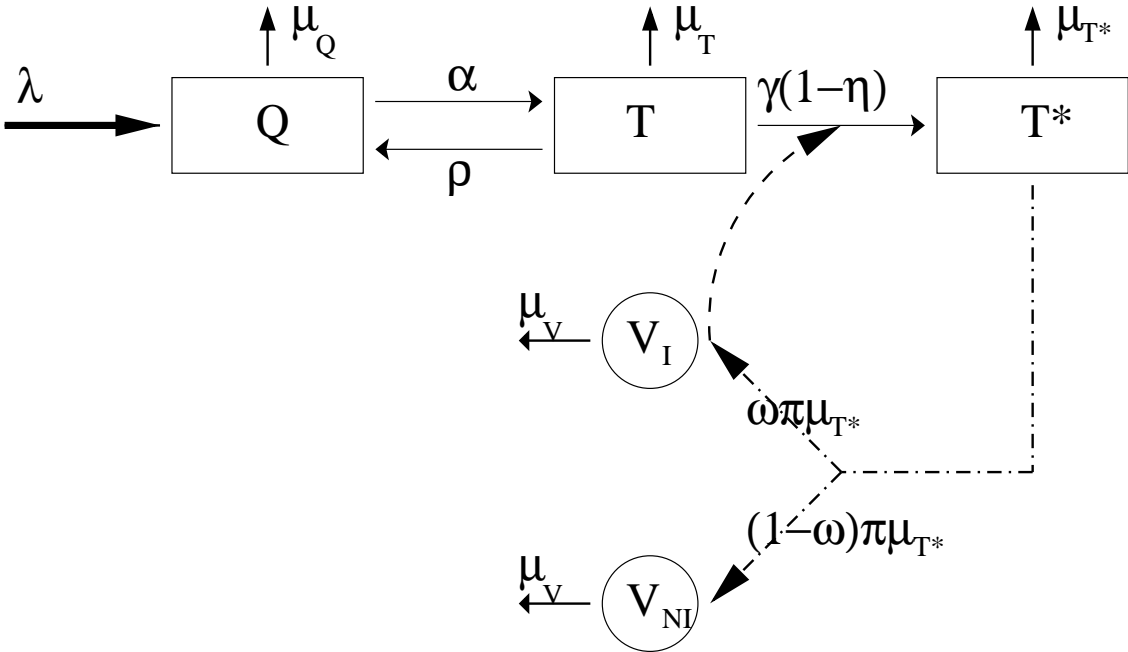**Figure 1.** Spaghetti plots of the data from ALBI ANRS 070 clinical trial.

23

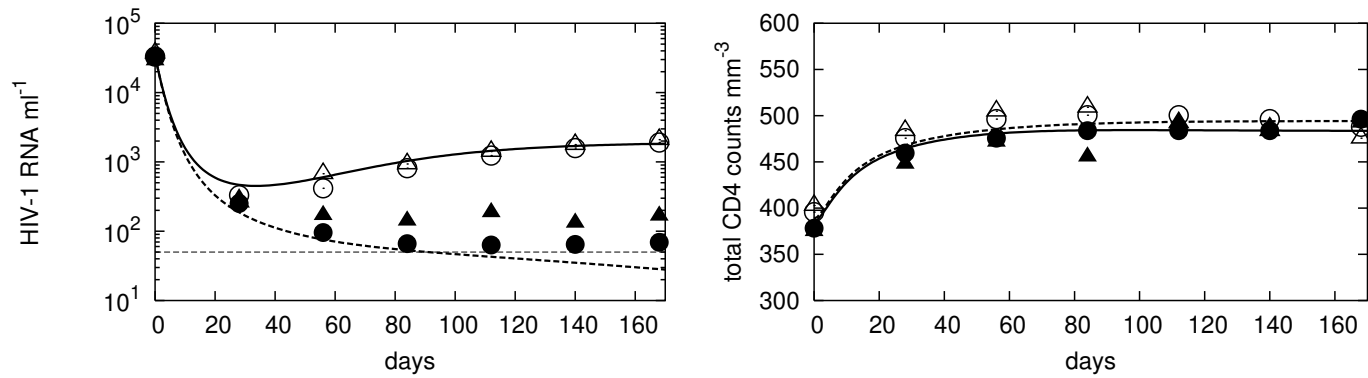**Figure 2.** Graphical representation of the system for HIV dynamics.

24

Figure 3: Mean observed values of HIV RNA level (left) and total CD4 count (right) according to the treatment group ($\triangle$ and $\blacktriangle$ for the group AZT+3TC and ddI+D4T respectively) and mean predicted values ($\circ$ and $\bullet$ for the group AZT+3TC and ddI+D4T respectively) in ALBI ANRS 070 trial. The plain line represents the expected trajectory for the group ddI+D4T whereas the dashed line represents the expected trajectory for the group AZT+3TC.
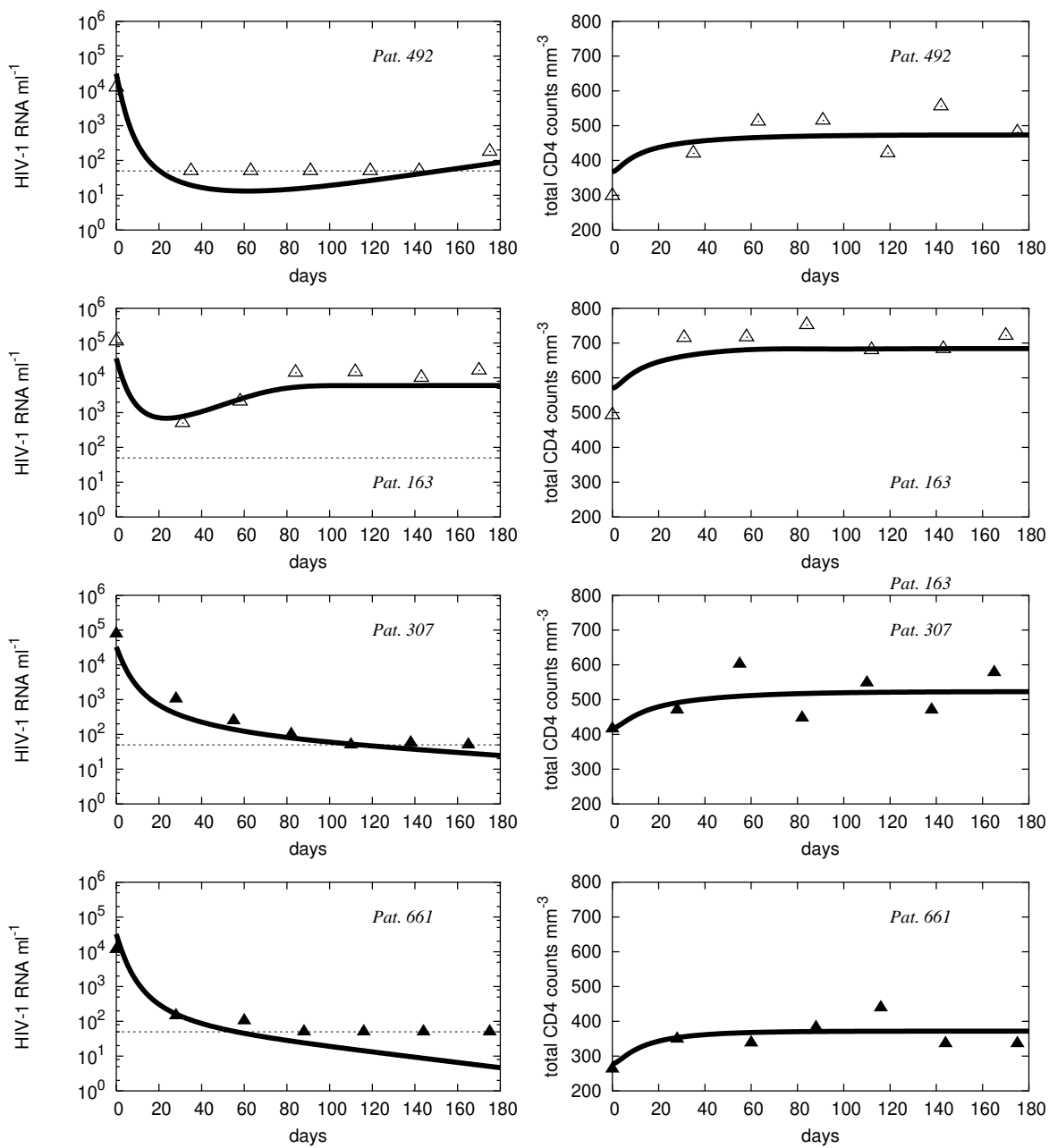
Figure 4. Individual predicted fit and observed values of HIV RNA level (left) and total CD4 count (right) according to the treatment group ($\triangle$ and $\blacktriangle$ for the group AZT+3TC and ddI+D4T respectively) in four patients.

Table 1: Natural model parameters and values for which were fixed.

Parameter	Meaning	Value	Reference
α	Activation rate of Q cells (day^{-1})		
λ	Rate of Q cells production $(\mu_L^{-1} day^{-1})$		
μ_{T*}	Death rate of T^* cells (day^{-1})		
π	Number of virions per T^* cell		
μ_T	Death rate of T cells (day^{-1})		
η	Efficacy of treatment (proportion)		
σ_{CD4}	Standard deviation of the observation error of $(Q + T + T^*)^{0.25}$		
σ_{CV}	Standard deviation of the observation error of $log_{10}(V_I + V_{NI})$		
γ	Infection rate of T cells per virion		
μ_Q	Death rate of Q cells	0.00014	Mclean and Michie, 1995
μ_v	Clearance of free virions	30.0	Ramratnam et al., 1999
ρ	Rate of reversion to the quiescent state	0.017	Ribeiro et al., 2002
ω	Proportion of non-infectious virions	0.20	Piatak et al., 1993

Table 2

Estimates of the model parameters and their standard deviation. ALBI ANRS 070 clinical trial.

Parameters	Estimated Value	STD
$\tilde{\alpha}$	-3.16	0.15
$\tilde{\lambda}$	2.62	0.12
$\tilde{\mu_{T*}}$	-0.40	0.11
$\tilde{\pi}$	4.64	0.12
$\tilde{\mu_T}$	-2.14	0.087
$\tilde{\eta}$	0.96	0.079
β	0.096	0.018
σ_α	0.31	0.025
σ_λ	0.043	0.0059
$\sigma_{\mu_{T*}}$	0.25	0.028
σ_{CV}	0.42	0.012
σ_{CD4}	0.18	0.0050

Table 3

*Absolute bias, 95% confidence interval coverage, mean of estimated standard
deviations (STD) and empirical STD (STD of parameter estimations) of the
estimators of model parameters for 100 simulated data sets of 100 patients.*

Parameter	True Value	Absolute Bias	Coverage (%)	STD of Estimates	Mean of STD
$\tilde{\alpha}$	-3.16	0.0015	98	0.12	0.14
$\tilde{\lambda}$	2.62	-0.00070	99	0.098	0.11
$\mu_{\tilde{T*}}$	-0.40	0.0078	94	0.10	0.095
$\tilde{\pi}$	4.64	0.014	99	0.11	0.12
$\tilde{\mu_T}$	-2.14	0.0062	93	0.082	0.081
$\tilde{\eta}$	0.96	0.0029	94	0.11	0.12
β	0.096	0.0098	92	0.019	0.016
σ_α	0.31	-0.0036	93	0.025	0.026
σ_λ	0.043	0.0016	90	0.0067	0.0057
$\sigma_{\mu_{T*}}$	0.25	0.0026	97	0.037	0.039
σ_{CV}	0.42	-0.0043	93	0.013	0.014
σ_{CD4}	0.18	-0.00014	92	0.0056	0.0053