Statistics and Its Interface Volume 6 (2013) 9–18

Parameter estimation for HIV ODE models incorporating longitudinal structure

Yao Yu and Hua Liang*

We apply nonlinear mixed-effects models (NLME) to estimate parameters in Perelson's HIV dynamic model, a system of mechanism-based ordinary differential equations (ODE). The unknown parameters of the dynamic model and the baseline of infected CD4$^+$ T cells are estimated simultaneously. Meanwhile variance components for random-effects and parameters for individuals are also estimated. Because we solve the ODE directly without making any model approximations or fixing any parameters to obtain close-form solutions as in literature, the drawing conclusion maintains biological interpretability for dynamic parameters which is critically helpful. Simulation studies are conducted to examine the performance of this approach, especially the influence of measurement errors and model assumptions underlying the parameter estimation method. Moreover, we apply this approach to real data collected from an AIDS clinical trial of HIV-1 study.

Keywords and phrases: HIV dynamic model, Measurement errors, Nonlinear mixed-effects model, Variance components, Random-effects.

1. INTRODUCTION

Over the past two decades, ordinary differential equation (ODE) models have been used to examine HIV viral dynamics for the study of pathogenesis and the development of new treatment strategies for AIDS clinical trials and patient cares. As mechanism-based models, HIV dynamic models are based on the process of HIV inflection on CD4$^+$ T cells [1–6]. Thus all parameters have biological meanings and can effectively interpret the infection progress and treatment effects. Therefore a variety of statistical methods have been proposed to estimate these parameters with the aim to guide clinical decisions for individual treatment. For example, [2] employed nonlinear least-square (NLS) regression to fit data from each patient separately and obtained parameter estimators for individuals, although NLS can't account for the between-patient variance and convergence problems may be encountered when observations for each individual subject are sparse.

*Corresponding author.

In order to efficiently use between-subject information, mixed-effects models were introduced for parameter estimation in the last two decades, for example, [6–8] for linear mixed-effect models (LME) and nonlinear mixed-effects models (NLME), and [9] for semiparametric nonlinear mixed-effects models (SNLME). [10] did a comparison of LME, NLME, SNLME and used them to study the relationship between the viral load baseline and viral decay rates. Moreover, the stochastic approximation expectation-maximization (SAEM) algorithm, proposed by [11], was employed to approximate the likelihood for NLME with the purpose of estimating parameters in HIV dynamic systems. [12] applied the SAEM algorithm to analyze left-censored longitudinal HIV data. [13] applied the SAEM algorithm to estimate population parameters for pharmacokinetic-pharmacodynamic-viral dynamic models. [14] employed the SAEM algorithm to estimate parameters for long-term HIV dynamic systems. [15] used a viral dynamics nonlinear mixed effects drug-disease model in NONMEM to fit HIV-RNA data. It is interesting that different models result in different conclusions. It is also worthwhile to mention that owing to the complexity of ODE , too many unknown parameters and the existence of unobservable state variables, reparameterization, approximation and simplification are utilized to estimate parameters in HIV dynamic models. Consequently, the parameters in the simplified mixed-effects models may not have the same biological meanings as in the original dynamic models.

To avoid approximating dynamic models and repeatedly solving ODE, [16] developed a nonparametric method based on local smoothing and the pseudo-least square principles to estimate unknown parameters for general ODE models. The method avoids the concerns we just commented on and has some advantages such as computational efficiency and easing the convergency problem. It has some limitations, namely, estimation errors are large and frequent measurements of state variables are required to estimate derivatives. In order to incorporate information from various subjects for better estimated values, [17] extended the method proposed in [16] to longitudinal data by adopting the mixed-effects modeling approach. However, their approach still inherits the limitations of [16].

When frequentists gained fruitful findings in HIV dynamic parameter estimation, the Bayesian principle was widely used in HIV dynamic parameter estimation also. [18]

introduced the Bayesian modeling approach to nonlinear random effects estimation problems. [19, 20] estimated constant coefficients and time-varying dynamic parameters for long-term HIV viral dynamic models. [21] incorporated clinical factors and baseline covariates into Bayesian models to study treatment efficacy in long-term HIV viral dynamics. It should be noted that all these Bayesian methods require prior information for most of the parameters, which may not be easy to obtain.

Facing the complexity of ODE models, the unobservable features of some state variables and the sparsity of experimental and clinical data for each individual, parameter estimation for HIV dynamic models is still challenging. In this paper, we introduce a novel application of NLME to estimate parameters in Perelson's HIV dynamic model [22], a dynamic system which has been widely used to enhance our understanding of immunological process against HIV infection [2, 23–25]. We solve Perelson's HIV dynamic model directly via Fortran ODE solver Isoda (livermore solver for ordinary differential equations, with automatic algorithm selection) [26] to obtain numerical solutions. Then we fit NLME to real data from a HIV clinical trial to estimate population and individual parameters. Using the estimated values, we design Monte Carlo simulation studies to assess the performance of this approach.

The rest of the paper is organized as follows. In Section 2, we introduce Perelson's HIV dynamic model and briefly describe the estimation method based on NLME. In Section 3, we analyze a real data set and estimate the fixed-effects and random-effects of dynamic parameters, and the baseline of infected CD4$^+$ T cells. In Section 4, we conduct Monte Carlo simulation studies to evaluate the performance of the approach. The paper ends with a discussion in Section 5.

2. MODELS AND ESTIMATION METHODS

2.1 The ODE model

Consider the following HIV dynamic model after antiretroviral therapy [22, 25]:

$$\begin{aligned}\frac{dT_U(t)}{dt} &= \lambda - \rho T_U(t) - \eta(t) T_U(t) V(t),\\ \frac{dT_I(t)}{dt} &= \eta(t) T_U(t) V(t) - \delta T_I(t),\\ \frac{dV(t)}{dt} &= N\delta T_I(t) - cV(t),\end{aligned} \tag{1}$$

where $T_U(t)$ is the concentration of uninfected CD4$^+$ T cells, $T_I(t)$ is the concentration of infected CD4$^+$ T cells, $V(t)$ is the plasma virus concentration, λ is the rate at which new CD4$^+$ T cells are continuously generated, ρ is the death rate of uninfected CD4$^+$ T cells, δ is the death rate of the infected CD4$^+$ T cells, c is the clearance rate of free virions, N is the number of virions produced from each infected cell and $\eta(t)$ is the time varying infection rate of T cells which depends on the antiviral drug efficacy. When the time range is small, $\eta(t)$ can be treated as a constant parameter η. In this system, $T_U(t), T_I(t)$ and $V(t)$ are state variables, and $(\lambda, \rho, \delta, c, N, \eta)$ are unknown parameters.

In AIDS clinical studies, generally only $V(t)$ and the total CD4$^+$ T cell counts, $T(t) = T_I(t) + T_U(t)$, are measured for cost concerns. By combining the first and second differential equations in the system (1), we remove $\eta(t)T_U(t)V(t)$. As a result, we obtain a simplified differential equation system (2),

$$\begin{aligned}\frac{dT(t)}{dt} &= \lambda - \rho\{T(t) - T_I(t)\} - \delta T_I(t),\\ \frac{dT_I(t)}{dt} &= \eta\{T(t) - T_I(t)\}V(t) - \delta T_I(t),\\ \frac{dV(t)}{dt} &= N\delta T_I(t) - cV(t).\end{aligned} \tag{2}$$

To obtain numerical solutions for $T_I(t), T(t)$ and $V(t)$, we apply the Fortran ODE solver lsoda to the first-order ODE system (2). Theoretically the numerical solution can be written as follows:

$$T(t) = g_1\big(\lambda, \rho, \delta, T_I(t), T(0), t\big), \tag{3}$$

$$T_I(t) = g_2\big(\eta, \delta, T(t), V(t), T_I(0), t\big), \tag{4}$$

$$V(t) = g_3\big(N, \delta, c, T_I(t), V(0), t\big), \tag{5}$$

where $T(0), T_I(0)$ and $V(0)$ are the baselines of state variables at $t = 0$, the onset time of the drug effect. $g_1(.)$, $g_2(.)$ and $g_3(.)$ are nonlinear functions. For simplicity, we use T_0, T_{I0} and V_0 to represent $T(0), T_I(0)$ and $V(0)$ in the rest of the paper.

By plugging (4) in (3) and (5) respectively, we re-express nonlinear functions (3) and (5) as follows:

$$T(t) = \tilde{g}_1(\lambda, \rho, \delta, \eta, N, c, T_{I0}, T_0, V_0, t), \tag{6}$$

$$V(t) = \tilde{g}_2(\lambda, \rho, \delta, \eta, N, c, T_{I0}, T_0, V_0, t). \tag{7}$$

Write $Y = \big(T(t), V(t)\big)^{\mathrm{T}}$, $Y_0 = \big(T_0, V_0\big)^{\mathrm{T}}$,$\theta = (\lambda, \rho, \delta, N, c, \eta, T_{I0})^{\mathrm{T}}$ and $G = (\tilde{g}_1, \tilde{g}_2)^{\mathrm{T}}$. Functions (6), (7) can be represented as

$$Y = G(\theta, Y_0, t), \quad t \geqslant 0. \tag{8}$$

In equation (8), the baselines Y_0 and the state variables Y are measurable. $\theta = (\lambda, \rho, \delta, N, c, \eta, T_{I0})^{\mathrm{T}}$ are unknown parameters and need to be estimated.

2.2 The mixed-effects model

Many AIDS clinical trial studies involve repeated observations from the same patient over a given period. Note that $G(.)$ in (8) is a complex function with seven unknown parameters, the estimation of unknown dynamic parameters

may not be satisfactorily attained by using sparse individual observations. Here we estimate the fixed-effects associated with the population and the random-effects related to the between-subject variations by fitting NLME to observed $T(t)$ and $V(t)$. To stabilize the computational algorithms, θ, T_0 and V_0 were taken logarithm transformation before analysis. The model framework can be described as follows.

Denote the number of subjects by n, the number of observations for i_{th} subject by n_i. Let t_{ij} be the measurement time for i_{th} subject at j_{th} time point; $\mathbf{y}_{ij} = (T_{ij}, V_{ij})^{\mathrm{T}}$ where T_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, n_i$ and V_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, n_i$ are the measurements of CD4$^+$ T cells a nd plasma viral load at time t_{ij} respectively; the parameter vector after logarithm transformation for i_{th} subject is $\phi_i = (\ln(\lambda_i), \ln(\rho_i), \ln(\delta_i), \ln(N_i), \ln(c_i), \ln(\eta_i), \ln(T_{I0i}))^{\mathrm{T}}$.

The function (8) can then be re-written as

$$\mathbf{y}_{ij} = \mathbf{f}_i(\phi_i, \ln(Y_{0i}), t_{ij}), \tag{9}$$

with $\mathbf{f}_i(\phi_i, \ln(Y_{0i}), t_{ij}) = (f_{1i}(\phi_i, \ln(Y_{0i}), t_{ij}), f_{2i}(\phi_i, \ln(Y_{0i}), t_{ij}))$ which satisfy $f_{1i}(\phi_i, \ln(Y_{0i}), t_{ij}) = \tilde{g}_1(\theta_i, Y_{0i}, t_{ij})$ and $f_{2i}(\phi_i, \ln(Y_{0i}), t_{ij}) = \tilde{g}_2(\theta_i, Y_{0i}, t_{ij})$.

Using the nlmeODE function (available in the `nlmeODE` R packages), we obtain the form for function $\mathbf{f}_i(.)$. We then apply the NLME model to estimate the fixed-effects and random-effects for unknown parameters ϕ_i. We briefly describe the model framework of NLME as follows [27–30]):

Stage 1.

$$\mathbf{y}_{ij} = \mathbf{f}_i\big(\phi_i, \ln(Y_{0i}), t_{ij}\big) + \epsilon_{ij}, \tag{10}$$

$\epsilon_{ij} \sim N(0, R_i)$ with $R_i = \sigma^2 I_{n_i}$, I_{n_i} is the n_i identity matrix, σ^2 is an unknown parameter and needs to be estimated.

Stage 2.

$$\phi_i = \phi + b_i, \quad b_i \sim N(0, \Psi), \tag{11}$$

where ϕ is k-dimensional vectors, and k is the number of unknown parameters. The random-effects b_i, $i = 1, \ldots, n$ follow multivariate normal distribution with mean 0 and covariance matrix Ψ. To reduce the number of unknown parameters, we assume Ψ is a diagonal matrix $\sigma_b = \mathrm{diag}(\sigma^2_{b,1}, \ldots, \sigma^2_{b,7})$, which needs to be estimated.

Let $\mathbf{Y}_i = (\mathbf{y}_{i1}, \mathbf{y}_{i2}, \ldots, \mathbf{y}_{in_i})$ be the measurements of CD4$^+$ T cells and plasma viral load for the i_{th} subject, $t_i = (t_{i1}, t_{i2}, \ldots, t_{in_i})$ and $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{in_i})$. Combining equations (10) and (11), we obtain

$$\mathbf{Y}_i = \mathbf{f}_i\big(b_i, \phi, \ln(Y_{0i}), t_i\big) + \varepsilon_i, \quad i = 1, 2, \ldots, n. \tag{12}$$

The nlme algorithm (available in the `nlme` package) was implemented to estimate unknown fixed-effects ϕ, variance components of random-effects $\sigma_{b,k}$, $k = 1, \ldots, 7$ and the standard error of residual σ. The restricted maximum likelihood (REML) is a function of the marginal density of Y. In the population level, the marginal density of Y can be defined as follows:

$$P(Y|\phi, \sigma^2, \sigma_b) = \int P(Y|\phi, \sigma^2, b) P(b|\sigma_b) db, \tag{13}$$

where $P(Y|\phi, \sigma^2, b)P(b|\sigma_b)$ is the joint density function of (Y, b), $P(Y|\phi, \sigma^2, b)$ is the conditional density of Y given the random-effects b and $P(b|\sigma_b)$ is the marginal distribution of b. Since the integral (13) generally has no close-form, the nlme algorithm uses two alternating steps: the penalized nonlinear least square (PNLS) step and the LME step proposed by [27] to approximate REML.

The PNLS step Fix the current estimate σ_b and then estimate the conditional estimator of random-effects b_i's and the conditional estimator of the fixed-effects ϕ based on current σ_b by minimizing the penalized nonlinear least-square objective function:

$$\sum_{i=1}^{n} \left\{ \frac{\|\mathbf{Y}_i - \mathbf{f}_i(b_i, \phi, \ln(Y_{0i}), t_i)\|^2}{\sigma^2} + b_i^{\mathrm{T}} \sigma_b^{-1} b_i \right\}. \tag{14}$$

The LME step Update σ_b based on the estimator of b_i's, ϕ and the first order Taylor expansion of $\mathbf{f}_i(.)$ around the estimator of b_i, ϕ which can be described as follows:

$$\begin{aligned} \mathbf{Y}_i \approx \mathbf{f}_i\big(\hat{\phi}, \hat{b}_i, \ln(Y_{0i}), t_i\big) &+ \left.\frac{\partial \mathbf{f}_i}{\partial \phi^{\mathrm{T}}}\right|_{\hat{\phi}, \hat{b}_i} \big(\phi - \hat{\phi}\big) \\ &+ \left.\frac{\partial \mathbf{f}_i}{\partial b_i^{\mathrm{T}}}\right|_{\hat{\phi}, \hat{b}_i} \big(b_i - \hat{b}_i\big) + \varepsilon_i. \end{aligned} \tag{15}$$

For notational simplicity, we use $\frac{\partial \mathbf{f}_i}{\partial \phi^{\mathrm{T}}}$ and $\frac{\partial \mathbf{f}_i}{\partial b_i^{\mathrm{T}}}$ to represent $\frac{\partial \mathbf{f}_i}{\partial \phi^{\mathrm{T}}}\big|_{\hat{\phi}, \hat{b}_i}$, $\frac{\partial \mathbf{f}_i}{\partial b_i^{\mathrm{T}}}\big|_{\hat{\phi}, \hat{b}_i}$ respectively and the approximate restricted log-likelihood $\log L^r_{LME}(\phi, \sigma^2, \sigma_b)$ for updating σ_b in the LME step can be written as:

$$\begin{aligned} &\log L^r_{LME}(\phi, \sigma^2, \sigma_b) \\ &= -\frac{1}{2} \sum_{i=1}^{M} \log \left\{ \left| \sigma^2 \left(\frac{\partial \mathbf{f}_i}{\partial \phi^{\mathrm{T}}} \right)^{\mathrm{T}} \left[\mathbf{I} + \frac{\partial \mathbf{f}_i}{\partial b_i^{\mathrm{T}}} \sigma_b \left(\frac{\partial \mathbf{f}_i}{\partial b_i^{\mathrm{T}}} \right)^{\mathrm{T}} \right]^{-1} \frac{\partial \mathbf{f}_i}{\partial \phi^{\mathrm{T}}} \right| \right\} \\ &\quad + \log L_{LME}(\phi, \sigma^2, \sigma_b), \end{aligned}$$

where $\mathbf{I}$ is the identity matrix and the approximate log-likelihood function $\log L_{LME}$ can be described as:

$$
\begin{aligned}
&\log L_{LME}(\phi, \sigma^2, \sigma_b) \\
&= -\frac{M}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\sum_{i=1}^{M}\left\{\log\left|\sigma^2\mathbf{I} + \frac{\partial \mathbf{f}_i}{\partial \phi^{\mathrm{T}}}\sigma_b\left(\frac{\partial \mathbf{f}_i}{\partial \phi^{\mathrm{T}}}\right)^{\mathrm{T}}\right|\right. \\
&\quad + \left(y_i - \mathbf{f}_i(\phi, b_i) + \frac{\partial \mathbf{f}_i}{\partial b^{\mathrm{T}}}\hat{b}_i\right)^{\mathrm{T}} \\
&\quad \times \left[\sigma^2\mathbf{I} + \frac{\partial \mathbf{f}_i}{\partial \phi^{\mathrm{T}}}\sigma_b\left(\frac{\partial \mathbf{f}_i}{\partial \phi^{\mathrm{T}}}\right)^{\mathrm{T}}\right]^{-1} \\
&\quad \times \left.\left(y_i - \mathbf{f}_i(\phi, b_i) + \frac{\partial \mathbf{f}_i}{\partial b_i^{\mathrm{T}}}\hat{b}_i\right)\right\},
\end{aligned}
$$

where $M = \sum_{i=1}^{n} n_i$, the total number of observations of n individuals. The algorithm alternates between the PNLS and LME steps until convergence criteria meet. As a consequence, ϕ, σ_b and σ^2 can be obtained. Meanwhile, the random-effects b_i's can be estimated as a by-product of the estimation procedure. The detailed description about this algorithm was reported in [31].

3. ANALYSIS OF GLAXWELL DATA

In this section, we analyzed a real data set from an AIDS clinical trial of HIV-1 study with patients treated with abacavir combined with 5 different protease inhibitors: prenavir (APV), indinavir (IDV), nelfinavir (NFV), ritonavir (RTV), and saquinavir (SAQ). Blood samples were collected from patients for quantitative analysis. Of all the 82 study subjects enrolled in this study, 74 patients accepted antiretroviral therapy. 3 patients who accepted the antiretroviral therapy were excluded from viral dynamic analysis for no confirmed viral decay during 8 weeks of treatment. 4 patients were dropped for being lack of baselines, T_0 and V_0. As a result, 67 subjects were included in this analysis. The measurements of CD4 $^+$ T cells and plasma HIV-1 RNA were collected before the start of study, at day 0, 3 or 4 ,7, and 10 or 11, week 2, 3, 4, 8, 12, 16, 24, 32, 40, 48 and every 8 weeks for up to 2 years. In this paper, we only analyzed the measurements collected from day 0 to day 54. More details about the study can be found in [32].

For the nlme algorithm, choosing start values is critical because improper start values may cause convergence problems, moreover, different start values may result in different estimated values. Here we generated 339 sets of start values based on crude parameter ranges from literature. For example, [33] pointed out that the memory T cells may live 2–6 weeks in the absence of antigen-stimulated replication, while [2] reported the approximate ranges of c and δ. [34] suggested that N is between 50 and 1,000. The summary of the parameter ranges are given in Table 1. The steps used to generate start values are as follows:

Step 1. Multiply the lower bound and upper bound of each interval given in Table 1 by 0.8 and 1.2 respectively and generate 113 sets of $(\lambda, \rho, \delta, N, c, \eta, T_{I0})$ from expanded intervals in an increasing consecutive order.

Table 1. The dynamic parameter ranges used to generate start values

Parameters	Ranges	ln(ranges)
λ	(10, 100)	(2.30, 4.61)
ρ	(0.071, 0.023)	(−2.65, −3.77)
δ	(0.26, 0.68)	(−1.35, −0.39)
N	(50, 1000)	(3.91, 6.91)
c	(2.06, 3.81)	(0.72, 1.34)
η	$(10^{-6}, 10^{-5})$	(−13.82, −11.51)
T_{I0}	(77.0, 192.5)	(4.34, 5.26)

Step 2. Fix the order of $\lambda, \rho, \delta, N, T_{I0}$ generated in step 1 and reverse the order of generated c, η simultaneously to obtain 113 sets of parameters with different combinations.

Step 3. Maintain the order $\lambda, \rho, \delta, \eta, N$ generate in step 1 and reverse the order of generated c and T_{I0} simultaneously, we got another 113 sets of parameters.

Taking natural logarithm transformation of the generated 339 sets of parameters, we obtain 339 sets of start values. Using the modeling approach described in Section 2, we estimate ϕ, σ and $\sigma_{b,k}$'s. Applying AIC and loglikelihood as selection criteria, we choose start values for NLME as follows: $\ln(\lambda) = 3.0910$, $\ln(\rho) = -3.6212$, $\ln(\delta) = -1.2588$, $\ln(N) = 5.8141$, $\ln(C) = 0.6999$, $\ln(\eta) = -10.7245$, and $\ln(T_{I0}) = 4.5164$. Employing the modeling approach described in Section 2 again, we obtain the estimates of the fixed-effects and random-effects. The dynamic parameter estimates for 67 individuals ϕ_i's can be obtained. To save the space, we only report the estimates of fixed-effects ϕ, standard errors of ϕ and $\sigma_{b,k}$'s in Table 2.

The following findings can be observed in Table 2: (i) uninfected CD4$^+$ T cells on average have a lower death rate compared with infected CD4$^+$ T cells which is consistent with previous research. The population estimates of the decay rates of uninfected CD4$^+$ T cells and infected CD4$^+$ T cells are $\hat{\rho} = 0.1396$ and $\hat{\delta} = 0.2136$ which correspond to $\log 2/\hat{\rho} = 4.9652$ and $\log 2/\hat{\delta} = 3.2451$ days' half-lives.; (ii) the clearance rate for free virions is $\hat{c} = 0.7105$ which corresponds a half-life of $\log 2/\hat{c} = 0.9756$ day. The estimated clearance rate is smaller than these from methods based on close form solutions; (iii) the supply rate of CD4$^+$ T cells, the number of free viruses produced by a CD4$^+$ T cell during its lifetime and the unobservable baseline of infected CD4$^+$ T cells for short-term HIV clinical trial study are $\lambda \approx 62$, $N \approx 183$ and $T_{I0} \approx 172$, respectively. In an AIDS clinical trial study, the concentration of infected CD4$^+$ T cells isn't observable. To make dynamic parameters identifiable, some estimation methods assume that for patients at the start of therapy, the concentration of CD4$^+$ T cells and plasma viruses are in the quasi-steady state; the uninfected CD4$^+$ T cells equal the observed CD4$^+$ T cells. Our result shows that the population mean of infected CD4$^+$ T

Table 2. The estimated values of fixed-effects and variance components of random-effects for Perelson's HIV dynamic models

	$\ln(\lambda)$	$\ln(\rho)$	$\ln(\delta)$	$\ln(N)$	$\ln(c)$	$\ln(\eta)$	$\ln(T_{I0})$
est	4.1271	−1.9687	−1.5439	5.2092	−0.3418	−12.6141	5.1443
se	0.1273	0.1415	0.1659	0.2621	0.0794	0.3149	0.4614
$\sigma_{b,k}$'s	0.3040	0.2780	0.5355	0.9146	0.3633	0.7680	1.1918
	λ	ρ	δ	N	c	η	T_{I0}
Exp(est)	62	0.1396	0.2136	183	0.7105	3.3249e-06	172

cell concentration $T_{I0} \approx 172$ which indicate that the proportion of infected CD4$^+$ T cells isn't ignorable. Moreover, although the change rates of CD4$^+$ T cells and plasma viral load are small, $\frac{dT_U(t)}{dt}$, $\frac{dT_I(t)}{dt}$ and $\frac{dV(t)}{dt}$ are not equal to 0 at the start time of therapy The estimated $\hat{\rho}$ and $\hat{c}$ are out of the ranges given in Table 1, but they are close to the estimated death rate of infected CD4$^+$ T cells and the clearance rate for free virions reported in [16]. One interesting point of this approach is that the rate of converting uninfected CD4$^+$ T cells to infected CD4$^+$ T cells η can be obtained. By fitting data collected before and after antiretroviral therapy with the approach described in this paper separately and comparing estimated η, evaluating drug efficiency becomes possible. In addition, the variance components $\sigma_{b,k}$'s show significant between-patient variations. Therefore, patients should be treated individually.

The curves of plasma viral load $V(t)$ and total CD4$^+$ T cell counts $T(t)$ for eight patients are shown in Figs 1 and 2, respectively. One can see that the individual curves (dashed lines) are closer to the observed values (dots) than the associated population curves (solid lines). The fitted population curves are defined by $Y(t_i) = f_i(\phi, \ln(Y_{0i}), t_i)$ where $Y(t_i) = (T(t_i), V(t_i))$, ϕ is the fix-effect, $t_i = (t_{i1}, t_{i2}, \ldots, t_{in_i})$ and $\ln(Y_{0i}) = (\ln(T_{0i}), \ln(V_{0i}))$ (the baselines of CD4$^+$ T cells and plasma viral load after logarithm transformation). It is worthwhile to mention that the population curves can be distinctive because $\ln(Y_{0i})$ may be different for each subject.

Although the between-subjects variation is obvious, the plasma virus concentration $V(t)$ for all patients share a similar pattern. After initiating antiretroviral therapy, the number of viruses falls to a low level, and then because of the drug resistance, inadequate drug concentration in certain tissues and the persistence of long-lived viruses, the decline rate of virus concentration tends to become slow and reaches a stable period during which the viral load changes little. Meanwhile the CD4$^+$ T cell counts T(t) rise steadily. The change trends of $V(t)$ and $T(t)$ are similar to the typical dynamic feature for HIV-infected patients reported in [3, 22–24]. Model assumption verifications are conducted by examining standard diagnostic plots: the Q-Q plots, and the observed values vs. fitted values. Figure 3 indicates that λ, ρ, δ and T_{I0} are normally distributed and that there is no serious violation to the normal assumption. Meanwhile Fig. 4 shows a good agreement between the observed measurements and predicted values.

4. SIMULATION STUDIES

In this section, we conduct Monte Carlo simulation studies to examine the performance of the modeling approach. Considering the fact that observations for individuals may be sparse, we vary the sample size n_i=15, 21 and 41, to examine the performance of the estimators. In AIDS clinical studies, because of measurement errors and the natural variation of state variables, observed state variables are very noisy with large variation. We further examine the performance of this approach by assuming measurements are collected with 10% measurement errors. In Section 2.2, in order to reduce the number of parameters for variance components, we assume Ψ is diagonal. Herein, we choose two different Ψ's for our simulation studies to examine the influence of this assumption. The detail of simulation studies is given as follows.

We generate parameters $(\phi_i, \ln(T_{0i}), \ln(V_{0i}))^{\mathrm{T}}$, $i = 1, \ldots, 12$ from multivariate normal distribution $N_9(\mu, \Sigma)$, with $\mu = (4.1271, -1.9687, -1.5439, 5.2092, -0.3418, -12.6141, 5.1443, 5.8665, 10.9956)$ and $\Sigma = \Sigma_1$ or $\Sigma = \Sigma_2$ respectively. Σ_1 is a diagonal matrix with diag $= (0.0428, 0.0215, 0.0370, 0.3788, 0.0745, 0.2804, 0.8843, 0.1963, 1.9312)$ and

$$\Sigma_2 = \begin{pmatrix} 0.0428 & -0.0232 & -0.0114 & -0.0035 & -0.0094 & -0.0537 & -0.0016 & 0.0531 & 0.0742 \\ -0.0232 & 0.0215 & 0.0002 & 0.0095 & -0.0001 & 0.0082 & 0.0215 & -0.0462 & -0.0035 \\ -0.0114 & 0.0002 & 0.0370 & 0.0581 & 0.0180 & -0.0024 & 0.0377 & -0.0107 & -0.0015 \\ -0.0035 & 0.0095 & 0.0581 & 0.3788 & 0.0655 & -0.0674 & 0.0104 & -0.0812 & 0.2711 \\ -0.0094 & -0.0001 & 0.0180 & 0.0655 & 0.0745 & 0.0489 & -0.0548 & 0.0098 & -0.0877 \\ -0.0537 & 0.0082 & -0.0024 & -0.0674 & 0.0489 & 0.2804 & -0.0011 & -0.0432 & -0.2998 \\ -0.0016 & 0.0215 & 0.0377 & 0.0104 & -0.0548 & -0.0011 & 0.8843 & -0.1762 & 0.4209 \\ 0.0531 & -0.0462 & -0.0107 & -0.0812 & 0.0098 & -0.0432 & -0.1762 & 0.1963 & -0.1140 \\ 0.0742 & -0.0035 & -0.0015 & 0.2711 & -0.0877 & -0.2998 & 0.4209 & -0.1140 & 1.9312 \end{pmatrix}$$

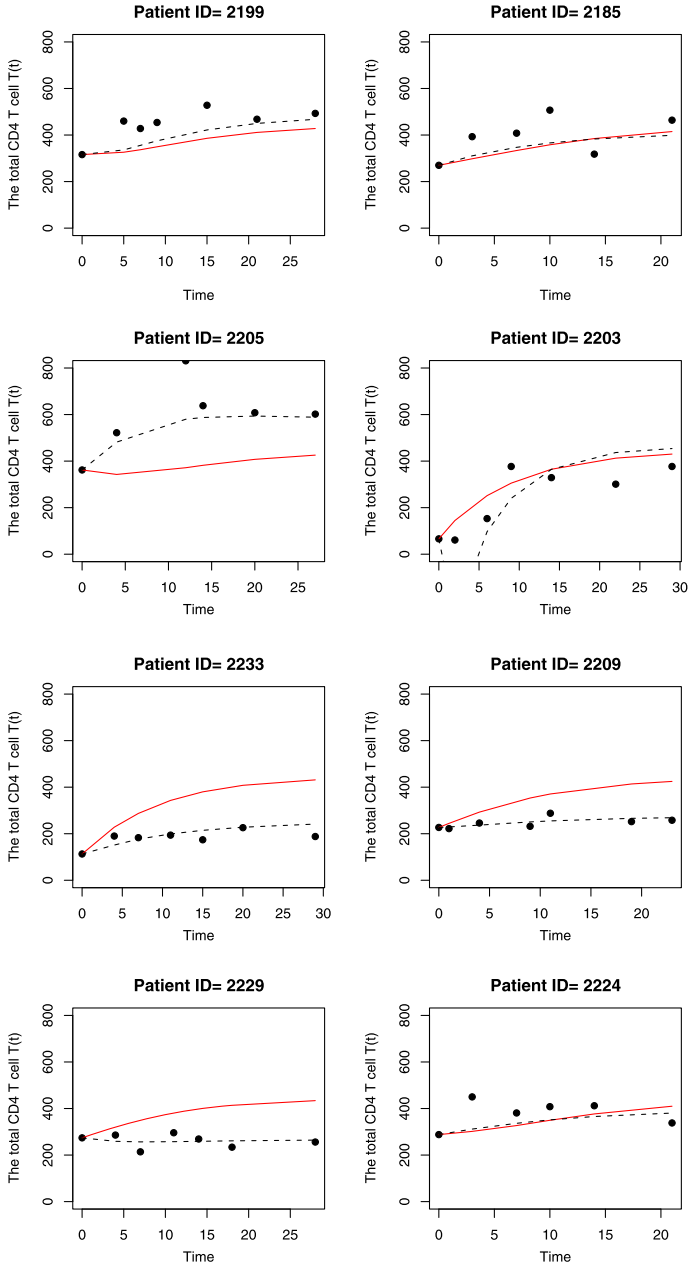

Figure 1. The observed (dots) and fitted population (solid) and individual (broken) curves of CD4 cell counts for 8 patients.

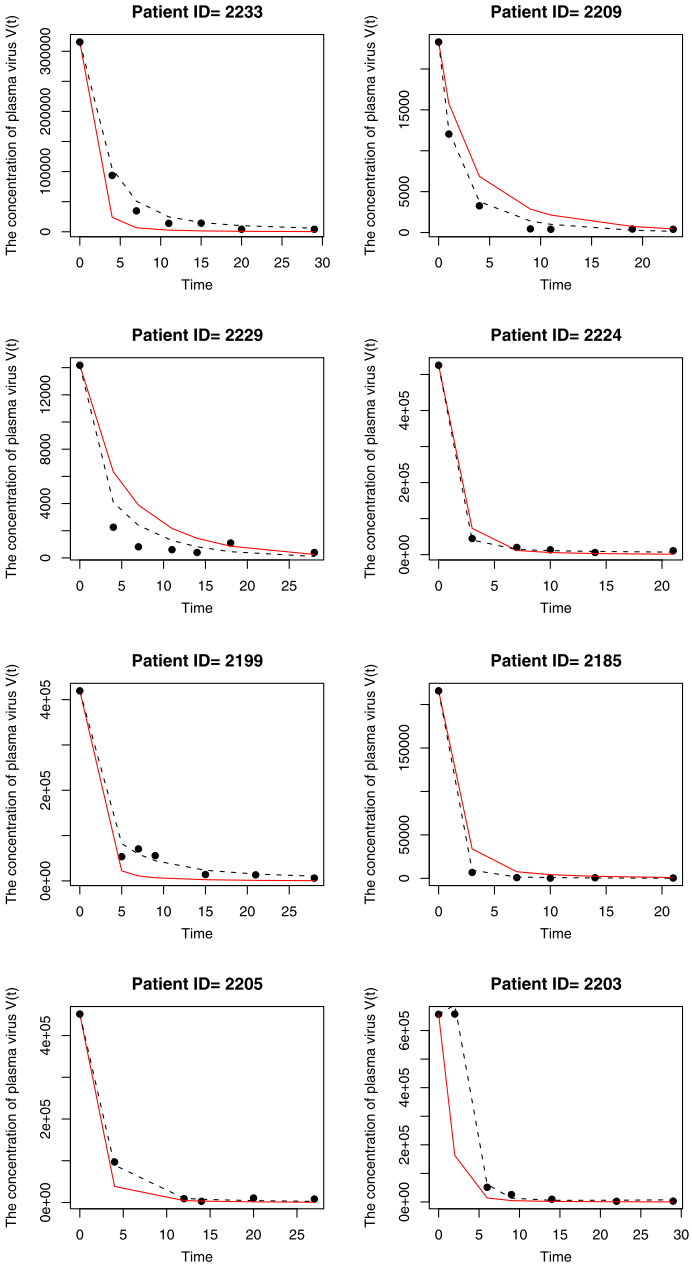

Figure 2. The observed (dots) and fitted population (solid) and individual (broken) curves of viral load for 8 patients.

Here, the mean and the variance-covariance matrix used for generating parameters are obtained from the estimates and observed baseline T_0 and V_0 in Section 3. By applying equations (6) and three outputting schedules: (1) $t = 0, 3, 4, 6, 7, 8, 9, 10, 11, 12, 14, 21, 27, 28, 29$ (the schedule is similar to that of the real data in Section 3, (2) at every 1 unit on the interval $[0, 20]$, (3) at every 0.5 unit on the interval $[0, 20]$, we generate $T^*(t_{ij})$ and $V^*(t_{ij})$ (the generated state variables without measurement error), with sample size $n_i = 15$, 21 and 41 respectively. We then add 10% measurement errors to $T^*(t_{ij})$ and $V^*(t_{ij})$:

$$T_{ij} = T^*(t_{ij})(1 + 0.1 * e_{1ij}),$$
$$V_{ij} = V^*(t_{ij})(1 + 0.1 * e_{2ij}),$$

where $e_{1ij} \sim N(0, 1)$, $e_{2ij} \sim N(0, 1)$.

Employing the procedure described above, we generate 150 data sets for eight combinations of measurement errors, covariance matrix Σ and n_i. Applying the approach given in Section 2 to the generated data sets, we obtain 150 sets of estimated fixed-effects ϕ, the standard errors of estimated fixed-effects ϕ and the variance components of random-effects $\sigma_{b,k}$'s for each combination. Here we summarize the

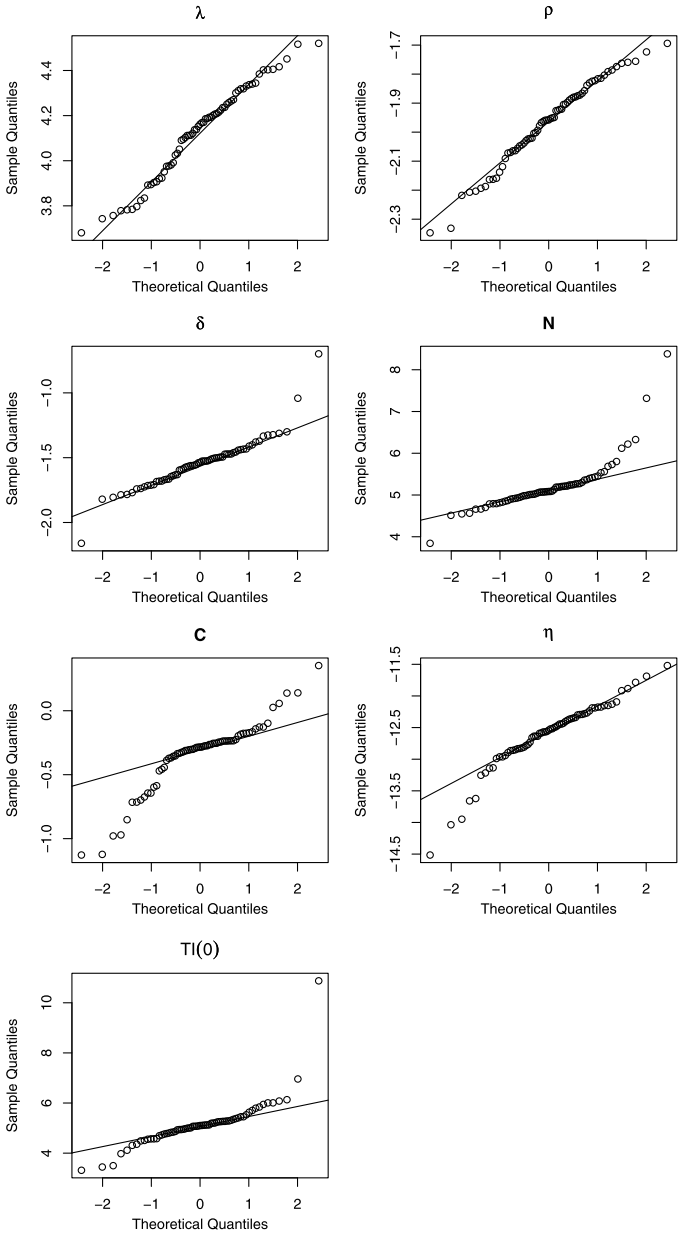

Figure 3. The Q-Q Plots of the estimated parameters in Perelson's HIV dynamic models for individuals.

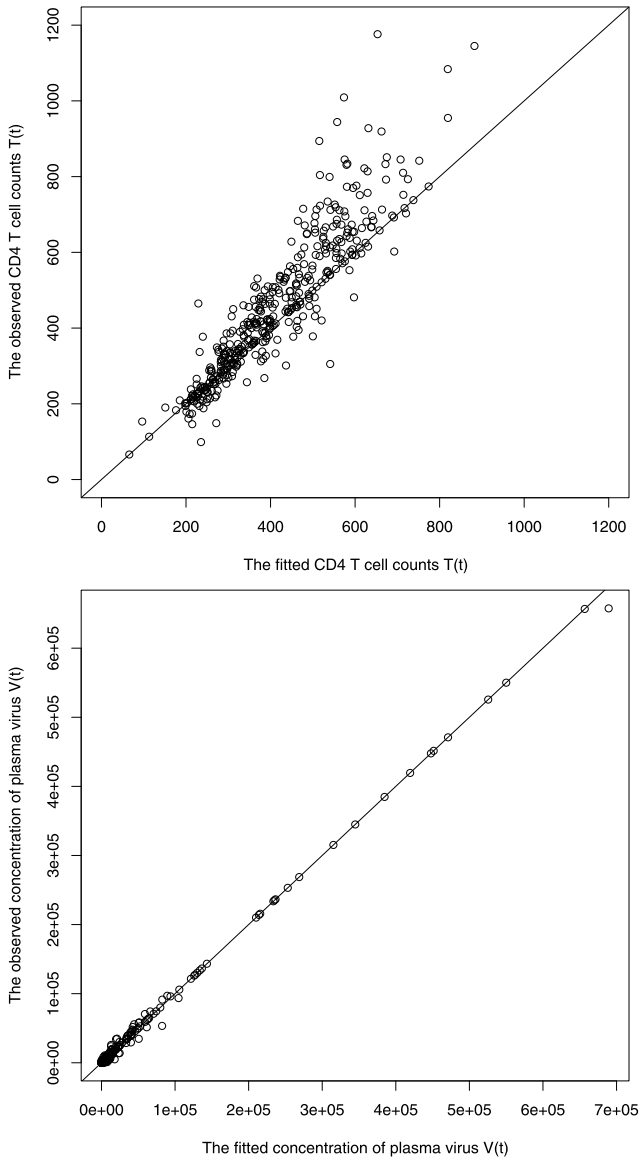

Figure 4. The plots of the observed versus individual fitted values.

10% trimmed means and standard deviations (the mean and standard deviations of estimates after discarding 5% of the lowest and the highest values) of the estimated ϕ and $\sigma_{b,k}$'s in Tables 3 and 4 respectively. Table 3 shows that the biases of ϕ are small and the standard errors of the estimated fixed-effects ϕ are reasonable. Meanwhile, Table 4 shows that for setting 1 and setting 2 in which the data are generated without measurement errors, the estimated $\sigma_{b,k}$'s are very close to the true values which indicates that the approach introduced in this paper has the potential capability to estimate the variance components of random-effects if we can improve the accuracy of measurements. For the settings with 10% errors, when n_i=15 the variance components for $\ln(\lambda)$, $\ln(\rho)$, $\ln(\delta)$, $\ln(c)$, and $\ln(T_{I0})$ can be roughly approximated. However when n_i=41, the variance components for $\ln(\lambda)$, $\ln(\rho)$, $\ln(\delta)$, $\ln(c)$ and $\ln(T_{I0})$ can be well estimated. Although the biases of the estimated variance components for $\ln(\eta)$ and $\ln(N)$ are large, we can see from Table 4 that when n_i increases from 15 to 41, the variance component estimates for both $\ln(\eta)$ and $\ln(N)$ have the tendency toward the true value. In practice, obtaining a HIV clinic trial data set with small measurement errors may be difficult. According to our simulation studies, we can see that increasing the frequency of collecting measurements can efficiently improve the estimation of the variance components of random-effects.

Generally speaking, the estimates of fixed-effects outperforms these of the random-effects. The fixed-effects ϕ, to-

Table 3. The 10% trimmed mean and standard deviation of the fixed-effects estimators $\hat{\phi}$ for Perelson's HIV dynamic model based on 150 replications

Setting	n_i	Errors	Σ	$\ln(\lambda) = 4.13$		$\ln(\rho) = -1.97$		$\ln(\delta) = -1.54$		$\ln(N) = 5.21$		$\ln(c) = -0.34$		$\ln(\eta) = -12.61$		$\ln(T_{I0}) = 5.14$	
				est	sd	est	sd	est	sd	est	sd	est	sd	est	sd	est	sd
1	15	0%	Σ_1	4.13	0.09	−1.97	0.08	−1.53	0.17	5.10	0.35	−0.33	0.11	−12.29	0.50	5.36	0.51
2	15	0%	Σ_2	4.15	0.11	−1.94	0.15	−1.49	0.17	5.18	0.45	−0.32	0.13	−12.44	0.55	5.21	0.63
3	15	10%	Σ_1	4.22	0.16	−1.88	0.16	−1.48	0.22	5.28	0.48	−0.29	0.17	−12.21	0.57	5.59	0.73
4	15	10%	Σ_2	4.24	0.22	−1.87	0.23	−1.43	0.33	5.02	0.59	−0.40	0.17	−12.23	0.76	5.26	1.02
5	21	10%	Σ_1	4.30	0.17	−1.78	0.19	−1.37	0.23	5.36	0.39	−0.25	0.13	−12.25	0.45	5.52	0.65
6	21	10%	Σ_2	4.29	0.18	−1.81	0.19	−1.39	0.24	5.24	0.49	−0.28	0.17	−12.37	0.56	5.44	0.78
7	41	10%	Σ_1	4.25	0.14	−1.83	0.16	−1.44	0.18	5.35	0.34	−0.25	0.14	−12.34	0.36	5.53	0.59
8	41	10%	Σ_2	4.29	0.18	−1.77	0.19	−1.37	0.25	5.22	0.37	−0.29	0.15	−12.32	0.43	5.42	0.77

Table 4. The 10% trimmed mean and standard deviation of the random-effects estimates $\sigma_{b,k}$'s for Perelson's HIV dynamic models based on 150 replications

Setting	n_i	Errors	Σ	$\ln(\lambda)$		$\ln(\rho)$		$\ln(\delta)$		$\ln(N)$		$\ln(c)$		$\ln(\eta)$		$\ln(T_{I0})$	
				$\sigma_{b,1} = 0.21$		$\sigma_{b,2} = 0.15$		$\sigma_{b,3} = 0.13$		$\sigma_{b,4} = 0.62$		$\sigma_{b,5} = 0.27$		$\sigma_{b,6} = 0.53$		$\sigma_{b,7} = 0.94$	
				est	sd	est	sd	est	sd	est	sd	est	sd	est	sd	est	sd
1	15	0%	Σ_1	0.18	0.10	0.18	0.09	0.29	0.18	0.57	0.39	0.31	0.14	0.66	0.43	0.89	0.53
2	15	0%	Σ_2	0.21	0.13	0.18	0.13	0.32	0.20	0.58	0.43	0.33	0.20	0.58	0.42	0.89	0.67
3	15	10%	Σ_1	0.13	0.10	0.12	0.10	0.13	0.12	0.17	0.26	0.23	0.16	0.09	0.19	0.42	0.42
4	15	10%	Σ_2	0.20	0.13	0.12	0.13	0.10	0.13	0.07	0.15	0.15	0.12	0.08	0.17	0.36	0.43
5	21	10%	Σ_1	0.14	0.09	0.14	0.11	0.10	0.11	0.26	0.30	0.23	0.12	0.15	0.21	0.70	0.48
6	21	10%	Σ_2	0.22	0.12	0.12	0.12	0.13	0.15	0.16	0.23	0.18	0.12	0.13	0.20	0.76	0.52
7	41	10%	Σ_1	0.16	0.08	0.15	0.10	0.11	0.12	0.31	0.30	0.25	0.14	0.25	0.27	1.08	0.63
8	41	10%	Σ_2	0.24	0.10	0.13	0.12	0.14	0.16	0.23	0.25	0.23	0.12	0.19	0.23	1.13	0.50

gether with the variance components of random-effects $\sigma_{b,k}$'s can be well estimated when plasma viral load and the total CD4$^+$ T cell counts can be observed precisely. For the generated data sets with 10% measurement errors, the fixed-effects ϕ can be well estimated whether the sample size for individuals is large or small. And when n_i=41, the estimation for most of the variance components $\sigma_{b,k}$'s can be well performed which indicates that increasing sample size can result in better estimates even for data with measurement errors. In addition, from Tables 3 and 4, one can see that there is no substantial difference among the estimates obtained from data generated by Σ_1 and Σ_2. As a result, we know that the assumption employed in Section 2.2 does not ruin the estimated results and estimates obtained in Section 3 are reliable.

5. CONCLUSION

In this paper, we apply NLME to estimate the fixed-effects and random-effects for parameters in Perelson's HIV dynamic model and the baseline of infected CD4$^+$ T cells based on partially observed state variables. All dynamic parameters and T_{I0} were directly estimated from experimental data. To the best of our knowledge, there is no literature considering the estimation of unobservable T_{I0}. In our paper, to maintain the biological meanings of dynamic parameters, we solve Perelson's HIV dynamic system directly to obtain numerical solutions. As a result, the numerical solutions are functions of unknown dynamic parameters and baselines of state variables. And the value of T_{I0} may affect the estimation of dynamic parameters. Although T_{I0} isn't our interest parameter, simply treating the baselines of state variables as functions of dynamic parameters may cause misleading results. When we perform the real data analysis, to overcome common problems for general NLME such as convergence problems, we tried multiple start values generated from the ranges given by previous research.

Moreover, we design eight different settings based on the estimates of real data analysis to verify the reliability of the obtained estimates and explore possible approaches to improve the accuracy of the estimators. Our simulation results demonstrate small biases of the fixed-effects estimates in settings where the observations for individual are rich as well as sparse. As to the challenging problem, the estimation of variance components for the random-effects, settings considering the data without measurement errors provide us with good estimates which are very close to the true values, therefore, indicating that the approach is promising in parameter estimation for the random-effects once we can improve the accuracy of clinical trial data. For the generated data considering 10% measurement errors, when the sample size for individuals is 41, the variance components of $\ln(N)$

and $\ln(\eta)$ can be estimated roughly and the variance components of the rest of the dynamic parameters can be well estimated.

Here, we only consider short-term data analysis while assuming η is constant. An algorithm being able to deal with time-varying dynamic parameters for long-term HIV dynamic model will be considered in the near future.

ACKNOWLEDGEMENT

The authors thank the editor and referees for their comments which improved this article. Liang's research was partially supported by NSF grants DMS1007167 and DMS1207444.

Received 28 December 2011

REFERENCES

[1] Perelson, A. S. (2002). Modelling viral and immune system dynamics. *Nature Reviews Immunology* **2**:28–36.

[2] Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M. and Ho, D. D. (1996). HIV-1 dynamics in vivo: Virion clearance rate, infected cell life-span and viral generation time. *Science* **271**:1582–1586.

[3] Wei, X., Ghosh, S. K., Taylor, M. E., Johnson, V. A., Emini, E. A., Deutsch, P., Lifsonparallel, J. D., Bonhoeffer, S., Nowak, M. A., Hahn, B. H., Saag, M. and Shaw, G. M. (1995). Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* **373**:117–122.

[4] DeBoer, R. J. and Boucher, C. A. B. (1996). Anti-CD4 therapy for aids suggested by mathematical models. *Proceedings of the Royal Society of London-Series B: Biological Sciences* 1996; **263**:899–905.

[5] Nowak, M. A. and Bangham, C. R. M. (1996). Population dynamics of immune responses to persistent viruses. *Science* **272**:74–79.

[6] Wu, H. L., Ding, A. A. and DeGruttola, V. (1998). Estimation of HIV dynamic parameters. *Statistics in Medicine* **17**:2463–2485.

[7] Wu, H. L. 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.

[8] Wu, H. L. and Wu, L. (2002). Identification of significant host factors for HIV dynamics modelled by non-linear mixed-effects models. *Statistics in Medicine* **21**:753–771.

[9] Wu, H. L. and Zhang, J. T. (2002). The study of long-term HIV dynamics using semi-parametric non-linear mixed-effects models. *Statistics in Medicine* **21**:3655–3675.

[10] Wu, H. L., Zhao, C. X. and Liang, H. (2004). Comparisons of linear model, nonlinear model and semiparametric model for estimating HIV dynamic parameters. *Biometrical Journal* **46**:233–245. MR2099362

[11] Delyon, B., Lavielle, M. and Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Statistics* **27**:94–128. MR1701103

[12] Samson, A., Lavielle, M. and Mentre, F. (2006). Extension of the SAEM algorithm to left-censored data in nonlinear mixed-effects model: Application to HIV dynamics model. *Computational Statistics & Data Analysis* **51**:1562–1574. MR2307526

[13] Chan, P. L. S., Jacqmin, P., Lavielle, M., McFadyen, L. and Weatherley, B. (2011). The use of the SAEM algorithm in MONOLIX software for estimation of population pharmacokinetic-pharmacodynamic-viral dynamics parameters of maraviroc in asymptomatic HIV subjects. *Journal of Pharmacokinetics and Pharmacodynamics* **38**:41–61.

[14] Lavielle, M., Samson, A., Fermin, A. K. and Mentre, F. (2011). Maximum Likelihood Estimation of Long-Term HIV Dynamic Models and Antiviral Response. *Biometrics* **67**:250–259. MR2898837

[15] Roshammar, D., Simonsson, U. S. H., Ekvall, H., Flamholc, L., Ormaasen, V., Vesterbacka, J., Wallmark, E., Ashton, M. and Gisslen, M. (2011). Non-linear mixed effects modeling of antiretroviral drug response after administration of lopinavir, atazanavir and efavirenz containing regimens to treatment-naive HIV-1 infected patients. *Journal of Pharmacokinetics and Pharmacodynamics* **38**:727–742.

[16] Liang, H. and Wu, H. L. (2008). Parameter estimation for differential equation models using a framework of measurement error in regression models. *Journal of the American Statistical Association* **103**:1570–1583. MR2504205

[17] Fang, Y., Wu, H. L. and Zhu, L. X. (2011). A two-stage estimation method for random coefficient differential equation models with application to longitudinal HIV dynamic data. *Statistica Sinica* **21**:1145–1170. MR2827518

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

[19] Huang, Y. X., Liu, D. C. and Wu, H. L. (2006). Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. *Biometrics* **62**:413–423. MR2227489

[20] Huang, Y. X. and Wu, H. L. (2006). A Bayesian approach for estimating antiviral efficacy in HIV dynamic models. *Journal of Applied Statistics* **33**:155–174. MR2223142

[21] Huang, Y. X. (2010). A Bayesian approach in differential equation dynamic models incorporating clinical factors and covariates. *Journal of Applied Statistics* **37**:181–189. MR2751325

[22] Perelson, A. S., Kirschner, D. E. and Boer, R. D. (1993). Dynamics of HIV infection of CD4+ T cells. *Mathematical Biosciences* **114**:81–125.

[23] 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.

[24] Perelson, A. S., Essunger, P., Cao, Y. Z., Vesanen, M., Hurley, A., Saksela, K., Markowitz, M. and Ho, D. D. (1997). Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* **387**:188–191.

[25] Perelson, A. S. and Nelson, P. W. (1999) Mathematical analysis of HIV-1 dynamics in vivo. *SIAM Review* **41**:3–44. MR1669741

[26] Petzold, L. (1983). Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. *SIAM Journal on Scientific and Statistical Computing* **4**:136–148. MR0689694

[27] Lindstrom, M. J. and Bates, D. M. (1990). Nonlinear mixed-effects models for repeated measures data. *Biometrics* **46**:673–687. MR1085815

[28] Davidian, M. and Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data*. New York: Chapman and Hall.

[29] Tornoe, C. W., Agerso, H., Jonsson, E. N., Madsen, H. and Nielsen, H. A. (2004). Non-linear mixed-effects pharmacokinetic/pharmacodynamic modelling in NLME using differential equations. *Computer Methods and Programs in Biomedicine* **76**:31–40.

[30] Wang, Y., Eskridge, K. M. and Zhang, S. P. (2008). Semiparametric mixed-effects analysis of PK/PD models using differential equations. *Journal of Pharmacokinetics and Pharmacodynamics* **35**:443–463.

[31] Pinheiro, J. C. and Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics* **4**:12–35.

[32] McMahon, D., Lederman, M., Haas, D. W., Haubrich, R., Stanford, J., Cooney, E., Horton, J., Kelleher, D., Ross, L., Cutrell, A., Lee, D., Spreen, W. and Mellors, J. (2001). Antiretroviral activity and safety of abacavir in combination with selected HIV-1 protease inhibitors in therapy-naive HIV-1-infected adults. *Antiviral Therapy* **6**:105–114.

[33] Gray, D. and Leanderson, T. (1990). Expansion, selection and maintenance of memory B-cell clones. *Current Topics in Microbiology and Immunology* 1990; **159**:1–17.

[34] Merrill, S. (1988). AIDS: Background and the dynamics of the decline of immunocompetence. In *Theoretical Immunology, Part 2*, Perelson, A. S. ed. Addison-Wesley, Redwood City, California, 59–75.

Yao Yu
Department of Biostatistics and Computational Biology
University of Rochester Medical Center
Rochester, New York 14642
USA
E-mail address: Yao_Yu@urmc.rochester.edu

Hua Liang
University of Rochester Medical Center
Rochester, New York 14642
USA
E-mail address: hliang@bst.rochester.edu