Published in final edited form as:

Stat Med. 2008 October 15; 27(23): 4722–4739. doi:10.1002/sim.3272.

Identifying significant covariates for anti-HIV treatment response: mechanism-based differential equation models and empirical semiparametric regression models

Yangxin Huang¹, Hua Liang², and Hulin Wu²

¹ Department of Epidemiology and Biostatistics, College of Public Health, MDC 56 University of South Florida, Tampa, FL33612, USA

² Department of Biostatistics and Computational Biology, School of Medicine and Dentistry, University of Rochester, Rochester, NY 14642, USA

Summary

In this paper, the mechanism-based ordinary differential equation (ODE) model and the flexible semiparametric regression model are employed to identify the significant covariates for antiretroviral response in AIDS clinical trials. We consider the treatment effect as a function of three factors (or covariates) including pharmacokinetics, drug adherence and susceptibility. Both clinical and simulated data examples are given to illustrate these two different kinds of modeling approaches. We found that the ODE model is more powerful to model the mechanism-based nonlinear relationship between treatment effects and virological response biomarkers. The ODE model is also better in identifying the significant factors for virological response, although it is a little bit liberal and there is a trend to include more factors (or covariates) in the model. The semiparametric mixed-effects regression model is very flexible to fit the virological response data, but it is too liberal to identify correct factors for virological response; sometimes it may miss the correct factors. The ODE model is also biologically justifiable and good for predictions and simulations for various biological scenarios. The limitations of the ODE models include the high cost of computation and the requirement of biological assumptions that sometimes may not be easy to validate. The methodologies reviewed in this paper are also generally applicable to studies of other viruses such as hepatitis B virus (HBV) or hepatitis C virus (HCV).

Keywords

Adherence; AIDS; antiretroviral therapy; Bayesian mixed-effects models; drug resistance; nonlinear mixed-effects models; pharmacodynamics; regression spline; semiparametric regression; viral dynamic models

1 Introduction

In the past two decades, many mathematical models and statistical methods have been developed to model the data from AIDS studies [1,2,3,4,5,6,7]. In particular, modeling HIV dynamics has played an important role in understanding pathogenesis of HIV infections [8,9,10,11]. Two different types of models have been proposed for modeling biomarker data, such as plasma HIV RNA levels (viral load) and CD4 T cell counts, from HIV/AIDS clinical studies. One class of models is the mechanism-based dynamic (differential equation)

model and another class of models is the regression model including linear, nonlinear, nonparametric/semiparametric regression models and time-varying coefficient models. In order to efficiently model the longitudinal biomarker data from AIDS clinical studies, the “mixed-effects” or “random-effects” coefficients have been incorporated into these two classes of models [11,12,13,14,15,16,17,18,19,20].

The regression models link the antiviral response, such as viral load, to other covariates (e.g., CD4 T cell counts) via a regression model which is determined after initial data exploration. The regression models, including linear, nonlinear, nonparametric/semiparametric regression models and time-varying coefficient models, are quite flexible to fit the experimental data. Usually, model fitting is very good since the appropriate models can be selected after exploring the data. The regression models are also robust to the biological assumptions, and the assumption of a parametric form may not be necessary if the nonparametric or semiparametric regression models are used. However, the drawbacks of the regression models include that the parameters may not have biological interpretations and the prediction may not be easy to perform, in particular for nonparametric regression models. In contrast, the mechanism-based ordinary differential equation (ODE) models are usually developed based on biological mechanisms of HIV infection. The mechanism-based ODE models are easier to interpret biologically. All the parameters have biological meaning and may be of a great interest to biomedical investigators. In addition, the mechanism-based ODE models appropriately capture the complicated nonlinear relationship and interactions among different factors in a meaningful way. A drawback of the mechanism-based models is that biological assumptions are required and some of these assumptions are difficult to validate.

In this paper, we give a review of the two classes of models for identifying the covariates for antiviral treatment response in AIDS clinical studies. We mainly consider three covariate factors, pharmacokinetics, drug adherence and susceptibility in an AIDS clinical study. In this study, the parameter C_{12h} (the drug concentration in plasma measured at 12 hours from dose taken) represents the pharmacokinetic properties, the drug adherence is measured from pill count data, and drug susceptibility is measured by IC_{50} (the median inhibitory concentration of the drug). In Section 2, we summarize the mechanism-based ODE models while the semiparametric mixed-effects regression models are reviewed in Section 3. Both an AIDS clinical study and simulated data analysis are presented to illustrate these two classes of modeling approaches in Section 4. We conclude the paper with a brief discussion and future direction in Section 5.

2 Mechanism-Based Differential Equation Models

Differential equations have a long and illustrious history in mathematical modeling. The use of a system of differential equations describing the interaction between HIV and its host cells can be traced back to late 1980s [1,2]. Many mathematical models for HIV dynamics have been developed over the past 20 years. Recent surveys can be found in Perelson and Nelson [3], Nowak and May [4], Perelson [5], Wu [6] and Tan and Wu [7]. HIV dynamic studies have significantly contributed to the understanding of HIV infection. However, most studies [3,8,9,12] are limited to short-term viral dynamics due to the difficulty of establishing a relationship of virologic response with multiple factors such as drug exposure and drug susceptibility during long-term treatment. A mechanism-based ODE model for characterizing long-term viral/T lymphocyte dynamics with antiretroviral therapy, has been developed and Bayesian method has been proposed for ODE parameter estimation by Huang *et al.* [20]. This model allows us to incorporate the factors such as drug exposure and drug susceptibility for predicting antiviral response in a natural way. For completeness, a brief summary of the models and methods [20] is given as follows.

2.1 Drug efficacy models

As Molla *et al.* [21] suggested, the phenotype marker, median inhibitory concentration (IC_{50}), can be used to quantify agent-specific drug susceptibility. We use the following model to approximate the within-host changes over time in IC_{50} ,

$$IC_{50}(t) = \begin{cases} I_0 + \frac{I_r - I_0}{t_r} t & \text{for } 0 < t < t_r, \\ I_r & \text{for } t \geq t_r, \end{cases} \quad (2.1)$$

where I_0 and I_r are respective values of $IC_{50}(t)$ at baseline and at time point t_r when resistant mutations dominate. In our study, t_r is the time of virological failure which is observed from clinical studies.

Poor adherence to a treatment regimen is one of the major causes of treatment failure [22]. The following model is used to represent adherence for a time interval $T_{k-1} < t \leq T_k$,

$$A(t) = \begin{cases} 1 & \text{if all doses are taken in } (T_{k-1}, T_k], \\ R_k & \text{if } 100 R_k \% \text{ doses are taken in } (T_{k-1}, T_k], \end{cases} \quad (2.2)$$

where $0 \leq R_k < 1$, with R_k indicating the adherence rate during the interval $(T_{k-1}, T_k]$, T_k denotes the adherence evaluation time at the k th clinical visit. In clinical studies, adherence can be measured using questionnaires, pill counts, and electronic compliance monitoring caps.

In most viral dynamic studies, investigators assumed that either drug efficacy was constant over treatment time [3,13] or antiviral regimens had perfect effect in blocking viral replication [8,9]. However, the drug efficacy may change as concentrations of antiretroviral drugs and other factors (*e.g.* drug resistance) vary during treatment. We employ the following modified E_{max} model [23] to represent the time-varying drug efficacy for two antiretroviral agents within a class,

$$\gamma(t) = \frac{C_1 A_1(t) / IC_{50}^1(t) + C_2 A_2(t) / IC_{50}^2(t)}{\varphi + C_1 A_1(t) / IC_{50}^1(t) + C_2 A_2(t) / IC_{50}^2(t)}, \quad (2.3)$$

where C_1 and C_2 represent the respective drug concentrations or any of the PK parameters such as C_{12h} , $IC_{50}^1(t)$ and $IC_{50}^2(t)$ indicate the median inhibitory concentrations of the two drugs, and $A_1(t)$ and $A_2(t)$ are adherence profiles of the two drugs (measured by pill counts in our real clinical example). The parameter φ can be regarded as a conversion factor between *in vitro* and *in vivo* IC_{50} s and will be estimated from the data. $\gamma(t)$ ranges from 0 to 1.

2.2 Antiviral response model

We consider a simplified HIV dynamic model with antiviral treatment as follows [20],

$$\begin{aligned}\frac{d}{dt}T(t) &= \lambda - d_T T(t) - [1 - \gamma(t)]kT(t)V(t), \\ \frac{d}{dt}T^*(t) &= [1 - \gamma(t)]kT(t)V(t) - \delta T^*(t), \\ \frac{d}{dt}V(t) &= N\delta T^*(t) - cV(t),\end{aligned}\quad (2.4)$$

where the three differential equations represent three compartments: target uninfected cells (T), infected cells (T*) and free virions (V). The parameter λ represents the rate at which new T cells are generated from sources within the body, such as the thymus, d_T is the death rate of T cells, k is the infection rate without treatment, δ is the death rate of infected cells, N is the number of new virions produced from each infected cell during its life-time, and c is the clearance rate of free virions. The time-varying parameter $\gamma(t)$ is the antiviral drug efficacy at treatment time t . If the regimen is not 100% effective (imperfect inhibition), the system of ODEs cannot be solved analytically. The solutions to (2.4) then have to be evaluated numerically. In the estimation procedure, we only need to evaluate the difference between observed data and numerical solutions of $V(t)$. So there is no need for an explicit solution of equations (2.4).

2.3 Bayesian nonlinear mixed-effects model

A number of studies investigated various statistical methods, including Bayesian approaches, to fit viral dynamic models and to predict virological responses using short-term viral load data [8,9,12,13,20,24]. Huang *et al.* [20] extended the existing methods to model long-term HIV dynamics of virological response.

We denote the number of subjects by n and the number of measurements on the i th subject by m_i . For notational convenience, let $\boldsymbol{\mu} = (\ln \varphi, \ln c, \ln \delta, \ln \lambda, \ln d_T, \ln N, \ln k)^T$, $\boldsymbol{\theta}_i = (\ln \varphi_i, \ln c_i, \ln \delta_i, \ln \lambda_i, \ln d_{Ti}, \ln N_i, \ln k_i)^T$ and $\mathbf{Y} = \{y_i(t_{ij}), i=1, \dots, n; j=1, \dots, m_i\}$. Let $f_i(\boldsymbol{\theta}_i, t_{ij}) = \log_{10}(V_i(\boldsymbol{\theta}_i, t_{ij}))$, where $V_i(\boldsymbol{\theta}_i, t_{ij})$ denotes the numerical solution of the differential equations (2.4) for the i th subject at time t_{ij} . Let $y_i(t_{ij})$ and $e_i(t_{ij})$ denote the common logarithmic viral load measurements over time for the i th subject and a measurement error with mean zero, respectively. The Bayesian nonlinear mixed-effects model can be written as the following three stages [20, 25].

Stage 1. Within-subject variation:

$$\mathbf{y}_i = \mathbf{f}_i(\boldsymbol{\theta}_i) + \mathbf{e}_i, \quad \mathbf{e}_i | \sigma^2, \boldsymbol{\theta}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{m_i}) \quad (2.5)$$

where $\mathbf{y}_i = (y_i(t_{i1}), \dots, y_i(t_{im_i}))^T$, $\mathbf{f}_i(\boldsymbol{\theta}_i) = (f_i(\boldsymbol{\theta}_i, t_{i1}), \dots, f_i(\boldsymbol{\theta}_i, t_{im_i}))^T$, $\mathbf{e}_i = (e_i(t_{i1}), \dots, e_i(t_{im_i}))^T$.

Stage 2. Between-subject variation:

$$\boldsymbol{\theta}_i = \boldsymbol{\mu} + \mathbf{b}_i, \quad [\mathbf{b}_i | \boldsymbol{\Sigma}] \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \quad (2.6)$$

Stage 3. Hyperprior distributions:

$$\sigma^{-2} \sim Ga(a, b), \quad \boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\eta}, \boldsymbol{\Lambda}), \quad \boldsymbol{\Sigma}^{-1} \sim Wi(\boldsymbol{\Omega}, \nu) \quad (2.7)$$

where the mutually independent Gamma (Ga), Normal (N) and Wishart (Wi) prior distributions are chosen to facilitate computations [25]. The hyper-parameters $a, b, \boldsymbol{\eta}, \boldsymbol{\Lambda}, \boldsymbol{\Omega}$

and v were determined from previous studies and the literature [3,4,8,9]. See Huang *et al.* [20] for a detailed discussion of the Bayesian modeling approach, including the choice of the hyper-parameters and the implementation of the Markov chain Monte Carlo (MCMC) procedures.

3 Semiparametric Regression Models

Regression models can also be used to establish the relationship between the covariates and the antiviral response. A variety of parametric models such as linear mixed-effects [13,17], nonlinear mixed-effects models [12,13,17] and semiparametric/nonparametric models [14,17,19] have been proposed to study the dynamics of HIV infection over the past decade. However, most parametric models and methods are applicable only to short-term viral dynamics data [8,9,10,13]. Since the long-term viral load data fluctuate significantly within-subject and patterns vary between-subject, it is difficult to find a parametric function to model the long-term viral load data. A non-parametric regression model is flexible to fit the long-term viral load data as a time function, but we also need to incorporate other covariates. In order to flexibly model the viral load trajectories as well as incorporate the covariates in simple parametric forms, the semiparametric regression models were proposed [14,17,19]. A time-varying non-parametric component can be used to flexibly model the time patterns of viral load trajectories while a linear model can be used to model covariate effects. To efficiently model the longitudinal data, random-effects (mixed-effects) were also introduced into both the nonparametric component and the linear coefficients [19]. Thus, a semiparametric mixed-effects (SPME) model can very flexibly model long-term HIV viral load data [17,19]. Note that, compared to the ODE models, the semiparametric regression models are easy to implement using existing software packages such as SAS and Splus/R. The SPME model can be specified as [17,19]:

$$y_i(t_{ij}) = x_{ij}^T \beta_i + \eta_i(t_{ij}) + e_i(t_{ij}), \quad j=1, \dots, m_i, \quad i=1, \dots, n, \quad (3.1)$$

where $\beta_i = (\beta + a_i)$, $\eta_i(t) = \eta(t) + \kappa_i(t)$, $y_i(t_{ij})$ is the common logarithm of the viral load measurements in our HIV/AIDS clinical study (response variable) and x_{ij} is a vector representing covariates such as drug exposure and drug susceptibility for the z th subject at time t_{ij} . The error term $e_i(t)$ is a zero mean stochastic process with covariance function $\rho_e(s, t) = \text{cov}\{e_i(s), e_i(t)\}$. β and $\eta(t)$ describe the population characteristics, while a_i and $\kappa_i(t)$ reflect individual variations from β and $\eta(t)$. We are concerned with the population parameter β and curve $\eta(t)$ as well as individual parameters β_i and curves $\eta_i(t) = \eta(t) + \kappa_i(t)$, for $i = 1, \dots, n$. The population parameter or curve is important because it reflects the overall trend or progress of an underlying population process and can be used as an important index for the population relationship between viral load levels and other covariates. The individual parameters/curves are also important, because they characterize the individual effects and may help biomedical investigators to develop individualized treatment strategies for AIDS patients.

Semiparametric models with a similar form as (3.1) for cross-sectional data have been extensively studied in the past two decades [26,27,28], and were extended recently to longitudinal data modeling [14,17,19,29,30]. To incorporate the correlations of longitudinal data as well as between-subject and within-subject variations, the model can be extended to include random-effects components. Zeger and Diggle [31] used a similar semiparametric model to analyze an AIDS data set. Semiparametric mixed-effects models for longitudinal data have been extensively studied in the recent literature [14,17,19,32,33,34]. Semiparametric nonlinear mixed-effects models have also been proposed and investigated [14,35]. Local polynomial, smoothing spline and regression spline methods, as well as other

smoothing techniques, have been proposed for estimation and inference for nonparametric/semiparametric regression models [19].

Model (3.1) combines a linear mixed-effects model [36,37] and nonparametric mixed-effects models [19,38,39,40,41] into the semiparametric mixed-effects model. Many smoothing techniques are available to estimate the nonparametric functions $\eta(t)$ and $\kappa_i(t)$, including local polynomial kernel methods, smoothing splines and regression splines [19]. A practical concern of the first two techniques is that they require intensive computation and sometimes they are not robust against smoothing parameters. In this paper we use the regression splines approach to approximate the nonparametric functions $\eta(t)$ and $\kappa_i(t)$ because of its simplicity and easy implementation. Shi *et al.* [39] and Rice and Wu [41] have also applied the regression splines approach for nonparametric mixed-effects estimates. The regression spline methods transform a nonparametric curve into a linear combination of basis functions so that we can operationally implement the model-fitting and inference using the standard linear mixed-effects (LME) model approach which is available from many statistical software packages such as SAS and SpluS/R.

Let $\mathbf{t}_i = (t_{i1}, \dots, t_{imi})^T$, $\mathbf{y}_i = (y_i(t_{i1}), \dots, y_i(t_{imi}))^T$, $\boldsymbol{\eta}(\mathbf{t}_i) = (\eta(t_{i1}), \dots, \eta(t_{imi}))^T$, $\boldsymbol{\kappa}_i(\mathbf{t}_i) = (\kappa_i(t_{i1}), \dots, \kappa_i(t_{imi}))^T$, $\mathbf{x}_i = (x_{i1}, \dots, x_{imi})^T$ and $\mathbf{e}_i = (e_i(t_{i1}), \dots, e_i(t_{imi}))^T$. Then, model (3.1) can be written in vector notation as

$$\mathbf{y}_i = \mathbf{x}_i^T \boldsymbol{\beta}_i + \boldsymbol{\eta}_i(\mathbf{t}_i) + \mathbf{e}_i, \quad i=1, \dots, n. \quad (3.2)$$

Our primary interest is to estimate both population coefficients $\boldsymbol{\beta}$ and $\eta(t)$, individual coefficients $\boldsymbol{\beta}_i$ and $\kappa_i(t)$ and identify which of the covariates is significant for predicting the response variable (viral load in our application example).

We approximate the nonparametric functions $\eta(t)$ and $\kappa_i(t)$ by the basis functions

$$\eta_p(t) = \sum_{k=0}^p \xi_k \theta_k(t) = \boldsymbol{\Theta}_p(t)^T \boldsymbol{\xi}_p \quad \text{and} \quad \kappa_{i,r}(t) = \sum_{k=0}^r b_{ik} \psi_k(t) = \boldsymbol{\Psi}_r(t)^T \mathbf{b}_i,$$

where $\boldsymbol{\Theta}_p(t) = (\theta_0(t), \dots, \theta_p(t))^T$, $\boldsymbol{\Psi}_r(t) = (\psi_0(t), \dots, \psi_r(t))^T$, $\boldsymbol{\xi}_p = (\xi_0, \dots, \xi_p)^T$ and $\mathbf{b}_i = (b_{i0}, \dots, b_{ir})^T$. Also note that for fixed p and r , the truncated vector $\mathbf{b}_i$ is a random vector with mean 0 and covariance matrix D , thus $\rho(s, t) = \text{cov}\{\kappa_{i,r}(s), \kappa_{i,r}(t)\} = \boldsymbol{\Psi}_r(s)^T D \boldsymbol{\Psi}_r(t)$. Replacing $\eta_i(t)$ in model (3.2) by $\eta_p(t) + \kappa_{i,r}(t)$, we obtain an approximate model

$$\mathbf{y}_i = \sum_{k=0}^p \xi_k \theta_k(\mathbf{t}_i) + \mathbf{x}_i^T \boldsymbol{\beta}_i + \sum_{k=0}^r b_{ik} \psi_k(\mathbf{t}_i) + \mathbf{e}_i(\mathbf{t}_i). \quad (3.3)$$

For given p and r , this is a linear mixed-effects (LME) model with fixed-effect terms

$\sum_{k=0}^p \xi_k \theta_k(\mathbf{t}_i) + \mathbf{x}_i^T \boldsymbol{\beta}$ and random-effects terms $\mathbf{x}_i^T \boldsymbol{\beta}_i + \sum_{k=0}^r b_{ik} \psi_k(\mathbf{t}_i)$. Let

$\mathbf{X}_i = (\mathbf{x}_i^T, \theta_0(\mathbf{t}_i), \dots, \theta_p(\mathbf{t}_i))$, $\mathbf{Z}_i = (\mathbf{x}_i^T, \psi_0(\mathbf{t}_i), \dots, \psi_r(\mathbf{t}_i))$, $\boldsymbol{\alpha} = (\boldsymbol{\beta}^T, \boldsymbol{\xi}_p^T)^T = (\boldsymbol{\beta}^T, \xi_0, \dots, \xi_p)^T$, and $\mathbf{u}_i = (a_i^T, b_{i0}, \dots, b_{ir})^T$. Model (3.3) can be expressed as a standard LME model:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{Z}_i \mathbf{u}_i + \mathbf{e}_i(t_i), \quad \mathbf{u}_i \sim (0, D^*), \quad \mathbf{e}_i \sim (0, R_i).$$

Thus, for given D^* and R_i , the closed forms for the estimates of $\boldsymbol{\alpha}$ and $\mathbf{u}_i$ can be written as follows [25,36]:

$$\hat{\boldsymbol{\alpha}} = \left(\sum_{i=1}^n \mathbf{X}_i^T \sum_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}_i^T \sum_i^{-1} \mathbf{y}_i \right), \quad \hat{\mathbf{u}}_i = D^* \mathbf{Z}_i^T \sum_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\alpha}}),$$

where $\sum_i = R_i + \mathbf{Z}_i D^* \mathbf{Z}_i^T$. Consequently, the estimates of $\eta_p(t)$ and $\kappa_{i,r}(t)$ can be expressed as $\hat{\eta}_p(t) = \boldsymbol{\Theta}_p(t)^T \hat{\boldsymbol{\xi}}_p$ and $\hat{\kappa}_{i,r}(t) = \boldsymbol{\Psi}_r(t)^T \hat{\mathbf{b}}_i$. The covariance matrix D^* may be specified as unstructured or with some special structures. The covariance matrix R_i may also have a special structure, but very often we simply set $R_i = \sigma^2 \mathbf{I}_{m_i}$, where σ^2 needs to be estimated. The unknown parameters in D^* and R_i can also be estimated using the maximum likelihood or restricted maximum likelihood method [25,42]. Also note that $\hat{\rho}(s, t) = \text{cov}(\kappa_{i,r}(s), \kappa_{i,r}(t)) = \boldsymbol{\Psi}_r(s)^T \hat{D} \boldsymbol{\Psi}_r(t)$.

The choice of basis functions is usually not as crucial as the selection of p and r [43]. We use the natural cubic spline basis due to its good property and easy implementation using existing software such as R/Splus [37]. Eubank [44] and others have proposed to locate the knots at the quantiles of the data which will be used in our estimation procedure. We use the model selection criteria AIC to determine p and r .

4 Application to An AIDS Clinical Study and Simulated Data Analysis

4.1 An AIDS clinical data analysis

4.1.1 Study design—In this section, we apply our methods to the data from an AIDS Clinical Trials Group (ACTG) protocol A5055 study to evaluate results from both the mechanism-based differential equation model and the semiparametric regression model. This study was a Phase I/II, randomized, open-label, 24-week comparative study of the pharmacokinetics, tolerability and antiretroviral effects of two regimens of indinavir (IDV), zidovudine (ZDV), plus two nucleoside analogue reverse transcriptase inhibitors (NRTIs) on HIV-1-infected subjects failing Protease inhibitor-containing antiretroviral therapies [45,46]. The 44 subjects were randomly assigned to the two treatment regimens, Arm A (IDV 800 mg q12h + ZDV 200 mg q12h) and Arm B (IDV 400 mg q12h + ZDV 400 mg q12h). Study visits occurred at pre-entry, entry (within 14 days of pre-entry), days 7, 14, 28, 56, 84, 112, 140 and 168 of follow-up. Plasma HIV RNA testing was conducted at each study visit. Clinical assessment and laboratory parameters, including CD4 and CD8 cell counts, were performed at all visit weeks with exception of week 1. Phenotypic determination (IC_{50}) of antiretroviral drug resistance was performed at baseline and at the time of virological failure. PK parameters of IDV and ZDV were determined using noncompartmental methods. Calculated pharmacokinetic parameters included the trough level of drug concentration in plasma (C_{trough}), the drug concentration in plasma measured after 12 hours from dose taken (C_{12h}), the maximum drug concentration in plasma (C_{max}) and the area under the plasma concentration-time curve (AUC). To monitor adherence, pill counts were performed at each study visit from week 2 to week 24. Of the 44 subjects, 42 subjects were included in this analysis; of the remaining two subjects, one was excluded from the analysis because the PK parameters were not obtained and the other was excluded because the phenotype assay could not be completed on this subject. More detailed descriptions of this study and data have been reported by Acosta *et al.* [45] and Wu *et al.* [46].

4.1.2 Results from the mechanism-based ODE models—We demonstrated in a previous report [47] that no significant differences were found when we compared the four PK parameters (C_{12h} , C_{trough} , C_{max} , AUC) as covariates of virological response under both the single-factor case (only PK factor included) and the three-factor case (adherence and drug susceptibility in addition to the PK factor included). Thus, C_{12h} will be used in the subsequent analysis. To implement the Bayesian approach, we only need to specify the priors at the population level.

We consider all the seven parameters to be unknown, but we assume that $\log \varphi$ has a non-informative prior and the other six parameters have informative priors. The prior distribution for $\mu = (\log \varphi, \log c, \log \delta, \log \lambda, \log d_T, \log N, \log k)^T$ is assumed to be $\mathcal{N}(\eta, \Lambda)$ with Λ being a diagonal matrix. We chose the values of the hyper-parameters as follows:

$$\begin{aligned} a &= 4.5, \quad b = 9.0, \quad v = 8.0, \quad \eta = (2.5, 1.1, -1.0, 4.6, -2.5, 6.9, -9.0)^T, \\ \Lambda &= \text{diag}(1000.0, 0.0025, 0.0025, 0.0025, 0.0025, 0.0025, 0.001), \\ \Omega &= \text{diag}(1.25, 2.5, 2.5, 2.0, 2.0, 2.0, 2.0). \end{aligned}$$

These values of the hyper-parameters were determined based on several studies in the literature [3,4,8,9,10,12,24]. We implemented our Bayesian approach using the MCMC procedure consisting of a series of Gibbs sampling and Metropolis-Hastings algorithms [20]. The convergence of the MCMC algorithms was carefully monitored [20,48].

In order to assess how adherence and drug susceptibility interact with drug pharmacokinetics to contribute to virological response, we fitted the models with all (eight) possible combinations of the three factors and compared the fitting results. Note that for each of the factors, we always consider the two protease inhibitor (PI) drugs simultaneously, i.e., if any of the three factors, PK (C_{12h}), adherence and drug susceptibility (IC_{50}), were considered in the drug efficacy model (2.3), this factor for both PI drugs was incorporated in the model fitting. The eight scenarios of the drug efficacy models that incorporate all possible combinations of the three factors are specified as follows (See Wu et al. [46] for details):

1. *Control model*: $\gamma_{no}(t) = \frac{2}{\varphi + 2}$;
2. $A: \gamma_A(t) = \frac{A_1(t) + A_2(t)}{\varphi + A_1(t) + A_2(t)}$;
3. $I: \gamma_I(t) = \frac{1/IC_{50}^1(t) + 1/IC_{50}^2(t)}{\varphi + 1/IC_{50}^1(t) + 1/IC_{50}^2(t)}$;
4. $C: \gamma_C(t) = \frac{C_{12h}^1 + C_{12h}^2}{\varphi + C_{12h}^1 + C_{12h}^2}$;
5. $IA: \gamma_{IA}(t) = \frac{A_1(t)/IC_{50}^1(t) + A_2(t)/IC_{50}^2(t)}{\varphi + A_1(t)/IC_{50}^1(t) + A_2(t)/IC_{50}^2(t)}$;
6. $CA: \gamma_{CA}(t) = \frac{C_{12h}^1 A_1(t) + C_{12h}^2 A_2(t)}{\varphi + C_{12h}^1 A_1(t) + C_{12h}^2 A_2(t)}$;
7. $CI: \gamma_{CI}(t) = \frac{C_{12h}^1/IC_{50}^1(t) + C_{12h}^2/IC_{50}^2(t)}{\varphi + C_{12h}^1/IC_{50}^1(t) + C_{12h}^2/IC_{50}^2(t)}$;

$$8. \quad CIA: \gamma_{CIA}(t) = \frac{C_{12h}^1 A_1(t)/IC_{50}^1(t) + C_{12h}^2 A_2(t)/IC_{50}^2(t)}{\varphi + C_{12h}^1 A_1(t)/IC_{50}^1(t) + C_{12h}^2 A_2(t)/IC_{50}^2(t)}.$$

The MCMC techniques [20] were used to fit the proposed models with eight possible combinations of the three factors to the viral load data from the A5055 study. For assessing which of the PK parameters, adherence and drug susceptibility (IC_{50}) or their combinations significantly contributes to virological response, we compared the sum of squared residuals (SSR) from the viral dynamic model fitting for individual subjects using the sign test. All p -values are 2-sided with a significance level of 0.05 in this exploratory analysis. The results are summarized as follows.

For the purpose of illustration and comparison, the model-fitting curves for four randomly selected subjects are plotted in Figure 1 (dotted lines show the control model fitting and the solid lines show the full “CIA” model fitting). From these model fitting results, we clearly see that both models fitted the early viral load data well if early treatment response is good, but the control model without considering the subject-specific information of drug exposure and drug susceptibility failed to fit viral load data with rebounds and fluctuations. By considering drug pharmacokinetics, adherence and drug susceptibility, the long-term viral dynamics data can be fitted well (Figure 1).

Figure 2 displays the Sum of Squared Residuals (SSR) from the fitting of the eight models. Compared to the control model, the SSRs of the model with PK and drug susceptibility (CI) and the model with all three factors (CIA) were significantly smaller ($p = 0.0007$ and $p = 0.0055$ respectively), and both models were also significantly better than any of other 5 models ($p < 0.05$). However, when we compared the model with PK and drug susceptibility (CI) to the model with all three factors (CIA), the SSRs were not significantly different ($p = 0.5371$). These results suggest that the best model for prediction of virological response is the model with PK and drug susceptibility (CI).

4.1.3 Results from the semiparametric regression models—Semiparametric regression model (3.1) was used to fit the data of the A5055 study. To stabilize the variance and computational algorithms, we took the $\log_{10}$ -transformation of HIV viral load data. Similar to the ODE model, we considered the seven covariates (A, I, C, IA, CA, CI and CIA) listed in the last subsection. Note that, if none of the covariates is significant, the regression model is reduced to the control model. The nonparametric component $\eta_i(t)$ was modeled by a 6-order spline. For comparisons, the fitting results from the SPME model with the three-factor covariate of CIA for the same 4 subjects in Figure 1 are presented in Figure 3. The fitted viral load curves for both population (dotted lines) and individuals (solid lines) are presented. Similar to the ODE model, the population fit (similar to the control model) was poor while the individual curve fitting with the three-factor covariate was very good since the subject-specific information of drug exposure and drug susceptibility was used.

We also considered each of the seven covariates in the SPME model separately. We summarize the estimates of the covariate coefficients in Table 1. We found that the effects of all the 7 covariates were significant. These results are different from those of the ODE models from which only the covariates “CI” and “CIA” were identified as the significant covariates. The question is whether the SPME model is too liberal in identifying the significant covariates. In the next subsection, we further investigate this problem using a simulated data set.

4.2 A simulation trial analysis

4.2.1 Study design—Computer simulations become an important tool to evaluate the trial design and analysis plan for clinical trials. In this subsection, we simulated a clinical trial of

20 HIV infected patients with 168 days of antiretroviral treatment. For each patient, we assumed that measurements of viral load are taken at days 0, 7, 14, 28, 56, 84, 112, 140 and 168 of follow-up. The design of this simulation experiment is similar to the A5055 study that we have described in Section 4.1.

We generated the true parameters using the between-subject variation model (2.6), $\theta_i = \mu + \mathbf{b}_i$ ($i=1, \dots, 20$), where we assumed that the population parameter vector $\mu = (2.5, 1.1, -1.0, 4.6, -2.5, 6.9, -11.0)^T$ and the random effects $\mathbf{b}_i$ are normally distributed with mean $\mathbf{0}$ and diagonal standard deviation matrix of $\text{diag}(0.1, 0.2, 0.2, 0.1, 0.2, 0.1, 0.1)$. In addition, the data for the pharmacokinetic factor (C_{12h}) phenotype marker (baseline and failure IC_{50} s) and adherence as well as baseline viral load were taken from the A5055 study (Section 4.1).

Based on generated true parameters and data (C_{12h} , IC_{50} , or $A(t)$) the observations $y_i(t_{ij})$ (the common logarithm of viral load measurements) were generated by perturbing the solution of the differential equations (2.4) with a within-subject measurement error, *i.e.*, $y_i(t_{ij}) = \log_{10} V_i(t_{ij}) + e_i(t_{ij})$, where $V_i(t_{ij})$ is the numerical solution of viral load to the differential equations (2.4) for the i th subject at time t_{ij} . It is assumed that the within-subject measurement error e_i is normally distributed with $\mathcal{N}(0, 0.15^2)$. Seven data sets were generated by considering the seven combinations of the three factors in the drug efficacy model (2.3). Thus, the seven models for treatment effect listed as (2)–(8) in Section 4.1.2 were used, respectively, to generate data. Note that simulated data for each model were used to validate and illustrate the proposed modeling approaches.

4.2.2 Results from the mechanism-based differential equation models—For each of the seven generated data sets, the MCMC techniques [20] were used to fit the model (2.4), respectively, while considering all seven possible combinations of the three factors in the drug efficacy model (2.3) as well as the control model. The simulation results are summarized as follows.

For each data set, it is interesting to find, from Figures 4 and 5 (left panel), that the CI and CIA models are significantly better than the other six models, and the true model is also better than other models, but not significantly better than the CI and CIA models (if the true model is not one of these two). The significantly better models from the ODE method for each data set are listed in Table 2 (the third column), in which the true model (covariate) is highlighted.

As an example, we present more detailed results for one data set generated from the CI model. The mean and standard deviation (SD) of SSRs and p-values obtained from the pairwise comparisons using the sign test for the 8 models are summarized in Table 3. It can be seen from Table 3 that the mean and standard deviation (11.28 ± 18.63) of SSRs for model (IA) were larger than those (10.98 ± 17.50) of the model (I), but the model (I) was not significantly better than the model (IA) ($p = 0.2513$). We also found that the models (IA) and (CA) were not significantly better than any of the models with a single factor, while the model (CI) was significantly better than any of the models with a single factor, the two-factor combination of drug susceptibility and adherence ($p = 0.0001$) as well as the combination of PK and adherence ($p < 0.0001$). The full three-factor model (CIA) was also better than any other models, but not significantly better than the true two-factor model (CI) (0.0593). Thus, the best parsimonious model for prediction of virological response is the model with PK and drug susceptibility (CI) which is the true model that was used to generate observation data (viral load). These results are consistent with those of the real data example in Section 4.1.2.

4.2.3 Results from the semiparametric regression models—We fit the same 7 data sets generated from the differential equation model (2.4) in the last subsection to the semiparametric regression model (3.1). We intended to evaluate whether the semiparametric regression model, similar to the mechanism-based differential equation model, could identify the true covariates or factors that were used to generate the data. Similar to Section 4.1.3, we approximated $\eta(\cdot)$ and $\kappa_i(\cdot)$ by 6-order splines. We considered the covariates γ_A , γ_I , γ_C , γ_{IA} , γ_{CA} , γ_{CI} , γ_{CIA} which are defined in Section 4.1.2. In each fitting, we separately regressed the response variable (viral load) to each of the seven covariates and a nonparametric function of time. The t-test was used to evaluate whether a covariate was significant or not. The significant covariates are presented in the fourth column of Table 2, in which the true model (covariate) is highlighted.

As discussed in Section 4.1.2, the ODE method always selects the true model in all 7 cases, and at the same time the CI and CIA models are also identified as significant models in all 7 cases. In contrast for the semiparametric regression model, when we performed the univariate regression (only one covariate enters in the semiparametric regression model each time), the true model was identified as significant for most of the 7 cases except in the case of model (A). In addition, both the ODE and the univariate semiparametric regression methods are quite liberal in the sense that many other models besides the true model are also identified as significant. From Figures 4 and 5 (right panel), we can see that the model fitting residuals from the SPME model are much smaller than those of the ODE models for all 7 cases. The residuals for model (I) and model (IA) are always larger than those of other models in all 7 cases. The reason for this may be that the drug resistance data and adherence data are very noisy and not informative in our studies. Further investigation on this problem is warranted.

For the semiparametric regression model, we also performed the multiple regression analyses with all possible combinations of covariates and we applied the AIC and BIC criteria to select the best combination of covariates for the model fitting. We also report the results in Table 2 (column 5). It is very interesting to notice that the best model, selected by AIC or BIC criterion, only identified the true covariates in 3 out of 7 cases. For instance, we identified, by both AIC and BIC criteria, that the best model for the data set generated from the CI model is of form “ β_{CIA} CIA+a spline function”, which indicates that the true covariate CI is not included in the best model. The corresponding parameter estimates for this case are reported in Table 4, in which Z_j for $j = 1, \dots, 6$ denote the basis functions deduced from a 6-order spline.

5 Discussion

The mechanism-based ODE models and empirical parametric/nonparametric/semiparametric regression models are very popular for HIV biomarker data analysis. Each of these two completely different classes of models has its pros and cons. Based on an AIDS clinical data and simulated data analyses, our results suggest that the ODE model is more powerful to model the nonlinear relationship between treatment effects and virological response. The ODE model is also better to identify the significant factors for virological response although it is a little bit liberal and intends to include more factors or covariates in the model. The mixed-effects semiparametric regression model is very flexible to fit the virological response data, but it is too liberal in identifying correct factors for virological response. Additionally, this method may miss the correct factors. The ODE model is also biologically justifiable and good for predictions and simulations for various biological scenarios. The limitations of the ODE models include the extensive computational effort and requirement of biological assumptions that sometimes may not be easy to validate. As indicated, for the

simulated data analysis, a single simulation run was conducted to evaluate and validate the proposed models and methods due to the computational cost. Formal theoretical justifications and intensive simulation studies for comparing these two different kinds of models are warranted for future research.

Acknowledgments

Wu and Huang's research was partially supported by NIAID/NIH research grants AI052765, AI055290, and AI27658, and Liang's research was partially supported by NIAID/NIH research grants AI62247 and AI059773. The authors are grateful to the Editor and referees for their helpful comments and suggestions. We also thank Ms. Jeanne Holden-Wiltse for her editorial assistance.

References

1. Anderson, RM.; May, RM. Complex dynamical behavior in the interaction between HIV and the immune system. In: Goldbeter, A., editor. *Cell to Cell Signaling: From Experiments to Theoretical Models*. Academic Press; New York: 1989. p. 335-349.
2. Perelson, AS. Modeling the interaction of the immune system with HIV. In: Castillo-Chavez, C., editor. *Mathematical and Statistical Approaches to AIDS Epidemiology, Volume 83, Lecture Notes in Biomathematics*. Springer-Verlag; New York: 1989. p. 350-370.
3. Perelson AS, Nelson PW. Mathematical analysis of HIV-1 dynamics in vivo. *SIAM Review*. 1999; 41:3-44.
4. Nowak, MA.; May, RM. *Virus Dynamics: Mathematical Principles of Immunology and Virology*. Oxford University Press; Oxford: 2000.
5. Perelson AS. Modelling viral and immune system dynamics. *Nat Rev Immunol*. 2002; 2:28-36. [PubMed: 11905835]
6. Wu H. Statistical methods for HIV dynamic studies in AIDS clinical trials. *Statistical Methods in Medical Research*. 2005; 14:171-192. [PubMed: 15807150]
7. Tan, WY.; Wu, H. *Deterministic and Stochastic Models of AIDS Epidemics and HIV Infections with Intervention*. World Scientific; Singapore: 2005.
8. Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature*. 1995; 373:123-126. [PubMed: 7816094]
9. Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 dynamics *in vivo*: virion clearance rate, infected cell life-span, and viral generation time. *Science*. 1996; 271:1582-1586. [PubMed: 8599114]
10. Perelson AS, Essunger P, Cao Y, Vesanen M, Hurley A, Saksela K, Markowitz M, Ho DD. Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature*. 1997; 387:188-191. [PubMed: 9144290]
11. Wu H, Kuritzkes DR, McClernon DR, et al. Characterization of viral dynamics in human immunodeficiency virus type 1-infected patients treated with combination antiretroviral therapy: relationships to host factors, cellular restoration and virological endpoints. *Journal of Infectious Diseases*. 1999; 179(4):799-807. [PubMed: 10068574]
12. Wu H, Ding AA, de Gruttola V. Estimation of HIV dynamic parameters. *Statistics in Medicine*. 1998; 17:2463-2485. [PubMed: 9819839]
13. Wu H, Ding AA. Population HIV-1 dynamics *in vivo*: applicable models and inferential tools for virological data from AIDS clinical trials. *Biometrics*. 1999; 55:410-418. [PubMed: 11318194]
14. Wu H, Zhang JT. The study of long-term HIV dynamics using semiparametric nonlinear mixed-effects models. *Statistics in Medicine*. 2002; 21:3655-3675. [PubMed: 12436462]
15. Liang H, Wu H, Carroll RJ. The relationship between virologic and immunologic responses in AIDS clinical research using mixed-effects varying-coefficient semiparametric models with measurement error. *Biostatistics*. 2003; 4:297-312. [PubMed: 12925523]
16. Wu H, Liang H. Backfilling random varying-coefficient models with time-dependent smoothing covariates. *Scan J Statist*. 2004; 31:3-19.

17. Wu H, Zhao C, Liang H. Comparison of linear, nonlinear and semiparametric models for estimating HIV dynamic parameters. *Biometrical Journal*. 2004; 46:233–245.
18. Park JG, Wu H. Backfitting and local likelihood methods for nonparametric mixed-effects models with longitudinal data. *Journal of Statistical Planning and Inference*. 2006; 136:3760–3782.
19. Wu, H.; Zhang, J. *Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches*. John Wiley & Sons, Inc.; New York: 2006.
20. Huang Y, Liu D, Wu H. Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. *Biometrics*. 2006; 62:413–423. [PubMed: 16918905]
21. Molla A, Korneyeva M, Gao Q, Vasavanonda S, Schipper PJ, Mo HM, Markowitz M, Chernyavskiy T, Niu P, Lyons N, Hsu A, Granneman GR, Ho DD, Boucher CA, Leonard JM, Norbeck DW, Kempf DJ. Ordered accumulation of mutations in HIV protease confers resistance to ritonavir. *Nature Medicine*. 1996; 2:760–66.
22. Ickovics JR, Meisler AW. Adherence in AIDS clinical trial: a framework for clinical research and clinical care. *Journal of Clinical Epidemiology*. 1997; 50:385–391. [PubMed: 9179096]
23. Sheiner, LB. Modeling pharmacodynamics: parametric and nonparametric approaches. In: Rowland, M., et al., editors. *Variability in Drug Therapy: Description, Estimation, and Control*. Raven Press; New York: 1985. p. 139–152.
24. Han, C.; Chaloner, K.; Perelson, AS. Bayesian analysis of a population HIV dynamic model. In: Gatsoiquiry, C.; Kass, RE.; Carriquiry, A.; Gelman, A.; Higdon, D.; Pauler, DK.; Verdine, I., editors. *Case Studies in Bayesian Statistics*. Vol. 6. Springer-Verlag; New York: 2002. p. 223–237.
25. Davidian, M.; Giltinan, DM. *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall; London: 1995.
26. Engle RF, Granger CWJ, Rice J, Weiss A. Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*. 1986; 81:310–320.
27. Speckman P. Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B*. 1988; 50:413–436.
28. Härdle, W.; Liang, H.; Gao, JT. *Partially Linear Models*. Springer Physica; Heidelberg: 2000.
29. Lin DY, Ying Z. Semiparametric and nonparametric regression analysis of longitudinal data (with discussion). *Journal of the American Statistical Association*. 2001; 96:103–113.
30. Fan JQ, Li RZ. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association*. 2004; 99:710–723.
31. Zeger SL, Diggle PJ. Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics*. 1994; 50:689–699. [PubMed: 7981395]
32. Hoover DR, Rice JA, Wu CO, Yang LP. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*. 1998; 85:809–822.
33. Wu CO, Chiang CT, Hoover DR. Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *Journal of the American Statistical Association*. 1998; 93:1388–1402.
34. Zhang DW, Lin XH, Raz J, Sowers MR. Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association*. 1998; 93:710–719.
35. Ke CL, Wang YD. Semiparametric nonlinear mixed-effects models and their applications. *Journal of the American Statistical Association*. 2001; 96:1272–1281.
36. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics*. 1982; 38:963–974. [PubMed: 7168798]
37. Pinheiro, J.; Bates, D. *Mixed-effects models in S and S-plus*. Springer; New York: 2000.
38. Wu H, Zhang JT. Local polynomial mixed-effects models for longitudinal data. *Journal of the American Statistical Association*. 2002; 97:883–897.
39. Shi MG, Weiss RE, Taylor JMG. An analysis of pediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *Applied Statistics*. 1996; 45:151–163.
40. Lin XH, Carroll RJ. Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association*. 2000; 95:520–534.
41. Rice JA, Wu CO. Nonparametric mixed-effects models for unequally sampled noisy curves. *Biometrics*. 2001; 57:253–259. [PubMed: 11252607]

42. Vonesh, EF.; Chinchilli, VM. Linear and nonlinear models for the analysis of repeated measurements. Marcel Dekker; New York: 1997.
43. Ruppert D, Carroll RJ. Spatially-adaptive penalties for spline fitting. Australian and New Zealand Journal of Statistics. 2000; 42:205–223.
44. Eubank, RL. Nonparametric Regression and Spline Smoothing. Marcel Dekker; New York: 1999.
45. Acosta EP, Wu H, Hammer SM, Yu S, Kuritzkes DR, Walawander A, Eron JJ, Fichtenbaum CJ, Pettinelli C, Neath D, Ferguson E, Saah AJ, Gerber JG. Adult AIDS Clinical Trials Group 5055 Protocol Team. Comparison of two indinavir/ritonavir regimens in the treatment of HIV-infected individuals. Journal of Acquired Immune Deficiency Syndromes. 2004; 37:1358–1366. [PubMed: 15483465]
46. Wu H, Huang Y, Acosta EP, Rosenkranz SL, Kuritzkes DR, Eron JJ, Perelson AS, Gerber JG. Modeling long-term HIV dynamics and antiretroviral response: effects of drug potency, pharmacokinetics, adherence and drug resistance. Journal of Acquired Immune Deficiency Syndromes. 2005; 39:272–283. [PubMed: 15980686]
47. Wu H, Huang Y, Acosta EP, Park JG, Yu S, Rosenkranz SL, Kuritzkes DR, Eron JJ, Perelson AS, Gerber JG. Pharmacokinetics of antiretroviral agents in HIV-1 infected patients: using viral dynamic models that incorporate drug susceptibility and adherence. Journal of Pharmacokinetics and Pharmacodynamics. 2006; 33:399–419. [PubMed: 16583266]
48. Wakefield JC. The Bayesian analysis to population Pharmacokinetic models. Journal of the American Statistical Association. 1996; 91:62–75.

Figure 1.

The estimates of viral load trajectory from the model fitting with (solid lines) and without (dotted lines) considering $6I_{2h}$, $A(t)$ and $IC_{50}(t)$ in the ODE model for four subjects from A5055 study. The observed viral load values are indicated by circles.

Figure 2.

Sum of squared residuals (SSR) from individual subjects for 8 ODE models with all possible combinations of the three factors ϕ_{12h} , $IC_{50}(t)$ and $A(t)$ based on an AIDS clinical data. The sign test was used for pairwise comparisons between the control model and each of the other seven models, and the p-values are given at the top of the box.

Figure 3. The observed (circle symbols), population (dotted lines) and individual (solid lines) estimates of viral load from four selected subjects for the data set from the A5055 study for the CIA SPME model.

Figure 4.

Sum of squared residuals (SSR) from individual subjects for models with all possible combinations of the three factors (C_{12h} , $IC_{50}(t)$ and $A(t)$) based on the seven generated data sets. “0” in the x-label indicates the control model. “*” indicates the true model.

Figure 5.

Sum of squared residuals (SSR) from individual subjects for models with all possible combinations of the three factors (C_{12h} , $IC_{50}(t)$ and $A(t)$) based on the seven generated data sets. “0” in the x-label indicates the control model. “*” indicates the true model.

Table 1

Estimation results from the data of the A5055 study

model	coef	Value	Std. Error	t-value	p-value
A	β_A	5.691	0.179	31.763	$<10^{-4}$
I	β_I	13.660	0.473	28.867	$<10^{-4}$
C	β_C	4.039	0.147	27.382	$<10^{-4}$
IA	β_{IA}	13.697	0.504	27.202	$<10^{-4}$
CA	β_{CA}	4.046	0.151	26.867	$<10^{-4}$
CI	β_{CI}	5.841	0.195	29.964	$<10^{-4}$
CIA	β_{CIA}	5.891	0.205	28.733	$<10^{-4}$

Table 2

The significant variables selected using the different models and methods.

No.	True model	ODE method	Semiparametric Methods	
			univariate	multiple
1	A	A, CA, CI, CIA	I, IA	(A, CA)
2	I	I, CI, CIA	I, IA	(CI, IA, CIA)
3	C	C, CI, CIA	A, I, C, IA, CA, CI, CIA	(A, IA)
4	IA	I, IA, CI, CIA	I, IA	(CI, CIA)
5	CA	CA, CI, CIA	A, C, CA, CI, CIA	(C, IA, CA)
6	CI	CI, CIA	A, C, CA, CI, CIA	CIA
7	CIA	CA, CI, CIA	CI, CIA	CIA

Pairwise comparisons of sum of squared residuals (SSR) from individual subjects for 8 ODE models with all possible combinations of the three factors C_{12th} , and $A(t)$. The p -values were obtained using the sign test.

Table 3

	Models	Control	A	I	C	IA	CA	CI	CIA
p -values	A	0.2513							
	I	0.1083	0.4913						
	C	1.0000	0.6547	1.0000					
	IA	0.8185	0.0593	0.2513	1.0000				
	CA	1.0000	0.6547	1.0000	0.6547	1.0000			
	CI	0.0001	0.0001	0.0001	<0.0001	0.0001	<0.0001		
	CIA	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.0593	
SSR	Mean	8.58	7.91	10.98	7.30	11.28	6.81	0.22	0.80
	$\pm$ SD	16.72	17.05	17.50	12.90	18.63	10.42	0.03	0.04

Table 4

Estimation results of the SPME model for the simulated data from the model (CI)

	Value	SE	t-value	p-value
β_{CIA}	3.749	0.208	18.044	$< 10^{-4}$
Z_1	-0.874	0.211	-4.146	$< 10^{-4}$
Z_2	-1.009	0.211	-4.778	$< 10^{-4}$
Z_3	-1.052	0.210	-5.000	$< 10^{-4}$
Z_4	-0.993	0.209	-4.742	$< 10^{-4}$
Z_5	-1.257	0.214	-5.884	$< 10^{-4}$
Z_6	-0.888	0.210	-4.235	$< 10^{-4}$