Biostatistics (2003), **4**, 2, *pp.* 297–312
Printed in Great Britain

The relationship between virologic and immunologic responses in AIDS clinical research using mixed-effects varying-coefficient models with measurement error

HUA LIANG
Department of Biostatistics, St. Jude Children's Research Hospital, 332 North Lauderdale St., Memphis, TN 38105, USA

HULIN WU*
Frontier Science & Technology Research Foundation, 1244 Boylston Street, Suite 303, Chestnut Hill, MA 02467, USA
wu@sdac.harvard.edu

RAYMOND J. CARROLL
Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA

SUMMARY

In this article we study the relationship between virologic and immunologic responses in AIDS clinical trials. Since plasma HIV RNA copies (viral load) and CD4+ cell counts are crucial virologic and immunologic markers for HIV infection, it is important to study their relationship during HIV/AIDS treatment. We propose a mixed-effects varying-coefficient model based on an exploratory analysis of data from a clinical trial. Since both viral load and CD4+ cell counts are subject to measurement error, we also consider the measurement error problem in covariates in our model. The regression spline method is proposed for inference for parameters in the proposed model. The regression spline method transforms the unknown nonparametric components into parametric functions. It is relatively simple to implement using readily available software, and parameter inference can be developed from standard parametric models. We apply the proposed models and methods to an AIDS clinical study. From this study, we find an interesting relationship between viral load and CD4+ cell counts during antiviral treatments. Biological interpretations and clinical implications are discussed.

Keywords: AIDS clinical trial; Conditionally parametric model; Error-in-variables; Functional linear model; HIV dynamics; Longitudinal data; Measurement error; Regression splines; Time-varying coefficient model.

1. INTRODUCTION

Both virologic and immunologic surrogate markers such as plasma HIV RNA copies (viral load) and CD4+ cell counts currently play important roles in evaluating antiviral therapies in AIDS clinical research. Before HIV RNA assays were developed in the mid-1990s, CD4+ cell counts served as a primary surrogate

*To whom correspondence should be addressed

marker in AIDS clinical trials. Later plasma HIV RNA level (viral load) was shown to be more predictive of clinical outcomes (Saag *et al.*, 1996; Mellors *et al.*, 1995, 1996), and thus replaced CD4+ cell counts as a new primary surrogate marker in most AIDS clinical trials. However, recently some investigators have suggested that the combination of these two markers may be better and more appropriate for evaluating HIV/AIDS treatments (Padierna-Olivos *et al.*, 2000). This makes it important to study their relationship during treatment.

It is well known that CD4+ cells are targets of HIV and decline to a lower level after HIV infection. Thus, when antiviral therapies suppress viral load, CD4+ cell counts may recover to a higher level (Lederman *et al.*, 1998). In general, it is believed that the virologic response (measured by viral load) and immunologic response (measured by CD4+ cell counts) are negatively correlated during antiviral treatments. However, we speculate that their relationship may not be a constant during the whole period of treatment. In fact, the discordance between virologic and immunologic responses has been observed from several recent clinical studies (Sabin *et al.*, 2000; Mallolas *et al.*, 2000; Wu *et al.*, 2000).

Figure 1 presents simple linear regression plots of viral load (in $\log_{10}$ scale) versus CD4+ cell counts at different treatment times from an AIDS clinical study conducted by the AIDS Clinical Trials Group (ACTG 315). In this study, 48 evaluable HIV-1 infected patients were treated with potent antiviral therapy consisting of ritonavir, 3TC and AZT (see Lederman *et al.* (1998) and Wu *et al.* (1999) for more details on this study). Both viral load and CD4+ cell counts were monitored simultaneously at treatment days 0, 2, 7, 10, 14, 28, 56, 84, 168, and 336. We also present the raw data of CD4+ cell counts and viral load (as well as their mean curve estimates obtained from the proposed method in later sections) in the upper panel of Figure 2.

From Figure 1, we see that the viral load and CD4+ cell counts are negatively and approximately linearly related in most of the treatment times, but the regression coefficient is not a constant during the treatment period. This feature of the data motivates us to consider a varying-coefficient model (Cleveland *et al.*, 1991; Hastie and Tibshirani, 1993). When we look at the relationship between viral load and CD4+ cell counts in individual patients (Section 4.2 and Figure 5), we find a nonlinear relationship with a large between-patient variation. Thus, we propose a mixed-effects varying-coefficient model, instead of a standard varying-coefficient model, to capture both population and individual relationships between viral load and CD4+ cell counts during antiviral treatments.

The primary goal of this paper is to model the relationship between viral load and CD4+ cell counts in HIV-infected individuals during potent antiviral treatments based on the data from ACTG 315 (Lederman *et al.*, 1998; Wu *et al.*, 1999). Here we only focus on the data for the first 24 weeks of treatment, since virological or immunologic responses during this period are popular endpoints for many AIDS clinical trials. Our analysis based on the proposed models shows that there exists a strong association (inversely related) between HIV viral load and CD4+ T cell counts at the beginning of antiviral treatment. The strength of this association attenuates after initiation of potent antiviral treatment. The association gradually recovers after 8 weeks of the treatment. Our models also allow us to evaluate the individual association patterns among the study patients, which exhibits a large between-patient variation. These results indicate that it is important to monitor and evaluate both virologic and immunologic markers longitudinally in AIDS clinical studies due to the varying association between the two markers.

Hastie and Tibshirani (1993) proposed a general varying-coefficient model, $\eta = \beta_0 + X_1\beta_1(R_1) + \cdots + X_p\beta_p(R_p)$, where η is a parameter of the distribution of response variable Y, and the X are covariates. An important special case of this model is when the R are the same variable such as time,

$$\eta = \beta_0 + X_1(t)\beta_1(t) + \cdots + X_p(t)\beta_p(t). \tag{1.1}$$

This is a time-varying coefficient model, which is also called a 'dynamic generalized linear model' by West *et al.* (1985). For a given time t, this model is a simple parametric linear model, although the $\beta(t)$ are assumed to be nonparametric functions.

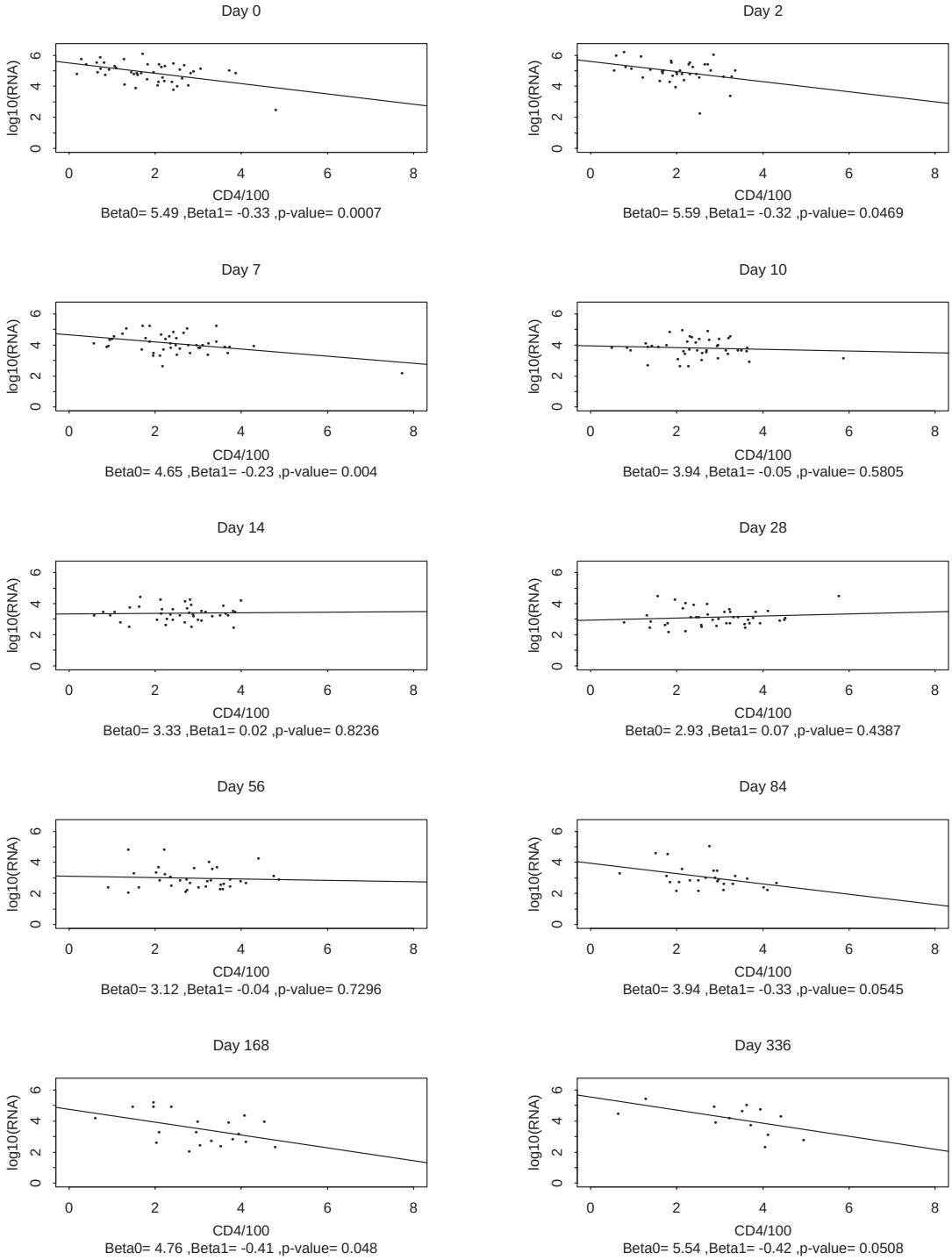

Fig. 1. Linear regression of viral load (HIV RNA) versus CD4+ cell counts at different treatment times. The fitted model at each time is a model $V = \beta_0 + CD4\beta_1 + e$, where the estimates of β_0, β_1, and p-values for testing $\beta_1 = 0$ are given for each of the plots. The data below the limit of quantification are excluded in the plots.

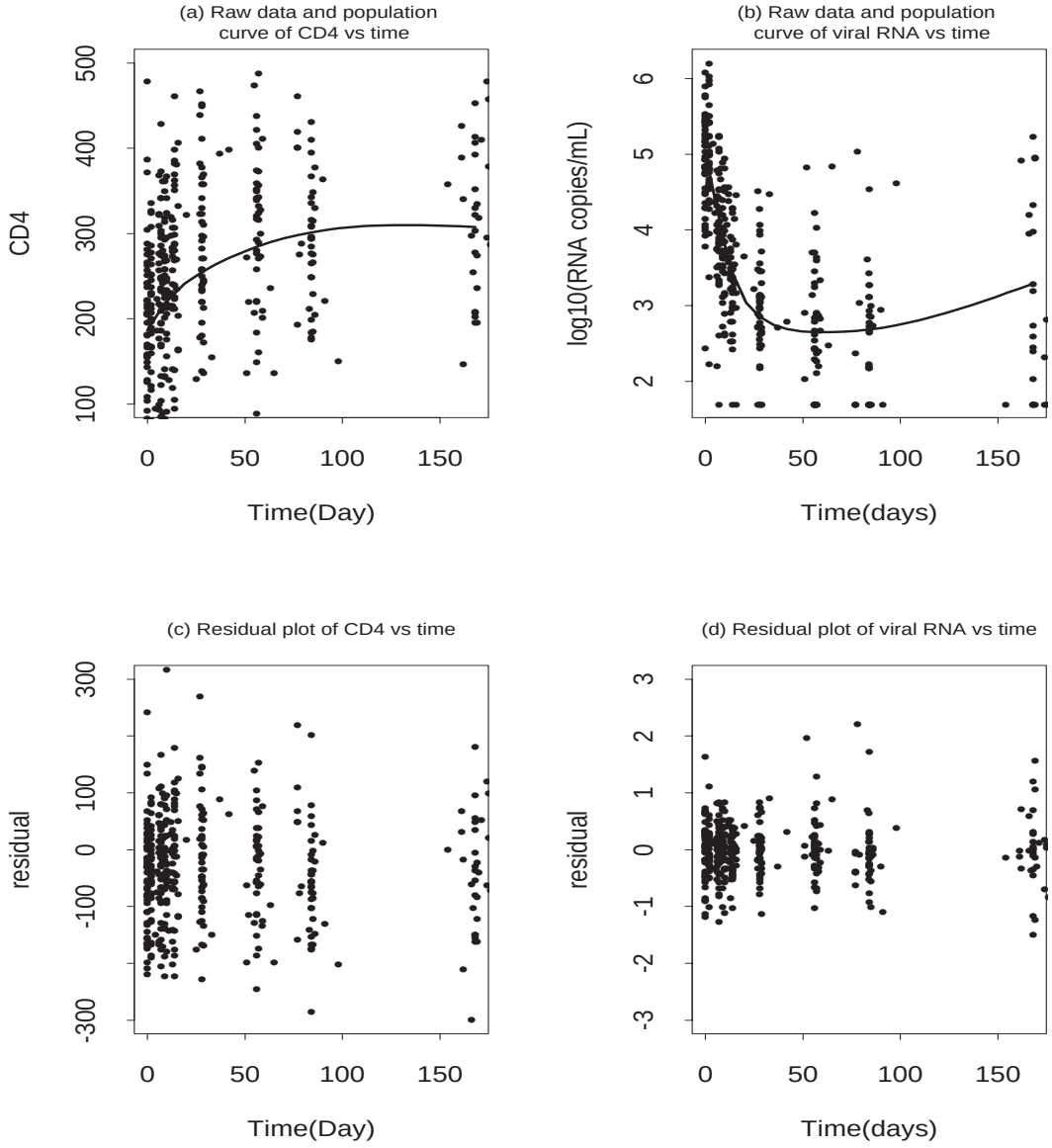

Fig. 2. Upper panel: CD4+ cell counts and viral load data, solid lines are the fitted curves from the proposed methods. Bottom panel: residual plots for CD4+ and RNA fit.

Recently, Hoover *et al.* (1998) and Wu *et al.* (1998) revisited this model for longitudinal data analysis. The standard time-varying coefficient model with a single covariate (Hoover *et al.*, 1998; Wu *et al.*, 1998) can be written as

$$y_{ij} = \beta_0(t_{ij}) + x_i(t_{ij})\beta_1(t_{ij}) + e_i(t_{ij}), \quad j = 1, \ldots, m_i, \quad i = 1, \ldots, n. \tag{1.2}$$

Both a smoothing spline method and a local polynomial kernel regression method were proposed in Hoover *et al.* (1998). Alternatively, Fan and Zhang (2000) proposed a two-step method for the same

model. However, none of these methods efficiently considered the important features of longitudinal data such as between-subject and within-subject variation and the special correlation structure of longitudinal data. Lin and Carroll (2000), however, show that using standard kernel methods to incorporate correlations is typically the wrong thing to do, as the methods inflate rather than deflate variability. Welsh *et al.* (2002) show that regression and smoothing splines do not suffer from this difficulty. In this paper we extend the time-varying coefficient model to include the random effects so that the association between CD4+ cell counts and viral load can be better quantified.

The remainder of the paper is organized as follows. In Section 2 we propose mixed-effects varying-coefficient models for the problem. Since both viral load and CD4+ cell counts are prone to measurement errors, we also consider measurement errors in covariates in our models. In Section 3 we consider measurement error calibration and propose regression spline methods for inferences. We apply the proposed methodologies to the HIV example that is the focus of this paper, and present the results in Section 4. The conclusions and some discussions are given in Section 5.

2. Mixed-effects varying-coefficient models with measurement error in covariates

For longitudinal data analysis, linear and nonlinear mixed-effects models have been proposed to incorporate the between-subject and within-subject variations: see, for example, Davidian and Giltinan (1995), Vonesh and Chinchilli (1996), and Pinheiro and Bates (2000). Since the relationship between viral load and CD4+ cell counts may be different for different subjects, we may naturally extend the standard varying-coefficient model (1.2) to a mixed-effects varying-coefficient model. Particularly for the study we introduced, we consider the time-varying coefficient $\beta_1(t)$, which is used to quantify the association between the CD4+ cell counts (covariate) and the viral load (response variable), to have a random component. That is, $\beta_{1i}(t) = \beta_1(t) + \gamma_i(t)$, where $\gamma_i(t)$ is a realization of a zero mean stochastic process $\gamma(t)$. Then the proposed model can be written as

$$y_{ij} = \beta_0(t_{ij}) + x_i(t_{ij})\beta_{1i}(t_{ij}) + e_i(t_{ij}), \quad j = 1, \ldots, m_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where y_{ij} is the viral load measurements in our HIV/AIDS clinical study (response variable) and $x_{ij} = x_i(t_{ij})$ is CD4+ cell counts for the ith subject at time t_{ij}. The error term $e_i(t)$ is a zero mean stochastic process with covariance function $\rho_e(s, t) = \mathrm{cov}\{e_i(s), e_i(t)\}$. Following the convention of mixed-effects models, we further assume that $e_i(t)$ and $\gamma_i(t)$ are independent, and $\{e_i(t), \gamma_i(t)\}$ and $\{e_k(t), \gamma_k(t)\}$ are independent for $i \neq k$. This model will be used to analyse the data from ACTG 315 study in Section 4. Note that the proposed model is similar in spirit to those in Zeger and Diggle (1994) and Zhang *et al.* (1998). However, a key difference is that these existing models did not consider the random time-varying coefficient.

Let $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{im_i})^T$, $\boldsymbol{y}_i = \{y_i(t_{i1}), \ldots, y_i(t_{im_i})\}^T$, $\beta_0(\boldsymbol{t}_i) = \{\beta_0(t_{i1}), \ldots, \beta_0(t_{im_i})\}^T$, $\beta_{1i}(\boldsymbol{t}_i) = \{\beta_{1i}(t_{i1}), \ldots, \beta_{1i}(t_{im_i})\}^T$, $\boldsymbol{x}_i(\boldsymbol{t}_i) = \{x_i(t_{i1}), \ldots, x_i(t_{im_i})\}^T$, and $\boldsymbol{e}_i = \boldsymbol{e}_i(\boldsymbol{t}_i) = \{e_i(t_{i1}), \ldots, e_i(t_{im_i})\}^T$. Then, the model (2.1) can be written in vector notation as

$$\boldsymbol{y}_i = \beta_0(\boldsymbol{t}_i) + \boldsymbol{x}_i(\boldsymbol{t}_i) * \beta_{1i}(\boldsymbol{t}_i) + \boldsymbol{e}_i, \quad i = 1, \ldots, n, \tag{2.2}$$

where $\boldsymbol{x}_i(\boldsymbol{t}_i) * \beta_{1i}(\boldsymbol{t}_i) = \{x_i(t_{i1})\beta_1(t_{i1}), \ldots, x_i(t_{im_i})\beta_1(t_{im_i})\}^T$, and $\beta_{1i}(\boldsymbol{t}_i) = \beta_1(\boldsymbol{t}_i) + \gamma_i(\boldsymbol{t}_i)$. Our primary interest is to estimate both population time-varying coefficient $\beta_1(t)$ and individual coefficients $\beta_{1i}(t)$. Notice that the proposed mixed-effects varying-coefficient model can also be regarded as an extension of semiparametric models or partial linear models for longitudinal data (Zeger and Diggle, 1994; Zhang *et al.*, 1998).

Although the time-varying intercept $\beta_0(t)$ is not our primary interest and is not considered as random in our study, the methodology that we will develop in the following is still applicable (with minor modifications) when the $\beta_0(t)$ also has a random component.

Since CD4+ cell counts are also measured with error, we need to consider measurement error in covariates in the proposed model (2.1). That is, we have a measurement error model as follows:

$$w_i(t) = x_i(t) + u_i(t), \tag{2.3}$$

where $w_i(t)$ is the observed CD4+ cell counts, and $x_i(t)$ is underlying true CD4+ cell counts for the ith patient at treatment time t. The error $u_i(t)$ represents measurement error in CD4+ cell counts. We assume that $u_i(t)$ is a mean zero process, and $\{x_i(t), u_i(t), \gamma_i(t), e_i(t)\}$ are mutually independent. Fuller (1987) and Carroll *et al.* (1995) gave a good survey on measurement error models. Tosteson *et al.* (1997) and Buonaccorsi *et al.* (2000) studied measurement errors in linear mixed-effects models (LMEs). Higgins *et al.* (1997) proposed a two-step approach to deal with measurement errors in nonlinear mixed-effects models. Wang *et al.* (1998) studied generalized linear mixed measurement error models.

Measurement error adjustment generally requires replications, validation data, or other information to estimate the error structure, none of which we have. However, we have repeated measurements on CD4+ cell counts. The repeatedly measured data along time t are similar to replications if we can assume that the measured variable is a smooth function of t. Shi *et al.* (1996) and Rice and Wu (2001) have modeled the natural history (untreated HIV-infected patients) of the CD4+ cell process using a mixed-effects regression spline model. We use a similar idea to model the CD4+ cell process as follows:

$$\boldsymbol{x}_i(\boldsymbol{t}_i) = \sum_{k=0}^{p} \xi_k \psi_k(t) + \sum_{k=0}^{q} \eta_{ik} \phi_k(t) = \boldsymbol{A}_i(\boldsymbol{t}_i)\boldsymbol{\alpha} + \boldsymbol{R}_i(\boldsymbol{t}_i)\boldsymbol{\zeta}_i, \tag{2.4}$$

where $\boldsymbol{A}_i(\boldsymbol{t}_i) = \{\psi_0(t), \psi_1(t), \ldots, \psi_p(t)\}$ and $\boldsymbol{R}_i(\boldsymbol{t}_i) = \{\phi_0(t), \phi_1(t), \ldots, \phi_q(t)\}$ are basis functions such as cubic B-spline basis, $\boldsymbol{\alpha} = (\xi_0, \xi_1, \ldots, \xi_p)^T$ is a fixed-effect parameter vector, and $\boldsymbol{\zeta}_i = (\eta_{i0}, \eta_{i1}, \ldots, \eta_{iq})^T$ is a random effect vector with mean zero and covariance matrix Σ_ζ (Σ_ζ may be unstructured or can be specified with a special structure). The selection of the number (p and q) and locations of knots for regression splines can be based on the suggestions in Eubank (1999) and Ramsay (1988). Rice and Wu (2001) suggest setting $p = q$ and $\psi_k(t) = \phi_k(t)$. Model (2.4) is a standard LME model, which can be fitted using either Splus (LME) or SAS (PROC MIXED). The population curve estimate of CD4+ cell counts from ACTG 315 is plotted in Figure 2(a).

3. ESTIMATION AND INFERENCE

3.1 *Measurement error calibration*

Higgins *et al.* (1997) have proposed a two-step approach to deal with measurement errors in time-dependent covariates in nonlinear mixed-effects models. The first step is to estimate the covariate function by fitting an appropriate model for covariate processes, and then fit the nonlinear mixed-effects model by plugging-in the estimates of covariates in the second step. This is similar in spirit to the regression calibration idea (Carroll *et al.*, 1995). By applying a similar idea to our model for measurement error in CD4+ cell counts, we may fit a model,

$$\boldsymbol{w}_i(\boldsymbol{t}_i) = \boldsymbol{A}_i(\boldsymbol{t}_i)\boldsymbol{\alpha} + \boldsymbol{R}_i(\boldsymbol{t}_i)\boldsymbol{\zeta}_i + \boldsymbol{u}_i(\boldsymbol{t}_i), \tag{3.1}$$

and obtain the estimates of $\boldsymbol{\alpha}$, $\boldsymbol{\zeta}_i$ and Σ_ζ, denoted by $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\zeta}}_i$ and $\widehat{\Sigma}_\zeta$, where

$$\widehat{\boldsymbol{\alpha}} = \left\{ \sum_{i=1}^{n} \boldsymbol{A}_i(\boldsymbol{t}_i)^T \Sigma_{w_i}^{-1} \boldsymbol{A}_i(\boldsymbol{t}_i) \right\}^{-1} \left\{ \sum_{i=1}^{n} \boldsymbol{A}_i(\boldsymbol{t}_i)^T \Sigma_{w_i}^{-1} \boldsymbol{w}_i(\boldsymbol{t}_i) \right\}$$

and

$$\widehat{\boldsymbol{\zeta}}_i = \Sigma_\zeta \boldsymbol{R}_i(\boldsymbol{t}_i)^T \Sigma_{w_i}^{-1}\{\boldsymbol{w}_i(\boldsymbol{t}_i) - \boldsymbol{A}_i(\boldsymbol{t}_i)\widehat{\boldsymbol{\alpha}}\}.$$

Thus, we can get an estimator for CD4+ cell count trajectory for each individual,

$$\begin{aligned}\widehat{\boldsymbol{x}}_i(\boldsymbol{t}_i) &= \boldsymbol{A}_i(\boldsymbol{t}_i)\widehat{\boldsymbol{\alpha}} + \boldsymbol{R}_i(\boldsymbol{t}_i)\widehat{\boldsymbol{\zeta}}_i \\ &= \{\boldsymbol{I} - \boldsymbol{R}_i(\boldsymbol{t}_i)\Sigma_\zeta \boldsymbol{R}_i(\boldsymbol{t}_i)^T \Sigma_{w_i}^{-1}\}\boldsymbol{A}_i(\boldsymbol{t}_i)\widehat{\boldsymbol{\alpha}} + \boldsymbol{R}_i(\boldsymbol{t}_i)\Sigma_\zeta \boldsymbol{R}_i(\boldsymbol{t}_i)^T \Sigma_{w_i}^{-1}\boldsymbol{w}_i(\boldsymbol{t}_i), \end{aligned} \tag{3.2}$$

which is the best linear unbiased predictor (BLUP). Then, similar to the method of Higgins *et al.* (1997), we may substitute the estimated covariates in the mixed-effects varying coefficient model (2.2) and proceed by treating the estimated covariates as true covariates, i.e.

$$\boldsymbol{y}_i = \beta_0(\boldsymbol{t}_i) + \widehat{\boldsymbol{x}}_i(\boldsymbol{t}_i) * \beta_{1i}(\boldsymbol{t}_i) + \boldsymbol{\varepsilon}_i(\boldsymbol{t}_i), \tag{3.3}$$

where $\boldsymbol{y}_i$ is viral load measurements and $\widehat{\boldsymbol{x}}_i(\boldsymbol{t}_i)$ is smoothed CD4+ cell count trajectory for the ith subject.

The second method for dealing with the measurement errors in covariates is to use the conditional distribution of $\boldsymbol{y}_i|\boldsymbol{w}_i$ under Normal assumptions. Denote $\mathrm{E}\{\boldsymbol{x}_i(\boldsymbol{t}_i)\} = \boldsymbol{\mu}_x(\boldsymbol{t}_i)$, $\mathrm{cov}\{\boldsymbol{x}_i(\boldsymbol{t}_i)\} = \Sigma_{x_i}$, and $\mathrm{cov}\{\boldsymbol{u}_i(\boldsymbol{t}_i)\} = \Sigma_{u_i}$. By equations (2.2)–(3.1), we have, $\boldsymbol{\mu}_{y|\beta_{1i}} = E(\boldsymbol{y}_i|\beta_{1i}) = \beta_0(\boldsymbol{t}_i) + \beta_{1i}(\boldsymbol{t}_i)\boldsymbol{\mu}_x(\boldsymbol{t}_i)$, $\boldsymbol{\mu}_w = E(\boldsymbol{w}_i) = \boldsymbol{\mu}_x(\boldsymbol{t}_i)$, $\Sigma_{w_i} = \mathrm{cov}(\boldsymbol{w}_i) = \Sigma_{u_i} + \boldsymbol{R}_i(\boldsymbol{t}_i)\Sigma_\zeta \boldsymbol{R}_i(\boldsymbol{t}_i)^T$, and $\Sigma_{yw|\beta_{1i}} = \mathrm{cov}\{(\boldsymbol{y}_i, \boldsymbol{w}_i)|\beta_{1i}\} = \boldsymbol{B}_{1i}\Sigma_{x_i}$, where $\boldsymbol{B}_{1i} = \mathrm{diag}\{\beta_{1i}(t_{i1}), \ldots, \beta_{1i}(t_{im_i})\}$. We assume that $\boldsymbol{y}_i$ and $\boldsymbol{w}_i$ follow a joint Normal distribution, conditional on β_{1i}. Then the conditional mean of $\boldsymbol{y}_i|(\boldsymbol{w}_i, \beta_{1i})$ is $\boldsymbol{\mu}_{y|w} = \boldsymbol{\mu}_{y|\beta_{1i}} + \Sigma_{yw|\beta_{1i}}\Sigma_{w_i}^{-1}(\boldsymbol{w}_i - \boldsymbol{\mu}_w)$ and the conditional covariance is $\Sigma_{y|w} = \Sigma_{y|\beta_{1i}} - \Sigma_{yw|\beta_{1i}}\Sigma_{w_i}^{-1}\Sigma_{yw|\beta_{1i}}^T$. Then a conditional regression model, given $\{\boldsymbol{w}_i, \beta_{1i}\}$, can be written as

$$\boldsymbol{y}_i = \beta_0(\boldsymbol{t}_i) + \beta_{1i}(\boldsymbol{t}_i) * \boldsymbol{\mu}_x(\boldsymbol{t}_i) + (\boldsymbol{B}_{1i}\Sigma_{x_i}\Sigma_{w_i}^{-1}) * (\boldsymbol{w}_i - \boldsymbol{\mu}_w) + \text{error},$$

where error is an error vector with mean zero and covariance matrix $\Sigma_{y|w}$. Also note that $*$ indicates the elementwise product of two vectors. Using $\Sigma_{x_i} = \boldsymbol{R}_i(\boldsymbol{t}_i)\Sigma_\zeta \boldsymbol{R}_i(\boldsymbol{t}_i)^T$ and replacing $\boldsymbol{\alpha}$ by $\widehat{\boldsymbol{\alpha}}$, this model can be written as

$$\begin{aligned}\boldsymbol{y}_i = \beta_0(\boldsymbol{t}_i) + \beta_{1i}(\boldsymbol{t}_i) * \{&[\boldsymbol{I} - \boldsymbol{R}_i(\boldsymbol{t}_i)\Sigma_\zeta \boldsymbol{R}_i(\boldsymbol{t}_i)^T \Sigma_{w_i}^{-1}]\boldsymbol{A}_i(\boldsymbol{t}_i)\widehat{\boldsymbol{\alpha}} \\ &+ \boldsymbol{R}_i(\boldsymbol{t}_i)\Sigma_\zeta \boldsymbol{R}_i(\boldsymbol{t}_i)^T \Sigma_{w_i}^{-1}\boldsymbol{w}_i(\boldsymbol{t}_i)\} + \boldsymbol{\varepsilon}_i. \end{aligned} \tag{3.4}$$

Then we can obtain the estimates of parameters $\beta_0(\boldsymbol{t}_i)$ and $\beta_{1i}(\boldsymbol{t}_i)$ using the methods proposed in the next section. Interestingly, if we plug the estimate of $\boldsymbol{x}_i(\boldsymbol{t}_i)$ from (3.2) into model (3.3), we exactly get model (3.4). This means that the foregoing two measurement error methods, the two-step substitution approach and the conditional distribution approach, are equivalent.

3.2 *Regression spline estimation*

Shi *et al.* (1996) and Rice and Wu (2001) have proposed regression spline methods to approximate nonparametric mixed-effect curves. The regression spline methods transform a nonparametric curve into a linear combination of basis functions so that model-fitting and inference can be based on standard parametric methods operationally.

Following similar ideas as in Shi *et al.* (1996) and Rice and Wu (2001), we approximate nonparametric functions $\beta_0(t)$, $\beta_1(t)$, and $\gamma_i(t)$ by basis functions. For simplicity, we use the same basis functions for

$\beta_0(t)$, $\beta_1(t)$, and $\gamma_i(t)$, which can then be approximated as

$$\beta_{0,p}(t) = \sum_{k=0}^{p} \xi_k \theta_k(t) = \Theta_p(t)^T \boldsymbol{\xi}_p,$$

$$\beta_{1,q}(t) = \sum_{k=0}^{q} \eta_k \phi_k(t) = \Phi_q(t)^T \boldsymbol{\eta}_q,$$

$$\gamma_{i,r}(t) = \sum_{k=0}^{r} b_{ik} \psi_k(t) = \Psi_r(t)^T \boldsymbol{b}_i.$$

Denote $\boldsymbol{\xi}_p = (\xi_0, \ldots, \xi_p)^T$, $\boldsymbol{\eta}_q = (\eta_0, \ldots, \eta_q)^T$, and $\boldsymbol{b}_i = (b_{i0}, \ldots, b_{ir})^T$. Also note that for fixed p, q and r, the truncated vector $\boldsymbol{b}_i = (b_{i0}, \ldots, b_{ir})^T$ is a random vector with mean 0 and covariance matrix D, $\rho(s,t) = \text{cov}\{\gamma_i(s), \gamma_i(t)\} = \Psi_r(t)^T D \Psi_r(t)$. Replacing $\{\beta_0(t), \beta_{1i}(t)\}$ by $\{\beta_{0,p}(t), \beta_{1,q}(t) + \gamma_{i,r}(t)\}$ in model (2.2), we obtain an approximate model

$$\boldsymbol{y}_i(\boldsymbol{t}_i) = \sum_{k=0}^{p} \xi_k \theta_k(\boldsymbol{t}_i) + \boldsymbol{x}_i(\boldsymbol{t}_i) * \sum_{k=0}^{q} \eta_k \phi_k(\boldsymbol{t}_i) + \boldsymbol{x}_i(\boldsymbol{t}_i) * \sum_{k=0}^{r} b_{ik} \psi_k(\boldsymbol{t}_i) + \boldsymbol{e}_i(\boldsymbol{t}_i). \tag{3.5}$$

For given p, q and r, this is a standard LME model with fixed-effect $\sum_{k=0}^{p} \xi_k \theta_k(\boldsymbol{t}_i) + \boldsymbol{x}_i(\boldsymbol{t}_i) * \sum_{k=0}^{q} \eta_k \phi_k(\boldsymbol{t}_i)$ and random-effect $\boldsymbol{x}_i(\boldsymbol{t}_i) * \sum_{k=0}^{r} b_{ik} \psi_k(\boldsymbol{t}_i)$, where $*$ indicates the elementwise product of two vectors. Let $\boldsymbol{X}_i = \{\theta_0(\boldsymbol{t}_i), \ldots, \theta_p(\boldsymbol{t}_i); \boldsymbol{x}_i(\boldsymbol{t}_i) * \phi_0(\boldsymbol{t}_i), \ldots, \boldsymbol{x}_i(\boldsymbol{t}_i) * \phi_q(\boldsymbol{t}_i)\}$, $\boldsymbol{Z}_i = \{\boldsymbol{x}_i(\boldsymbol{t}_i) * \psi_0(\boldsymbol{t}_i), \ldots, \boldsymbol{x}_i(\boldsymbol{t}_i) * \psi_r(\boldsymbol{t}_i)\}$, $\boldsymbol{\alpha} = (\boldsymbol{\xi}_p^T, \boldsymbol{\eta}_q^T)^T = (\xi_0, \ldots, \xi_p, \eta_0, \ldots, \eta_q)^T$. Model (3.5) can be expressed as a standard LME model for given p, q and r,

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\alpha} + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{e}_i(\boldsymbol{t}_i), \quad \boldsymbol{b}_i \sim (0, D), \quad \boldsymbol{e}_i \sim (0, R_i).$$

Thus, for given D and R_i, the closed forms for the estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{b}_i$ can be written as follows (Laird and Ware, 1982; Davidian and Giltinan, 1995; Vonesh and Chinchilli, 1996):

$$\widehat{\boldsymbol{\alpha}} = \left(\sum_{i=1}^{n} \boldsymbol{X}_i^T \Sigma_i^{-1} \boldsymbol{X}_i \right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{X}_i^T \Sigma_i^{-1} \boldsymbol{y}_i \right), \qquad \widehat{\boldsymbol{b}}_i = D \boldsymbol{Z}_i^T \Sigma_i^{-1} \left(\boldsymbol{y}_i - \boldsymbol{X}_i \widehat{\boldsymbol{\alpha}} \right),$$

where $\Sigma_i = R_i + \boldsymbol{Z}_i D \boldsymbol{Z}_i^T$. Consequently, the estimates of $\beta_{0,p}(t)$, $\beta_{1,q}(t)$ and $\gamma_{i,r}(t)$ can be expressed as $\widehat{\beta}_{0,p}(t) = \Theta_p(t)^T \widehat{\boldsymbol{\xi}}_p$, $\widehat{\beta}_{1,q}(t) = \Phi_q(t)^T \widehat{\boldsymbol{\eta}}_q$, and $\widehat{\gamma}_{i,r}(t) = \Psi_r(t)^T \widehat{\boldsymbol{b}}_i$. The covariance matrix D may be specified as unstructured or with a special structure. The covariance matrix R_i may also have a special structure, but very often we simply set $R_i = \sigma^2 \boldsymbol{I}_{n_i}$, where σ^2 needs to be estimated. The unknown parameters in D and R_i can be estimated using the maximum likelihood (ML) or restricted maximum likelihood (RML) method (Davidian and Giltinan, 1995; Vonesh and Chinchilli, 1996). Also note that $\widehat{\rho}(s,t) = \widehat{\text{cov}}\{\gamma_i(s), \gamma_i(t)\} = \Psi_r(t)^T \widehat{D} \Psi_r(t)$.

The choice of basis functions is usually not as crucial as the determination of smoothing parameters p, q and r. We use a natural cubic spline basis due to its optimality property and easy implementation using existing software such as SPLUS (Green and Silverman, 1994). Ramsay (1988), Eubank (1999) and others have proposed locating the knots at the quantiles of the data. To determine the smoothing parameters p, q and r, model selection criteria such as AIC, BIC, and cross-validation methods may be used.

3.3 *Large-sample inference*

In general, the fixed-effect estimator $\widehat{\boldsymbol{\alpha}}$ is consistent, and this is still true when the finite number of unknown parameters in covariance matrices D and R_i are substituted by their consistent estimators such

as MLEs (Vonesh and Chinchilli, 1996). Under Normality assumptions, one can show that $\sqrt{n}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}) \xrightarrow{d} N(0, \Omega)$, where $\boldsymbol{\alpha} = (\boldsymbol{\xi}_p^T, \boldsymbol{\eta}_q^T)^T$ and $\Omega = \{\lim_{n\to\infty}(n^{-1}\sum_{i=1}^n \boldsymbol{X}_i^T \Sigma_i^{-1} \boldsymbol{X}_i)\}^{-1}$, provided the limit exists (Vonesh and Chinchilli, 1996, Section 6.2.3). For given p, q and basis functions, we thus have that $\Theta_p(t)^T\widehat{\boldsymbol{\xi}}_p$ and $\Phi_q(t)^T\widehat{\boldsymbol{\eta}}_q$ are consistent estimators of $\beta_{0,p}(t)$ and $\beta_{1,q}(t)$. Furthermore, under Normality assumptions, we have

$$\sqrt{n}\{\Theta_p(t)^T\widehat{\boldsymbol{\xi}}_p - \beta_{0,p}(t)\} \xrightarrow{d} N\{0, \Theta_p(t)^T \Sigma_\xi \Theta_p(t)\},$$

$$\sqrt{n}\{\Phi_q(t)^T\widehat{\boldsymbol{\eta}}_q - \beta_{1,q}(t)\} \xrightarrow{d} N[0, \Phi_q(t)^T \Sigma_\eta \Phi_q(t)],$$

where Σ_ξ and Σ_η are corresponding block diagonal matrices of Ω with $p+1$ and $q+1$ dimensions respectively.

If nonparametric functions $\beta_0(t)$ and $\beta_1(t)$ are smooth enough and smoothing parameters p and q are appropriately selected, $\beta_{0,p}(t)$ and $\beta_{1,q}(t)$ will approximate $\beta_0(t)$ and $\beta_1(t)$ very well. That is, we can ignore the basis function approximation error at least for many practical purposes. Thus, inferences regarding $\beta_0(t)$ and $\beta_1(t)$ can be based on $\beta_{0,p}(t)$ and $\beta_{1,q}(t)$. Exact tests and confidence intervals for $\beta_{0,p}(t)$ and $\beta_{1,q}(t)$ are not available under the mixed-effects models unless the data are balanced and complete (Vonesh and Chinchilli, 1996). Instead, inferences for mixed-effects models have to be based on the aforementioned asymptotic results.

For example, we are interested in testing whether the association between viral load and CD4+ cell responses is a function of treatment time t or a constant, that is, we want to test the hypothesis

$$H_0 : \beta_1(t) = \text{Constant} \;\text{ versus }\; H_a : \beta_1(t) \neq \text{Constant}.$$

Under the above regression spline approximation, the hypothesis '$\beta_1(t)$ =constant' is approximately equivalent to '$\beta_{1,q}(t)$ =constant' or '$\eta_k = 0$ for $k \geqslant 1$' in the LME model (3.5). Then, the standard Wald chi-square test can be used to test this hypothesis (Davidian and Giltinan, 1995; Vonesh and Chinchilli, 1996):

$$T_\eta^2 = \widehat{\boldsymbol{\eta}}_{q-1}^T \widehat{\Sigma}_{\boldsymbol{\eta}_{q-1}}^{-1} \widehat{\boldsymbol{\eta}}_{q-1} \sim \text{approx.} \quad \chi_{q-1}^2,$$

where $\boldsymbol{\eta}_{q-1} = (\eta_1, \ldots, \eta_q)^T$, and $\widehat{\boldsymbol{\eta}}_{q-1}$ and $\widehat{\Sigma}_{\boldsymbol{\eta}_{q-1}}$ are the estimators of $\boldsymbol{\eta}_{q-1}$ and the covariance of $\widehat{\boldsymbol{\eta}}_{q-1}$. Similarly, we can test whether the correlation exists at all, that is to test $\beta_1(t) = 0$ for all t, which is approximately equivalent to testing '$\eta_k = 0$ for $k \geqslant 0$'. For large samples, the Wald test statistic T_η^2 is approximated well by a χ^2 distribution. The bootstrap method may be used if the sample size is not large enough.

4. Data analysis

In this section, we present the analysis results for ACTG 315 data. In this study, both viral load and CD4+ cell counts were scheduled to be measured on days $t = 0, 2, 7, 10, 14, 28, 56, 84, 168$ after initiation of an antiviral therapy. We obtained 441 complete pairs of viral load and CD4+ cell count observations from 48 evaluable patients. The number of data points on individual patients ranges from 3 to 9.

First we deal with measurement errors in the covariate, CD4+ cell counts. As suggested in Section 3.1, we fit model (3.1) to the CD4+ cell count data using the natural cubic spline basis. Then the substitution method (Section 3.1) is used to accommodate the covariate measurement errors in the mixed-effects varying coefficient model. We fitted the viral load and CD4+ cell count data in the four scenarios: (i) considering $\beta_1(t)$ as random and allowing measurement error in covariate $x_i(t)$; (ii) considering $\beta_1(t)$

as random, but no measurement error in covariate $x_i(t)$; (iii) considering $\beta_1(t)$ as fixed and allowing measurement error in covariate $x_i(t)$; (iv) considering $\beta_1(t)$ as fixed, but no measurement error in covariate $x_i(t)$. The smoothing parameters (the number of knots) are determined by the model selection criterion BIC and the location of knots is selected at the quantiles of the data (Ramsay, 1988; Eubank, 1999). We obtained that $p = q = r = 3$. To stabilize the variance and computational algorithms, we used $\log_{10}$ scale in viral load (this is commonly used in AIDS clinical trials), centered the covariate CD4+ cell counts, and took a log-transformation for time t in our model fitting. The viral load data below the limit of quantification (100 copies per ml plasma) are imputed by the mid-value of the quantification limit (50 copies per ml plasma). Since only 29 out of 441 (6.6%) measurements are below the limit of assay detection, we do not expect a large effect of these left-censored data on the analysis results. We fitted the model using the Splus LME function. The goodness-of-fit plots as outputs of the function show that the model fits the data very well (also see Figure 2).

Also note that, as suggested by one of the referees, the square-root transformation of CD4 data may be more appropriate for the assumptions of Normality and homogeneous variance. We also repeated our analysis using the square-root of CD4 data. However, the results are quite similar to those using the original CD4 data. This is probably because the departure from the Normality and homogeneous variance assumption is not serious in the original CD4 data. Thus, for better interpretation, we only report the results based on the original CD4 data.

4.1 *Results for population estimates*

We plot the estimates of $\beta_0(t)$ and $\beta_1(t)$ in Figure 3. Both cases, with/without measurement error in the covariate, are given. We can see that the estimates of $\beta_0(t)$ and $\beta_1(t)$ from different methods have similar shapes, although differences in the magnitude of the estimates from different methods can be seen clearly, especially in the estimates of $\beta_1(t)$. The results show that the viral load and CD4+ T cell count responses are inversely related. The association between the virologic and immunologic responses, measured by $\beta_1(t)$, is stronger at the beginning of the treatment, but gradually dampens to become weakest at week 4. However, after week 4, the association is recovered and finally becomes strongest at week 24.

We compared the estimation results between the cases with/without considering measurement error in CD4+ cell counts. We found that the coefficient estimates attenuate toward zero when the measurement error in the covariate is not considered. This is similar to the case in standard linear or nonlinear regression models with measurement error (Carroll *et al.*, 1995). Thus, the stronger relationship between viral load and CD4+ cell counts can be established when the measurement error in CD4+ cell counts is calibrated by a smoothing method as introduced in Section 3.1. When we compare the results (data not shown) between the cases with/without considering the random component in $\beta_1(t)$, interestingly we see that without considering a random-effect in $\beta_1(t)$, the estimates of coefficients also attenuate toward zero, which is similar to the effect of measurement error in the covariate. Thus, by considering a random-effect in $\beta_1(t)$, we can efficiently establish a significant relationship between viral load and CD4+ cell counts. Note that the wiggle at the boundary is exaggerated (Figure 3(d)) when the measurement error in the covariate is calibrated. This is because the attenuating factor (reliability ratio) for the noisy covariate CD4+ cell counts is very small and the variation is large at the boundary.

As also suggested by one of the referees, in order to confirm our conclusions we repeated the above analysis after changing the role of viral load and CD4+ cell counts, i.e. taking CD4+ cell counts as the response variable and viral load as the covariate, or fit a model

$$CD4_i(t) = \beta_i^*(t) + RNA_i(t)\beta_{1i}^*(t) + error_i.$$

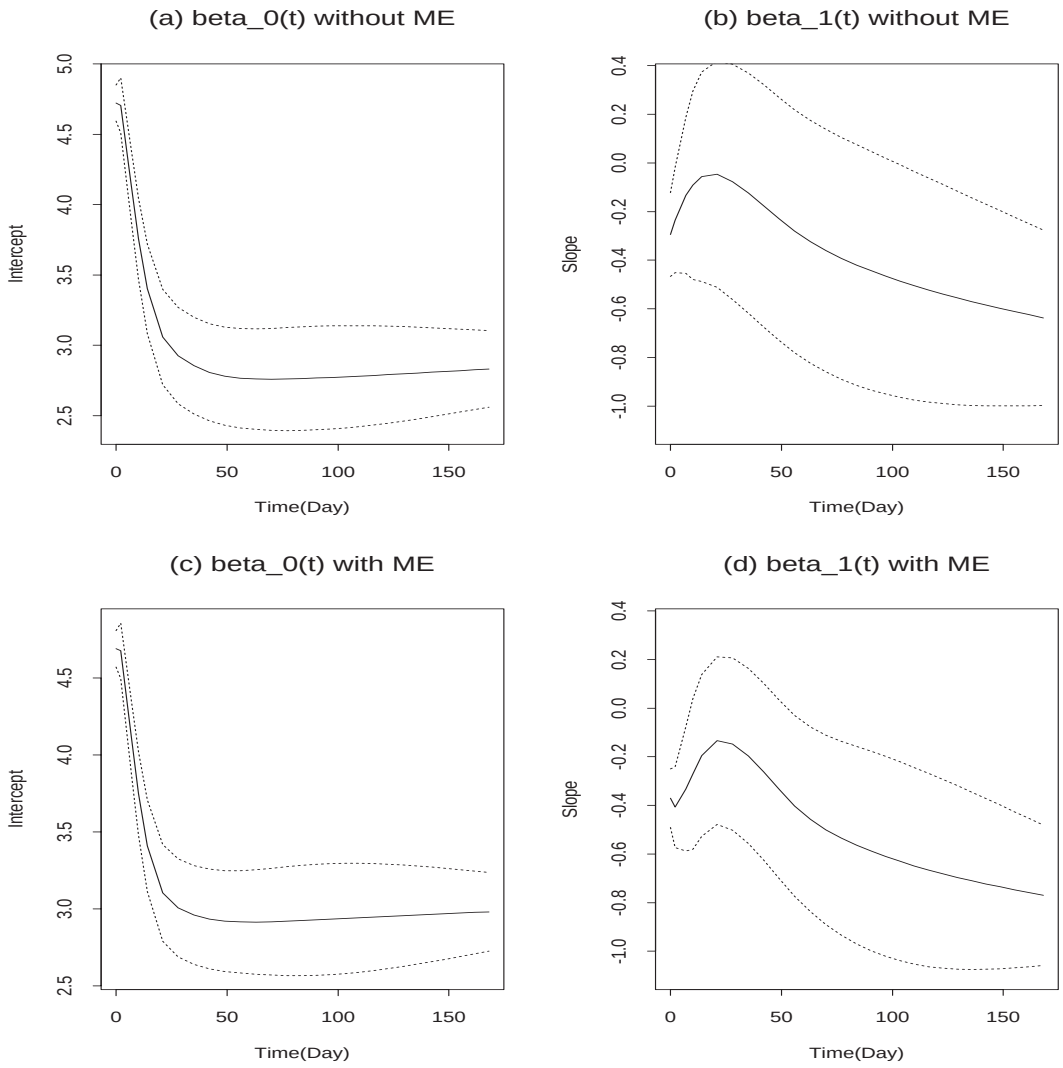

Fig. 3. The estimates of $\beta_0(t)$ and $\beta_1(t)$ from the cubic B-spline (CBS) method (solid curves). The estimation results with (lower panels) and without (upper panels) considering measurement errors in the covariate are presented. The dotted lines indicate ± 2 standard errors.

Thus, our results can be confirmed if $\beta_1^*(t)$ is similar to $1/\beta_{1i}(t)$ in the original model (or at least in a similar pattern). We plot the estimates of $\beta_1^*(t)$ and $1/\beta_{1i}(t)$ in Figure 4. We can see that the two curves are quite similar. This further confirms our results.

4.2 *Results for individual estimates*

One of the advantages of the proposed mixed-effects time-varying coefficient models is that parameter estimates for both population and individuals can be obtained. However, the parameter estimates for individuals may not exactly follow the patterns of the population if the between-subject variation is large.

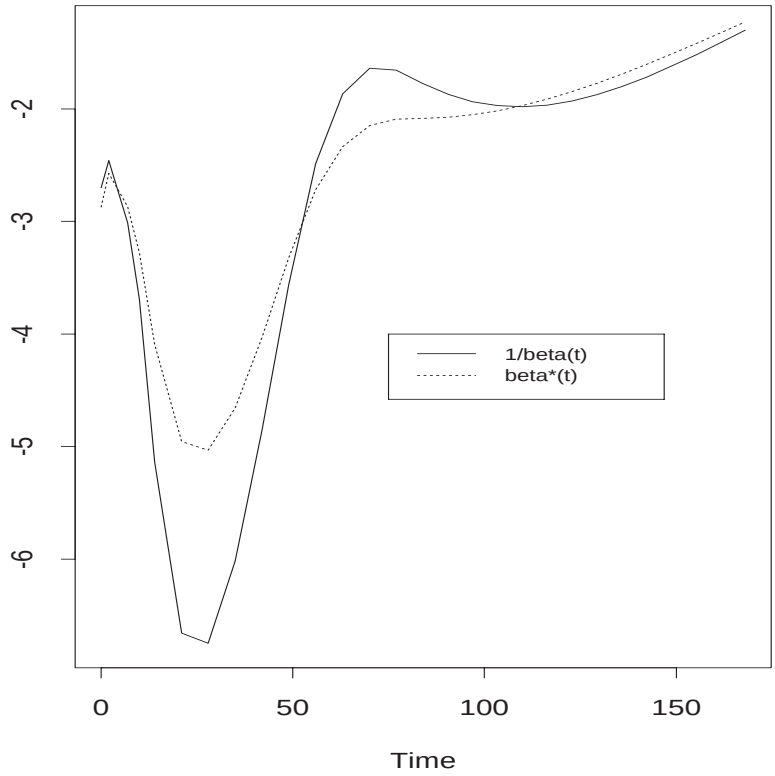

Fig. 4. The estimates of $\beta_1^*(t)$ (dotted curve) and $1/\beta_{1i}(t)$ (solid curve), which show similar patterns.

Figure 5 shows the individual estimates of $\beta_1(t)$ from four patients. For comparison, the corresponding population estimate of $\beta_1(t)$ is also plotted. From Figure 5, we can see that not only the magnitude but also the patterns are different between the population and individual estimates of $\beta_1(t)$. The individual estimates for subjects 1 and 44 show similar patterns to the population estimate. However, subject 1 has a positive correlation between viral load and CD4+ cell counts, while subject 44, similar to the population, has a strong negative correlation. Interestingly we also observed the discordance in patterns between the population estimate and individual estimates of $\beta_1(t)$. Two examples, subjects 10 and 24 are shown in Figure 5. In the early stage, their estimates of $\beta_1(t)$ have a similar pattern to the population. However, the estimated curves of $\beta_1(t)$ rebound rapidly later. Due to the large between-subject variation, the individual estimates of parameters become very important for individualizing treatment management and care for AIDS patients.

5. CONCLUSIONS AND DISCUSSION

In order to study the relationship between virologic and immunologic responses, repeatedly measured by HIV RNA levels (viral load) and CD4+ cell counts respectively in AIDS clinical trials, we proposed a mixed-effects varying-coefficient model. This model captured both population and individual relationships between the two longitudinal variables. We found that, from a clinical study of ACTG 315 (Lederman *et al.*, 1998), the viral load and CD4+ cell counts are inversely related in the study population

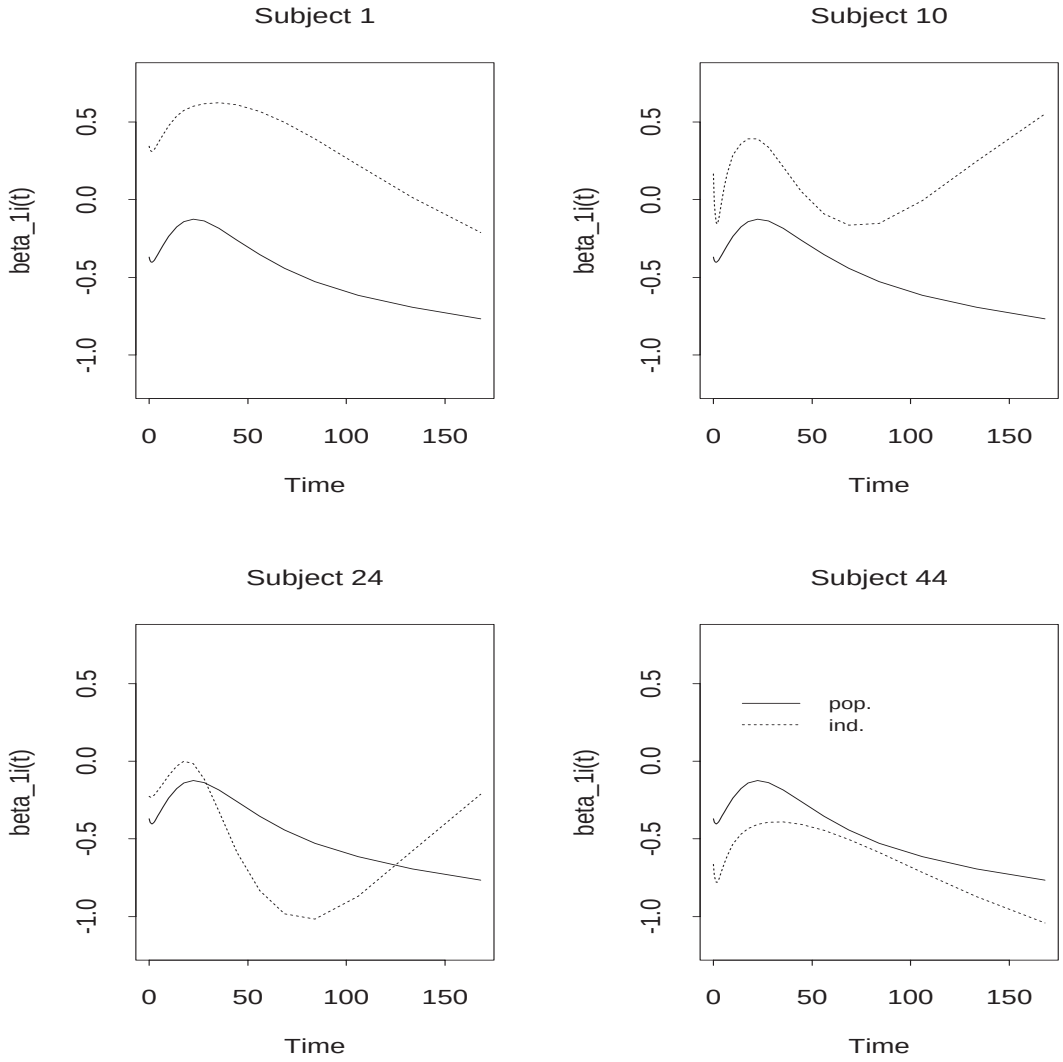

Fig. 5. The estimates of $\beta_1(t)$ from four selected subjects (dotted curves) and the corresponding population estimates (solid curves). The CBS method is used, allowing for measurement errors in the covariate.

during the treatment. However, the strength of the association varies smoothly over time. It is very strong at the beginning of the treatment, becomes weakest at about four weeks of treatment, and then gradually recovers to become strongest at week 24. When we change the role of viral load and CD4+ cell counts in the model, that is, treat CD4+ cell counts as the response variable and viral load as the covariate, we obtain a similar association pattern which further confirms our results.

Although the negative correlation between viral load and CD4+ cell counts is expected since CD4+ cells are major targets of HIV infection, the strong–weak–strong pattern of this correlation during potent antiviral treatments is interesting. One of the biological theories for the negative correlation is that HIV infection may trap CD4+ cells in lymphoid tissue from releasing to blood (Pakker *et al.*, 1998). Therefore, patients with high viral load (more viruses) may trap more CD4+ cells in lymphoid tissue which results

in a low CD4+ cell counts in blood, or a negative correlation between viral load and CD4+ cell counts in peripheral blood. After initiation of antiviral treatment, however, viral load is suppressed to a lower level or even below the limit of quantification, so that the trapped CD4+ cells may redistribute from lymphoid tissue to blood (Pakker *et al.*, 1998). Since the rapid redistribution of CD4+ cells from tissue to blood may end at week 4 or week 8 of treatment, and at this time viral load may reach a very low level (Figure 1) or below the detection limit, this may weaken the inverse association between the viral load and CD4+ cell counts. However, after week 4, some patients' viral load may rebound due to treatment failure (Figure 2), and the rebounded HIV viruses may trap CD4+ cells in lymphoid tissue again and make the inverse association come back. On the other hand, some patients' viral loads continue to decline after week 4 while their CD4+ cell counts continue to increase. This may also produce a negative association between the viral load and CD4+ cell counts in a later stage. From our findings, the association between viral load and CD4+ cell counts is not constant during antiviral treatment, and their relation may be disconnected at some time (at about week 4 for the study we considered). Thus, it is important to monitor both virologic and immunologic markers, especially for evaluating short-term (such as week 4) responses of an antiviral treatment, although one of the markers may be defined as a primary endpoint for long-term antiviral responses.

We believe that the proposed statistical models and the model-fitting approaches are methodologically valuable. The proposed mixed-effects varying-coefficient model allows us to establish a complicated nonlinear relationship between the two longitudinal variables, not only for the whole population of a study, but also for individual subjects. The proposed regression spline method is simple to implement, and naturally incorporates the typical features of longitudinal data such as between-subject and within-subject variations in the parameter estimates.

Besides the regression spline method considered above, other important nonparametric regression approaches such as smoothing splines and local polynomial kernel regression have also been suggested for varying-coefficient models (Hastie and Tibshirani, 1993; Hoover *et al.*, 1998; Wu *et al.*, 1998; Wu and Zhang, 2002). These methods may also apply to our models, and they are worthy subjects for future research.

ACKNOWLEDGEMENTS

HW and HL's research was supported by the National Institute of Allergy and Infectious Diseases/NIH grants R29 AI43220, R01 AI45356, R01 AI52765, and U01 AI38855. RJC's research was supported by a National Cancer Institute grant CA-57030 and a National Institute of Environmental Health Sciences grant P30-ESO9106. The authors thank Dr Jinting Zhang for insightful discussion and helpful comments, and the team of AIDS Clinical Trials Group Protocol 315 for permission to use the clinical data from their study. The authors also thank the AE and a referee for their valuable comments and suggestions.

REFERENCES

BUONACCORSI, J. P., DEMIDENKO, E. AND TOSTESON, T. D. (2000). Estimation in longitudinal random effects models with measurement error. *Statistica Sinica* **10**, 885–904.

CARROLL, R. J., RUPPERT, D. AND STEFANSKI, L. A. (1995). *Nonlinear Measurement Error Models*. New York: Chapman and Hall.

CLEVELAND, W. S., GROSSE, E. AND SHYU, W. M. (1991). Local Regression Models. Chambers, J. M. and Hastie, T. (eds), *Statistical Models in S*. Pacific Grove: Wadsworth and Brooks/Cole.

DAVIDIAN, M. AND GILTINAN, D. (1995). *Nonlinear Models for Repeated Measurement Data*. New York: Chapman and Hall.

EUBANK, R. L. (1999). *Nonparametric Regression and Spline Smoothing*. New York: Dekker.

FAN, J. AND ZHANG, J. T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society, Series B* **62**, 303–322.

FULLER, W. A. (1987). *Measurement Error Models*. New York: Wiley.

GREEN, P. J. AND SILVERMAN, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. London: Chapman and Hall.

HASTIE, T. AND TIBSHIRANI, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society, Series B* **55**, 757–796.

HIGGINS, K. M., DAVIDIAN, M. AND GILTINAN, D. M. (1997). A two-step approach to measurement error in time-dependent covariates in nonlinear mixed-effects models, with application to IGF-I pharmacokinetics. *Journal of the American Statistical Association* **92**, 436–448.

HOOVER, D. R., RICE, J. A., WU, C. O. AND YANG, L. P. (1998). Nonparametric smoothing estimates of timing-varying coefficient models with longitudinal data. *Biometrika* **85**, 809–822.

LAIRD, N. M. AND WARE, J. H. (1982). Random effects models for longitudinal data. *Biometrics* **38**, 963–974.

LEDERMAN, M. M., CONNICK, E. AND LANDAY, A., *et al.* (1998). Immunologic responses associated with 12 weeks of combination antiretroviral therapy consisting of zidovudine, lamivudine and ritonavir: results of AIDS clinical trials group protocol 315. *The Journal of Infectious Diseases* **178**, 70–79.

LIN, X. AND CARROLL, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association* **95**, 520–534.

MALLOLAS, J., LI, W., DEL RIO, A., PUMAROLA, T., GATELL, J. M. AND ERICE, A. (2000). Clinical Outcome, CD4+ Cell Count, and HIV-1 Reverse Transcriptase and Protease Sequences in Patients Remaining Viremic during HAART, *7th Conference on Retroviruses and Opportunistic Infections*, Abstract#334, Jan 30–Feb 2, 2000. San Francisco, CA.

MELLORS, J. W., KINGSLEY, L. A. AND RINALDO, C. R., *et al.* (1995). Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion. *Annals of Internal Medicine* **122**, 573–579.

MELLORS, J. W., RINALDO, C. R. AND GUPTA, P., *et al.* (1996). Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science* **272**, 1167–1170.

PADIERNA-OLIVOS, L., MORENO-ALTAMIRANO, M. M. B., SANCHEZ-COLON, S., MASSO-ROJAS, F. AND SANCHEZ-GARCIA, F. J. (2000). Connectivity and HIV-1 infection: role of CD4(+) T-cell counts and HIV-1 RNA copy number. *Scandinavian Journal of Immunology* **52**, 618–627.

PAKKER, N. G., NOTERMANS, D. W. AND DE BOER, R. J., *et al.* (1998). Biphasic kinetics of peripheral blood T cells after triple combination therapy in HIV-1 infection: a composite of redistribution and proliferation. *Nature Medicine* **4**, 208–214.

PINHEIRO, J. C. AND BATES, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer: New York.

RAMSAY, J. (1988). Monotone regression splines in action (with discussion). *Statistical Science* **4**, 425–461.

RICE, J. A. AND WU, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57**, 253–259.

SAAG, M. S., HOLODNIY, M., KURITZKES, D. R., O'BRIEN, W. A., COOMBS, R., POSCHER, M. E., JACOBSEN, D. M., SHAW, G. M., RICHMAN, D. D. AND VOLBERDING, P.A. (1996). HIV viral load markers in clinical practice. *Nature Medicine* **2**, 625–629.

SABIN, C., STASZEWSKI, S., PHILLIPS, A., RABENAU, H., LEPRI, A. C., WEIDMAN, E. AND MILLER, V. (2000). Discordant Immunological and Virological Responses to HAART. *7th Conference on Retroviruses and Opportunistic Infections*, Abstract#333, Jan 30–Feb 2, 2000. San Francisco, CA.

SHI, M., WEISS, R. E. AND TAYLOR, J. M. G. (1996). An analysis of pediatric CD4+ counts for acquired immune deficiency syndrome using flexible random curves. *Applied Statistics* **45**, 151–163.

TOSTESON, T. D., BUONACCORSI, J. P. AND DEMIDENKO, E. (1997). Covariate measurement error and the estimation of random effect parameters in a mixed model for longitudinal data. *Statistics in Medicine* **17**, 1959–1971.

VONESH, E. F. AND CHINCHILLI, V. M. (1996). *Linear and nonlinear models for the analysis of repeated measurements*. New York: Dekker.

WANG, N. S., LIN, X. H., GUTIERREZ, R. G. AND CARROLL, R. J. (1998). Bias analysis and SIMEX approach in generalized linear mixed measurement error models. *Journal of the American Statistical Association* **93**, 249–261.

WELSH, A., LIN, X. AND CARROLL, R. J. (2002). Marginal longitudinal nonparametric regression: locality and efficiency of spline and kernel methods. *Journal of the American Statistical Association* **97**, 482–493.

WEST, M., HARRISON, P. J. AND MIGON, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting (with discussion). *Journal of the American Statistical Association* **80**, 73–97.

WU, C. O., CHIANG, C. T. AND HOOVER, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *Journal of the American Statistical Association* **93**, 1388–1402.

WU, H., CONNICK, E., KURITZKES, D. R., LANDAY, A., SPRITZLER, J., ZHANG, B., KESSLER, H., LEDERMAN, M. M. AND THE ACTG 315 TEAM (2000). Cell Kinetic Patterns and Their Relationships with Virologic Responses in HIV-1-Infected Patients Treated with HAART. *7th Conference on Retroviruses and Opportunistic Infections*, Abstract#340, Jan 30–Feb 2, 2000. San Francisco, CA.

WU, H. AND DING, A. (1999). Population HIV-1 dynamics in vivo: applicable models and inferential tools for virological data from AIDS clinical trials. *Biometrics* **55**, 410–418.

WU, H., KURITZKES, D. R. AND MCCLERNON, D. R., *et al.* (1999). 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* **179**, 799–807.

WU, H. AND ZHANG, J. (2002). Local polynomial mixed-effects models for longitudinal data. *Journal of the American Statistical Association*, **97**, 883–897.

ZEGER, S. L. AND DIGGLE, P. J. (1994). Semiparametric models for longitudinal data with application to CD4+ cell numbers in HIV seroconverters. *Biometrics* **50**, 689–699.

ZHANG, D., LIN, X., RAZ, J. AND SOWERS, M. (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association* **93**, 710–719.

[*Received 21 August, 2001; first revision 17 April, 2002; second revision 15 July, 2002; accepted for publication 15 August, 2002*]