Published in final edited form as:

Biometrics. 2017 June ; 73(2): 441–451. doi:10.1111/biom.12626.

Generalized Semiparametric Varying-Coefficient Model for Longitudinal Data with Applications to Adaptive Treatment Randomizations

Li Qi¹, Yanqing Sun², and Peter B. Gilbert^{3,4}

¹Biostatistics and Programming, Sanofi, Bridgewater, NJ 08807, U.S.A

²Department of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223, U.S.A

³Department of Biostatistics, University of Washington, Seattle, WA 98195, U.S.A

⁴Fred Hutchinson Cancer Research Center, Seattle, WA 98109, U.S.A

SUMMARY

This paper investigates a generalized semiparametric varying-coefficient model for longitudinal data that can flexibly model three types of covariate effects: time-constant effects, time-varying effects, and covariate-varying effects. Different link functions can be selected to provide a rich family of models for longitudinal data. The model assumes that the time-varying effects are unspecified functions of time and the covariate-varying effects are parametric functions of an exposure variable specified up to a finite number of unknown parameters. The estimation procedure is developed using local linear smoothing and profile weighted least squares estimation techniques. Hypothesis testing procedures are developed to test the parametric functions of the covariate-varying effects. The asymptotic distributions of the proposed estimators are established. A working formula for bandwidth selection is discussed and examined through simulations. Our simulation study shows that the proposed methods have satisfactory finite sample performance. The proposed methods are applied to the ACTG 244 clinical trial of HIV infected patients being treated with Zidovudine to examine the effects of antiretroviral treatment switching before and after HIV develops the T215Y/F drug resistance mutation. Our analysis shows benefits of treatment switching to the combination therapies as compared to continuing with ZDV monotherapy before and after developing the 215-mutation.

Keywords

ACTG 244 AIDS clinical trial; Covariate-varying effects; Link function; Longitudinal data analysis; Hypothesis tests; Adaptive treatment randomization; Profile weighted least squares

Correspondence to: Yanqing Sun.

Supplementary Materials

Web Appendices for the proofs of the theorems referenced in Section 4, additional simulations and data analysis referenced in Section 7 and Section 8, along with the data and computer code, are available with this paper at the *Biometrics* website on the Wiley Online Library.

1. Introduction

Longitudinal data are common in medical and public health research. In AIDS clinical trials of HIV infected patients, for example, viral loads and CD4 counts are measured repeatedly during the course of studies. These biomarkers have long been known to be prognostic for both secondary HIV transmission and progression to clinical disease, cf. Mellors et al. (1997). Semiparametric regression models for longitudinal data have been intensively studied, in which the covariate effects are constant over time for some covariates and time-varying for others; see recent works by Fan and Li (2004), Qu and Li (2006), Fan et al. (2007) and Sun et al. (2013) among others. Semiparametric time-varying coefficient models allow effective simultaneous modeling of both types of covariate effects. However, in many applications there may be a third type of covariate effect – the effect that varies with an exposure variable. The statistical methods developed here apply for examining the possible exposure-varying effects of the adaptive treatment randomization strategy on longitudinal biomarkers such as CD4 cell counts and HIV viral load, with the exposure modifying variable the time since randomization.

A motivating example is a historical case study of antiretroviral treatment regimens, the ACTG 244 clinical trial. Zidovudine (ZDV) was the first drug approved for treatment of HIV infection. Initial approval was based on evidence of a short-term survival advantage over placebo when zidovudine was given to patients with advanced HIV disease. Shortly after that, zidovudine resistance was associated with disease progression measured by a rise in plasma virus and decline in CD4 cell counts in both children and adults receiving zidovudine monotherapy, cf. Japour et al. (1995). Subsequent studies suggested benefits of switching patients to treatments that combined ZDV with didanosine (ddI) or with ddI plus nevirapine (NVP). ACTG 244 enrolled subjects receiving ZDV monotherapy and monitored their HIV in plasma bi-monthly for the T215Y/F mutation. When a subject's viral population developed the 215 mutation, the subject was randomized to continue ZDV, add ddI or add ddI plus NVP. An important question is whether the treatment switching has any beneficial effects in treating the HIV infected patients.

We investigate the effect of the adaptive treatment randomization under the generalized semiparametric model for longitudinal data with a general link function. Our model considers three types of covariate effects: time-varying effects, covariate-varying effects and constant effects. The covariate-varying effects are specified as parametric functions of effect modifiers while the time-varying effects are modeled as nonparametric functions. While the nonparametric approach for the covariate-varying effects is more flexible, parametric modeling also has its advantage, in particular, when the dimension of the covariate that it depends on is moderately high. The parametric approach avoids the curse of dimensionality. Parametric forms are more interpretable when models are built based on the knowledge of the underline biological processes. Although in some cases one can include linear interaction terms to account for covariate-varying effects, this is limited to situations where the covariate-varying effects are linear functions of the exposure variables. It is also difficult to make inference for individual model parameters separately when higher order interactions are present. We develop hypothesis testing procedures to test the parametric functions of the covariate-varying effects. To the best of our knowledge, such procedures do not exist for

longitudinal analysis. The generalized semiparametric model allows selection of different link functions. Thus our methods can be applied for continuous as well as categorical longitudinal responses.

The rest of the paper is organized as follows. In Section 2, we introduce the generalized semiparametric varying-coefficient regression model. The profile local linear estimation method for the proposed model is developed in Section 3. In Section 4, we establish the asymptotic results for the nonparametric and parametric estimators. The hypothesis testing procedures are developed in Section 5. The finite sample performances of the proposed estimators and test statistics are examined in simulations in Section 6. The methods are applied to the ACTG 244 data in Section 7. Concluding remarks are given in Section 8.

2. Generalized semiparametric varying-coefficient model

Suppose there is a random sample of n subjects and τ is the end of follow-up. The longitudinal responses $Y_i(t)$ for subject i are observed at the sampling time points $0 = T_{i1} < T_{i2} < \dots < T_{in_i} \leq \tau$, where n_i is the total number of observations from subject i . The sampling times can be irregular and dependent on covariates. In addition, some subjects may drop out

of the study early. Let $N_i(t) = \sum_{j=1}^{n_i} I(T_{ij} \leq t)$ be the number of observations taken from the i th subject by time t , where $I(\cdot)$ is the indicator function. Let C_i be the end of follow-up time or censoring time, whichever comes first. The responses for subject i can only be observed at time points before C_i . Thus $N_i(t)$ can be written as $N_i^*(t \wedge C_i)$, where $N_i^*(t)$ is the counting process of potential sampling times. Let $\mathbf{X}_i(t)$ and $\mathbf{U}_i(t)$ be p and r dimensional vectors of possibly time-dependent covariates, respectively. Assume that $\{Y_i(\cdot), \mathbf{X}_i(\cdot), \mathbf{U}_i(\cdot), N_i(\cdot), i = 1, \dots, n\}$ are independent identically distributed (iid) random processes. The censoring time C_i is noninformative in the sense that

$E\{dN_i^*(t) | \mathbf{X}_i(t), \mathbf{U}_i(t), C_i \geq t\} = E\{dN_i^*(t) | \mathbf{X}_i(t), \mathbf{U}_i(t)\}$ and $E\{Y_i(t) | \mathbf{X}_i(t), \mathbf{U}_i(t), C_i \geq t\} = E\{Y_i(t) | \mathbf{X}_i(t), \mathbf{U}_i(t)\}$. Assume that $dN_i^*(t)$ is independent of $Y_i(t)$ conditional on $\mathbf{X}_i(t)$, $\mathbf{U}_i(t)$ and $C_i \geq t$. The censoring time C_i is allowed to depend on $\mathbf{X}_i(\cdot)$ and $\mathbf{U}_i(\cdot)$.

Suppose that covariates $\mathbf{X}_i(t) = (\mathbf{X}_{1i}^T(t), \mathbf{X}_{2i}^T(t), \mathbf{X}_{3i}^T(t))^T$ consist of three parts, $\mathbf{X}_{1i}(t)$, $\mathbf{X}_{2i}(t)$, $\mathbf{X}_{3i}(t)$, of dimensions p_1 , p_2 and p_3 , respectively, each of which has a different role in the model. We study the following generalized semiparametric varying-coefficient model:

$$\mu_i(t) = E\{Y_i(t) | \mathbf{X}_i(t), \mathbf{U}_i(t)\} = g^{-1} \left\{ \boldsymbol{\alpha}^T(t) \mathbf{X}_{1i}(t) + \boldsymbol{\beta}^T \mathbf{X}_{2i}(t) + \boldsymbol{\gamma}^T(\mathbf{U}_i(t); \boldsymbol{\theta}) \mathbf{X}_{3i}(t) \right\} \quad (1)$$

for $0 \leq t \leq \tau$, where $g(\cdot)$ is a known link function, $\boldsymbol{\alpha}(\cdot)$ is a p_1 -dimensional vector of completely unspecified functions representing the time-varying effects of $\mathbf{X}_{1i}(t)$, $\boldsymbol{\beta}$ is a p_2 -dimensional vector of parameters, $\boldsymbol{\theta}$ is a q -dimensional vector of parameters, and $\boldsymbol{\gamma}(\mathbf{u}, \boldsymbol{\theta})$ is a p_3 -dimensional vector of possibly nonlinear parametric functions defined on the range U of $\mathbf{U}_i(\cdot)$. Setting the first component of $\mathbf{X}_{1i}(t)$ as 1 gives a nonparametric baseline function. $\boldsymbol{\gamma}(\mathbf{u}) = \boldsymbol{\gamma}(\mathbf{u}, \boldsymbol{\theta})$ is the effect of $\mathbf{X}_{3i}(t)$ at covariate level $\mathbf{U}_i(t) = \mathbf{u}$. Both categorical and continuous longitudinal responses can be modelled with appropriately chosen link functions.

For example, the identity and logarithm link functions can be used for continuous response variables while the logit link function can be used for binary responses.

For the motivating example of the ACTG 244 study, it is of interest to know how biomarkers such as viral load and CD4 cell count respond to the new treatments. The effects of the new treatments are likely to depend on the time duration, $U_{\lambda}(t) = t - S_i$ since the switching, where t is the time since initiation of antiretroviral therapy (ART) and S_i is the time of treatment randomization. Letting $X_{3\lambda}(t) = I(t > S_i)$ in model (1), $\gamma(u)$ represents the change in the conditional mean response at time u after treatment randomization adjusting for other covariates $X_{1i}(t)$ and $X_{2i}(t)$. On the other hand, if we let $X_{3i}(t) = X_{3i}^o(t)I(t > S_i)$ where $X_{3i}^o(t)$ are the indicators for the new treatments after randomization, then $\gamma(u)$ are the effects of new treatments starting from treatment switching.

3. Statistical estimation

3.1 Estimation procedure

This section develops the estimation procedure for model (1). The approach utilizes the local linear estimation technique which has been shown to be design-adaptive and more efficient in correcting boundary bias than the kernel smoothing approach (Fan and Gijbels, 1996).

At each t_0 , let $\alpha(t) = \alpha(t_0) + \dot{\alpha}(t_0)(t - t_0) + O((t - t_0)^2)$ be the first order Taylor expansion of $\alpha(t)$ for $t \in N_{t_0}$ a neighborhood of t_0 . Where $\dot{\alpha}(t_0)$ is the derivative of $\alpha(t)$ at $t = t_0$. Denote $\alpha^*(t_0) = (\alpha^T(t_0), \dot{\alpha}^T(t_0))^T$, $X_{1i}^*(t, t - t_0) = X_{1i}(t) \otimes (1, t - t_0)^T$, where $\otimes$ is the Kronecker product. Let $\zeta = (\beta^T, \theta^T)^T$. For $t \in N_{t_0}$, model (1) can be approximated by

$$\tilde{\mu}(t, t_0, \alpha^*(t_0), \zeta | X_i, U_i) = g^{-1} \left\{ \alpha^{*T}(t_0) X_{1i}^*(t, t - t_0) + \beta^T X_{2i}(t) + \gamma^T(U_i(t), \theta) X_{3i}(t) \right\}. \quad (2)$$

The local linear estimating function for $\alpha^*(t_0)$ at each t_0 and for fixed ζ is given by

$$U_{\alpha}(\alpha^*; \zeta, t_0) = \sum_{i=1}^n \int_0^T W_i(t) [Y_i(t) - \tilde{\mu}(t, t_0, \alpha^*(t_0), \zeta | X_i, U_i)] X_{1i}^*(t, t - t_0) K_h(t - t_0) dN_i(t), \quad (3)$$

where $W_{\lambda}(t) = W(t, X_{\lambda}(t), U_{\lambda}(t))$ is a nonnegative weight process, $K(\cdot)$ is a kernel function, $h = h_n > 0$ is a bandwidth parameter and $K_h(\cdot) = K(\cdot/h)/h$. The solution to the equation $U_{\alpha}(\alpha^*; \zeta, t_0) = 0$ is denoted by $\tilde{\alpha}^*(t_0, \zeta)$.

Let $\tilde{\alpha}(t, \zeta)$ be the first p_1 components of $\tilde{\alpha}^*(t, \zeta)$. The profile weighted least squares estimator $\hat{\zeta}$ is obtained by solving the following estimating function

$$U_{\zeta}(\zeta) = \sum_{i=1}^n \int_{t_1}^{t_2} W_i(t) \left[Y_i(t) - g^{-1} \left\{ \tilde{\alpha}^T(t, \zeta) X_{1i}(t) + \eta^T(U_i(t), \zeta) X_{2i}^*(t) \right\} \right] \left\{ \left(\frac{\partial \tilde{\alpha}(t, \zeta)}{\partial \zeta} \right)^T X_{1i}(t) + \left(\frac{\partial \eta(U_i(t), \zeta)}{\partial \zeta} \right)^T X_{2i}^*(t) \right\} dN_i(t), \quad (4)$$

where $\eta(U_i(t), \zeta) = (\beta^T, \gamma^T(U_i(t), \theta))^T$, $\eta(U_i(t), \zeta)/\zeta = \text{diag} \{I_{p_2}, \gamma(U_i(t), \theta)/\theta\}$, $X_{2i}^*(t) = (X_{2i}^T(t), X_{3i}^T(t))^T$, and $\tilde{\alpha}(t, \zeta)/\zeta$ is the first p_1 rows of $\alpha^*(t, \zeta)/\zeta$. We take $[t_1, t_2] \subset (0, \tau)$ in order to avoid possible instability near the boundary. The profile estimator of $\alpha(t_0)$ is obtained by $\hat{\alpha}(t_0) = \tilde{\alpha}(t_0, \hat{\zeta})$ through substitution.

The partial derivatives $\alpha^*(t, \zeta)/\zeta$ needed to evaluate (4) can be expressed in terms of the partial derivatives of $U_{\alpha}(\alpha^*, \zeta, t)$ at $\alpha^* = \tilde{\alpha}^*(t, \zeta)$. Specifically, since $U_{\alpha}(\tilde{\alpha}^*(t, \zeta); \zeta, t) \equiv 0_{2p_1}$, it follows that $\tilde{\alpha}^*(t, \zeta)$ satisfies

$$\left\{ \frac{\partial U_{\alpha}(\alpha^*; \zeta, t)}{\partial \alpha^*} \frac{\partial \tilde{\alpha}^*(t, \zeta)}{\partial \zeta} + \frac{\partial U_{\alpha}(\alpha^*; \zeta, t)}{\partial \zeta} \right\} \Big|_{\alpha^* = \tilde{\alpha}^*(t, \zeta)} = 0_{2p_1}. \text{ Let } \phi(\cdot) = g^{-1}(\cdot) \text{ be the inverse function of the link function } g(\cdot) \text{ and } \dot{\phi}(\cdot) \text{ be the derivative of } \phi(\cdot). \text{ Then,}$$

$$\frac{\partial \tilde{\alpha}^*(t, \zeta)}{\partial \zeta} = - \left\{ \frac{\partial U_{\alpha}(\alpha^*; \zeta, t)}{\partial \alpha^*} \right\}^{-1} \frac{\partial U_{\alpha}(\alpha^*; \zeta, t)}{\partial \zeta} \Big|_{\alpha^* = \tilde{\alpha}^*(t, \zeta)}, \quad (5)$$

where $\partial U_{\alpha}(\alpha^*; \zeta, t_0)/\partial \alpha^* = - \sum_{i=1}^n \int_0^T W_i(t) \dot{\mu}_i^*(t_0, t, \zeta) X_{1i}^*(t, t-t_0)^{\otimes 2} K_h(t-t_0) dN_i(t)$, $\partial U_{\alpha}(\alpha^*; \zeta, t_0)/\partial \zeta = - \sum_{i=1}^n \int_0^T W_i(t) \dot{\mu}_i^*(t_0, t, \zeta) X_{1i}^*(t, t-t_0) \{X_{2i}^*(t)\}^T (\partial \eta(U_i(t), \zeta)/\partial \zeta) K_h(t-t_0) dN_i(t)$. Here, $\dot{\mu}_i^*(t_0, t, \zeta) = \dot{\phi}\{\alpha^{*T}(t_0) X_{1i}^*(t, t-t_0) + \eta^T(U_i(t), \zeta) X_{2i}^*(t)\}$.

When the link function is the identity function, $\tilde{\alpha}^*(t_0, \zeta)$ can be solved explicitly as the root of the estimating function (3). Under a general link function, $\tilde{\alpha}^*(t_0, \zeta)$ can be solved using an iterative algorithm. The estimation procedure iteratively updates estimates of the nonparametric component $\tilde{\alpha}^*(t_0, \zeta)$ and the parametric component ζ . Specifically, the estimators $\hat{\alpha}(t_0)$ and $\hat{\zeta}$ are obtained through the following iterated algorithm:

Computational algorithm

1. Let $\hat{\alpha}(t)^{(0)}$ and $\hat{\zeta}^{(0)}$ be initial values.
2. For each jump point of $\{N_{\lambda}(\cdot), i = 1, \dots, n\}$, say t , the m th step estimator $\hat{\alpha}^{*\{m\}}(t) = \hat{\alpha}^*(t, \hat{\zeta}^{(m-1)})$ is the root of the estimating function (3) satisfying $U_{\alpha}(\hat{\alpha}^{*\{m\}}(t), \hat{\zeta}^{(m-1)}, t) = 0$, where $\hat{\zeta}^{(m-1)}$ is the estimate of ζ at the $(m-1)$ th step.
3. The m th step estimator $\hat{\zeta}^{(m)}$ is the root of $U_{\zeta}(\zeta) = 0$ defined in (4) obtained after replacing $\tilde{\alpha}(t, \zeta)$ with $\hat{\alpha}^{*\{m\}}(t)$, where $\hat{\alpha}^{*\{m\}}(t)$ is the first p_1 components of $\hat{\alpha}^{*\{m\}}(t)$.

4. Repeating steps 2 and 3, the estimators $\hat{\alpha}^{*\{m\}}(t)$ and $\hat{\zeta}^{m\}$ are updated at each iteration until convergence. The estimates $\hat{\zeta}$ and $\hat{\alpha}(t)$ are $\hat{\zeta}^{m\}$ and the first p_1 components of $\hat{\alpha}^{*\{m\}}(t)$, respectively, at convergence.

In our numerical study, the main program of the estimation procedure is implemented using Matlab while part of the program that solves the estimating equation (3) for $\alpha(t)$ at all jump points is implemented using C++ to save computing time. The estimators usually converge in 5 iterations, and take less than one minute for a sample of size 400.

3.2 Bandwidth selection

The optimal theoretical bandwidth is difficult to achieve since it would involve estimating the second derivative $\ddot{\alpha}(t)$; see Theorem 2 in the next section and also Fan and Gijbels (1996) and Sun et al. (2013). In practice, the appropriate bandwidth can be based on a cross-validation method. This approach is widely used in the nonparametric function estimation literature; see Rice and Silverman (1991) for a leave-one-subject-out cross-validation approach, and Tian et al. (2005) for a K -fold cross-validation approach for survival data. The K -fold cross-validation approach in the current setting divides the data into K approximately equal-sized groups. With D_k denoting the k th subgroup of data, the k th prediction error is given by

$$PE_k(h) = \sum_{i \in D_k} \left\{ \int_{t_1}^{t_2} W_i(t) \left[Y_i(t) - g^{-1} \left\{ \hat{\alpha}_{(-k)}^T(t) \mathbf{X}_{i1}(t) + \boldsymbol{\eta}^T(\mathbf{U}_i(t), \hat{\zeta}_{(-k)}) \mathbf{X}_{2i}^*(t) \right\} \right] dN_i(t) \right\}^2,$$

for $k = 1, \dots, K$, where $\hat{\alpha}_{(-k)}(t)$ and $\hat{\zeta}_{(-k)}$ are estimated using the data from all subgroups other than D_k . The K -fold cross-validation bandwidth selection is obtained by minimizing

$$\text{the total prediction error } PE(h) = \sum_{k=1}^K PE_k(h) \text{ with respect to } h.$$

We also investigated an alternative bandwidth selection method. The bandwidth selection formula $h = C\hat{\sigma}_T n^{-1/3}$ has been examined for nonparametric density estimation and for semiparametric failure time regression in Jones et al. (1991) and Zhou and Wang (2000) among others, where C is a constant and $\hat{\sigma}_T$ is the estimated standard error of the sampling times in the domain of the nonparametric functions to be estimated. To adopt the formula for longitudinal data, we note that the observation times $\{T_{ij}, j = 1, \dots, n_i\}$ for a subject i are likely dependent. Suppose that ϕ_i is the random effect that induces such dependence. Then the variance of the observation times can be expressed as

$$\sigma_T^2 = \text{Var}(T_{ij}) = E\{\text{Var}(T_{ij}|\phi_i)\} + \text{Var}\{E(T_{ij}|\phi_i)\}, \text{ which can be estimated by } \\ \hat{\sigma}_T^2 = n^{-1} \sum_{i=1}^n S_i^a + S^b, \text{ where } S_i^a \text{ is the within-subject sample variance of } \{T_{ij}, j = 1, \dots, n_i\}, \\ \text{is } S^b \text{ the between-subject sample variance of } \bar{T}_i = n_i^{-1} \sum_{j=1}^{n_i} T_{ij} \text{ for } i = 1, \dots, n.$$

Our numerical study for both the bandwidth selection methods shows that the bandwidth selection with $h = C\hat{\sigma}_T n^{-1/3}$ by using C in the range from 3 to 5 is close to the bandwidth selected using the K -fold cross-validation for K in the range of 3 to 10. Our study used the K -fold cross-validation bandwidth selection as the bench mark to calibrate the constant C . The simulation results presented in Section 6 using $C = 4$ suggest that the formula $h =$

$4\hat{\sigma}_T n^{-1/3}$ works well. A larger C can be used if the distribution of the sampling times is skewed or sparse in some areas.

4. Asymptotic properties

Let ζ_0 and $\alpha_0(t)$ be the true values of ζ and $\alpha(t)$ under model (1), respectively. Let

$$\mu_i(t) = \phi \left\{ \alpha_0^T(t) \mathbf{X}_{1i}(t) + \eta^T(\mathbf{U}_i(t), \zeta_0) \mathbf{X}_{2i}^*(t) \right\},$$

$\dot{\mu}_i(t) = \dot{\phi} \left\{ \alpha_0^T(t) \mathbf{X}_{1i}(t) + \eta^T(\mathbf{U}_i(t), \zeta_0) \mathbf{X}_{2i}^*(t) \right\}$ and $\varepsilon_i(t) = Y_i(t) - \mu_i(t)$. Let $w_i(t) = w(t, \mathbf{X}_i(t), \mathbf{U}_i(t))$, where $w(t, \mathbf{x}, \mathbf{u})$ is the deterministic limit of $W(t, \mathbf{x}, \mathbf{u})$ in probability as $n \rightarrow \infty$. Define $e_{11}(t) = E[w_i(t) \dot{\mu}_i(t) \mathbf{X}_{1i}(t) \otimes \lambda_i(t) \xi_i(t)]$ and

$$e_{12}(t) = E \left[w_i(t) \dot{\mu}_i(t) \mathbf{X}_{1i}(t) \{ \mathbf{X}_{2i}^*(t) \}^T (\partial \eta(\mathbf{U}_i(t), \zeta_0) / \partial \zeta) \lambda_i(t) \xi_i(t) \right], \text{ where } \xi_i(t) = I(C_i = t) \text{ and } \lambda_i(t) \text{ is the conditional mean rate of } N_i(t) \text{ defined by } \lambda_i(t) dt = E(dN_i(t) | \mathbf{X}_i(t), \mathbf{U}_i(t)).$$

Let $\mathbf{Q}_i(t) = - (e_{12}(t))^T (e_{11}(t))^{-1} \mathbf{X}_{1i}(t) + (\partial \eta(\mathbf{U}_i(t), \zeta_0) / \partial \zeta)^T \mathbf{X}_{2i}^*(t)$. Denote by $\dot{\alpha}_0(t)$, $\ddot{\alpha}_0(t)$ the first and second derivatives of the true $\alpha_0(t)$ with respect to t , respectively.

$$\text{Let } \hat{\mu}_i(t) = \phi \left\{ \hat{\alpha}^T(t) \mathbf{X}_{1i}(t) + \eta^T(\mathbf{U}_i(t), \hat{\zeta}) \mathbf{X}_{2i}^*(t) \right\},$$

$$\hat{\dot{\mu}}_i(t) = \dot{\phi} \left\{ \hat{\alpha}^T(t) \mathbf{X}_{1i}(t) + \eta^T(\mathbf{U}_i(t), \hat{\zeta}) \mathbf{X}_{2i}^*(t) \right\} \text{ and } \hat{\varepsilon}_i(t) = Y_i(t) - \hat{\mu}_i(t). \text{ Let}$$

$$\hat{E}_{11}(t) = n^{-1} \sum_{i=1}^n \int_0^T W_i(s) \hat{\mu}_i(s) \mathbf{X}_{1i}(s) \otimes K_h(s-t) dN_i(s) \text{ and}$$

$$\hat{E}_{12}(t) = n^{-1} \sum_{i=1}^n \int_0^T W_i(s) \hat{\mu}_i(s) \mathbf{X}_{1i}(s) \{ \mathbf{X}_{2i}^*(s) \}^T (\partial \eta(\mathbf{U}_i(s), \hat{\zeta}) / \partial \zeta) K_h(s-t) dN_i(s).$$

$$\text{Let } \hat{\mathbf{Q}}_i(t) = - (\hat{E}_{12}(t))^T (\hat{E}_{11}(t))^{-1} \mathbf{X}_{1i}(t) + (\partial \eta(\mathbf{U}_i(t), \hat{\zeta}) / \partial \zeta)^T \mathbf{X}_{2i}^*(t).$$

The following theorems characterize the asymptotic properties of the estimators $\hat{\zeta}$ and $\hat{\alpha}(t)$ under Condition A given in the Web-based Supplementary Material.

Theorem 1

Under Condition A, $\hat{\zeta} \xrightarrow{p} \zeta_0$, and $\sqrt{n}(\hat{\zeta} - \zeta_0)$ converges in distribution to a mean zero Gaussian random vector with covariance matrix $\mathbf{A}^{-1} \Sigma \mathbf{A}^{-1}$, where

$$\mathbf{A} = E \left[\int_{t_1}^{t_2} w_i(t) \dot{\mu}_i(t) \{ \mathbf{Q}_i(t) \}^{\otimes 2} dN_i(t) \right] \text{ and } \Sigma = E \left[\int_{t_1}^{t_2} w_i(t) \mathbf{Q}_i(t) \varepsilon_i(t) dN_i(t) \right]^{\otimes 2}.$$

The matrices $\mathbf{A}$ and Σ can be consistently estimated respectively by

$$\hat{\mathbf{A}} = n^{-1} \sum_{i=1}^n \int_{t_1}^{t_2} W_i(t) \hat{\mu}_i(t) \{ \hat{\mathbf{Q}}_i(t) \}^{\otimes 2} dN_i(t) \text{ and}$$

$$\hat{\Sigma} = n^{-1} \sum_{i=1}^n \left(\int_{t_1}^{t_2} W_i(s) \hat{\varepsilon}_i(s) \hat{\mathbf{Q}}_i(s) dN_i(s) \right)^{\otimes 2}.$$

Theorem 2

Under Condition A, $\hat{\alpha}(t) \xrightarrow{p} \alpha_0(t)$, uniformly in $t \in [t_1, t_2]$, and

$$(nh)^{1/2} (\hat{\alpha}(t) - \alpha_0(t) - \frac{1}{2} \mu_2 h^2 \ddot{\alpha}_0(t)) \xrightarrow{D} N(0, \sum_{\alpha}(t)), \quad (6)$$

where $\mu_2 = \int_{-1}^1 t^2 K(t) dt$, $\sum_{\alpha}(t) = (e_{11}(t))^{-1} \sum_e(t) (e_{11}(t))^{-1}$, and $\sum_e(t) = \lim_{n \rightarrow \infty} h E \left\{ \int_0^T w_i(s) \varepsilon_i(s) \mathbf{X}_{1i}(s) K_h(s-t) dN_i(s) \right\}^{\otimes 2}$.

The variance-covariance matrix $\Sigma_{\alpha}(t)$ can be estimated consistently by replacing $e_{11}(t)$ with $\hat{E}_{11}(t)$ and $\Sigma_e(t)$ with $\sum_e(t) = n^{-1} h \sum_{i=1}^n \{\hat{g}_i(t)\}^{\otimes 2}$, where

$$\hat{g}_i(t) = \int_0^T W_i(s) \hat{\varepsilon}_i(s) \mathbf{X}_{1i}(s) K_h(s-t) dN_i(s) - \hat{E}_{12}(t) \hat{A}^{-1} \int_{t_1}^{t_2} W_i(s) \hat{Q}_i(s) \hat{\varepsilon}_i(s) dN_i(s).$$

5. Testing the covariate-varying effects

The generalized semiparametric varying-coefficient model (1) assumes that the covariate-varying effects are parametric functions of an effect modifier $U_{\lambda}(t)$. The parametric functions $\gamma(U_{\lambda}(t), \theta)$ can be specified based on knowledge of the underlying biological processes in some cases, and in others, they can be chosen as polynomial functions or linear combinations of basis functions such as the B-spline basis. This section develops hypothesis testing procedures to test the parametric forms $\gamma(U_{\lambda}(t), \theta)$. We construct the test process based on the weighted residual process that is closely related to the score function (4).

To test $H_0: \gamma(u) = \gamma(u, \theta)$, we consider the test process

$$\mathbf{R}(u, \hat{\zeta}) = n^{-1/2} (\mathbf{I}_r \otimes \hat{\mathbf{A}}^{-1}) \sum_{i=1}^n \int_{t_1}^{t_2} W_i(t) I \{U_i(t) \leq u\} \otimes \hat{Q}_i(t) \hat{\varepsilon}_i(t) dN_i(t), \quad (7)$$

for $u \in R^r$, where $\mathbf{I}_r$ is the $r \times r$ identity matrix, $I \{U_{\lambda}(t) \leq u\}$ is the column vector of the indicator functions, and $\otimes$ is the Kronecker product of matrices. By stratifying the score function for the values of $U_{\lambda}(t)$, the process $\mathbf{R}(u, \hat{\zeta})$ is sensitive to the misspecifications of $\gamma(u, \theta)$. The factor $\mathbf{I}_r \otimes \hat{\mathbf{A}}^{-1}$ balances the contributions from the covariates that might have larger or smaller variations.

We consider the supremum test statistic $T_1 = \sup_{u \in R^r} \|\mathbf{R}(u, \hat{\zeta})\|$ and the sum of the absolute deviation test statistic $T_2 = \sum_{u \in R^r} \|\mathbf{R}(u, \hat{\zeta})\|$ where R^r is a set of grid points in R^r , and $\|\cdot\|$ is the Euclidean norm in $R^{(p_2+q)}$.

By the first order approximation, $\mathbf{R}(u, \hat{\zeta}) = \mathbf{R}(u, \zeta_0) + (\mathbf{R}(u, \zeta_0) / \zeta_0) (\hat{\zeta} - \zeta_0) + o_p(1)$. Let $\mathbf{A}_u = E \left\{ \int_{t_1}^{t_2} w_i(t) \dot{\mu}_i(t) (I \{U_i(t) \leq u\} \otimes \mathbf{Q}_i(t)) \mathbf{Q}_i^T(t) dN_i(t) \right\}$. Following the proof of Theorem 1, we have

$$\begin{aligned} n^{-1/2} \partial \mathbf{R}(u, \zeta_0) / \partial \zeta &\xrightarrow{P} \\ &= (\mathbf{I}_r \otimes \mathbf{A}^{-1}) \mathbf{A}_u, \mathbf{R}(u, \zeta_0) \\ &= n^{-1/2} (\mathbf{I}_r \otimes \mathbf{A}^{-1}) \sum_{i=1}^n \int_{t_1}^{t_2} w_i(t) I \{U_i(t) \leq u\} \otimes \mathbf{Q}_i(t) \varepsilon_i(t) dN_i(t) + o_p(1), \text{ and} \\ n^{-1/2} (\hat{\zeta} - \zeta_0) &= \mathbf{A}^{-1} n^{-1/2} \sum_{i=1}^n \int_{t_1}^{t_2} w_i(t) \varepsilon_i(t) \mathbf{Q}_i(t) dN_i(t) + o_p(1). \text{ Hence,} \\ \mathbf{R}(u, \hat{\zeta}) &= n^{-1/2} \sum_{i=1}^n \mathbf{D}_i(u) + o_p(1), \text{ where} \end{aligned}$$

$$D_i(\mathbf{u}) = (\mathbf{I}_r \otimes \mathbf{A}^{-1}) \int_{t_1}^{t_2} w_i(t) [I\{U_i(t) \leq \mathbf{u}\} \otimes \mathbf{Q}_i(t) - \mathbf{A}_u \mathbf{A}^{-1} \mathbf{Q}_i(t)] \varepsilon_i(t) dN_i(t). \quad (8)$$

By Lemma 1 in the Appendix of Sun et al. (2016), $\mathbf{R}(\mathbf{u}, \hat{\mathbf{G}})$ converges weakly to a mean-zero Gaussian process $\mathbf{G}(\mathbf{u})$, for $\mathbf{u} \in \mathcal{R}^r$. It follows from the continuous mapping theorem that

$$T_1 \xrightarrow{\mathcal{D}} \sup_{\mathbf{u} \in \Delta} \|\mathbf{G}(\mathbf{u})\| \text{ and } T_2 \xrightarrow{\mathcal{D}} \sum_{\mathbf{u} \in \Delta} \|\mathbf{G}(\mathbf{u})\|.$$

The critical values of the test statistics T_1 and T_2 , or their asymptotic distributions, can be approximated by using the Gaussian multipliers resampling method, cf., Lin et al. (1993) and Sun et al. (2013). Let

$$\hat{\mathbf{A}}_u = n^{-1} \sum_{i=1}^n \int_{t_1}^{t_2} W_i(t) \hat{\mu}(t) (I\{U_i(t) \leq \mathbf{u}\} \otimes \hat{\mathbf{Q}}_i(t)) \hat{\mathbf{Q}}_i^T(t) dN_i(t), \text{ and}$$

$$\hat{D}_i(\mathbf{u}) = (\mathbf{I}_r \otimes \hat{\mathbf{A}}^{-1}) \int_{t_1}^{t_2} W_i(t) [I\{U_i(t) \leq \mathbf{u}\} \otimes \hat{\mathbf{Q}}_i(t) - \hat{\mathbf{A}}_u \hat{\mathbf{A}}^{-1} \hat{\mathbf{Q}}_i(t)] \hat{\varepsilon}_i(t) dN_i(t). \quad (9)$$

Define $\mathbf{G}^*(\mathbf{u}) = n^{-1/2} \sum_{i=1}^n \hat{D}_i(\mathbf{u}) \varphi_i$, where $\phi_1, \dots, \phi_n$ are independent standard normal random variables. The distribution of $\mathbf{G}(\cdot)$ is asymptotically equivalent to the distribution of $\mathbf{G}^*(\cdot)$ given the observed data sequence. Hence, the distributions of T_1 and T_2 under the null hypothesis can be approximated respectively by the conditional distributions of

$T_1^* = \sup_{\mathbf{u} \in \Delta} \|\mathbf{G}^*(\mathbf{u})\|$ and $T_2^* = \sum_{\mathbf{u} \in \Delta} \|\mathbf{G}^*(\mathbf{u})\|$, which can be approximated by repeatedly generating, say 1000, sets of independent normal random variables $\phi_1, \dots, \phi_n$ while holding the observed data sequence fixed.

6. Simulation studies

We conducted a simulation study to assess the finite-sample performance of the proposed methods. Performance is illustrated under the following model with three popular link functions:

$$E\{Y_i(t) | \mathbf{X}_i, S_i\} = g^{-1} \{ \alpha_0(t) + \alpha_1(t) X_{1i}(t) + \beta X_{2i} + \gamma(t - S_i, \boldsymbol{\theta}) X_{3i} I(t > S_i) \}, \quad (10)$$

for $0 \leq t \leq \tau$ with $\tau = 3.5$, where $\alpha_0(t) = 0.2 \sqrt{t}$, $\alpha_1(t) = 0.1 \sin(t)$, $\gamma(u, \boldsymbol{\theta}) = \theta_1 \exp(-\theta_2 u)$, and $\boldsymbol{\zeta} = (\beta, \theta_1, \theta_2) = (0.1, 1.0, 0.5)$. The covariate X_{2i} is a Bernoulli random variable with success probability of 0.5, $X_{1i}(t) = (t/3 - X_{2i} + a_i)/6$ is time-dependent with a_i from the normal distribution $N(0, 0.3)$, X_{3i} is a uniform random variable on $[-1, 1]$, and S_i is uniform on $[0, 1]$. The proposed methods are examined under the identity link function, logarithm link function and logit link function. For the identity link and the logarithm link, the error $\varepsilon_i = Y_i(t) - E\{Y_i(t) | \mathbf{X}_i, S_i\}$ has a normal distribution with mean ϕ_i and variance 0.5^2 , and ϕ_i is $N(0, 1)$. For the logit link, $Y_i(t)$ is generated from Bernoulli distribution with the probability of success of $E\{Y_i(t) | \mathbf{X}_i, S_i\}$. The observation times follow a Poisson process with the proportional mean rate model $h(t | \mathbf{X}_i, S_i) = 1.5 \exp(0.7 X_{2i})$. The censoring times C_i are

generated from a uniform distribution on $[1.5, 8]$. There are approximately six observations per subject in $[0, \tau]$ and about 30% of subjects are censored before $\tau = 3.5$. The Epanechnikov kernel $K(t) = 0.75(1 - t^2)K(|t| - 1)$, the bandwidth formula $h = 4\sigma_T n^{-1/3}$, and unit weight function $W_A(t) = 1$ are used. We take $t_1 = h/2$ and $t_2 = \tau - h/2$ in the estimating functions (4) to avoid larger variations on the boundaries.

The performances of the estimators for ζ and $\alpha(t) = (\alpha_0(t), \alpha_1(t))$ at a fixed time t are measured through the Bias, the sample standard error of the estimators (SEE), the sample mean of the estimated standard errors (ESE), and the 95% empirical coverage probability (CP). Table 1 summarizes the Bias, SEE, ESE and CP for ζ under the three different link functions, and three different sample sizes ($n = 200; 400$ and 600). The bandwidth formula $h = 4\sigma_T n^{-1/3}$ yields $h = 0.68$ for $n = 200$, $h = 0.54$ for $n = 400$ and $h = 0.47$ for $n = 600$. Each entry of the table is calculated based on 1000 repetitions. Table 1 shows that the estimates are unbiased and there is good agreement between the estimated and empirical standard errors. The bias and variance of the estimators decreases as the sample size increases. The coverage probabilities are close to the 95% nominal level. Additional simulations not presented here show that the proposed methods are not overly sensitive to the bandwidth; they work well for C in the range of $[3, 5]$.

Figure 1 plots the bias, SSE, ESE and CP for the estimators of $\alpha_0(t)$ and $\alpha_1(t)$ over the time interval $[0, 3.5]$ with $n = 400$ and bandwidth $h = 0.54$ for the three different link functions. The plots show that the estimates are close to the true values and the ESE provides a good approximation for the SSE of the pointwise estimators. The empirical coverage probabilities are close to the 95% nominal level.

Next, we examine the finite-sample performance of the proposed test statistics under model (10). The simulation examines the parametric forms of the covariate-varying effects $\gamma(u, \theta)$ while leaving other model specifications unchanged. The sizes of the test statistics T_1 and T_2 are examined under the null model $M_0: \gamma(u, \theta) = \theta_1 \exp(-\theta_2 u)$ with $\theta = (\theta_1, \theta_2) = (1.0, 0.5)$. The powers of the tests are examined under three alternative models $M_1: \gamma(u) = 1 - 0.4u$, $M_2: \gamma(u) = 1 - 0.8u + 0.25u^2$ and $M_3: \gamma(u) = 1 - \sin(2u)$, respectively. Table 2 shows the observed sizes and powers of the test statistics T_1 and T_2 at significance level 0.05. Each entry is based on 1000 Gaussian multiplier samples and 1000 simulations. The observed sizes are close to their nominal level for all three link functions and for sample sizes $n = 200, 400$ and 600 . The powers of the tests increase as sample size increases. The powers of the tests also increase as the alternative models move from M_1 to M_3 which represent increasing departure from the null model M_0 . The power of T_2 is slightly higher than the power of T_1 for all cases.

7. Real data application

We apply the newly developed methods to investigate treatment strategies dependent on the development of drug resistance in the ACTG 244 trial. ACTG 244 was a randomized, double-blind trial that evaluated the clinical utility of monitoring HIV infected patients taking Zidovudine (ZDV) monotherapy for occurrence of the T215Y/F ZDV resistance mutation. When a subject developed the 215 mutation, the subject was randomized to

continue ZDV, add ddI, or add both ddI and NVP. ACTG 244 began enrollment in February 1994, and among the 289 enrollees, 284 were dispensed ZDV, of whom 57 developed T215Y/F. Forty-nine of these subjects were randomized to ZDV ($n=17$), ZDV+ddI ($n=15$), or ZDV+ddI+NVP ($n=17$), and the other 8 subjects went off treatment prior to randomization. T215Y/F mutation status was determined by RT-PCR (Larder et al., 1991) and was measured at study entry and every 8 weeks thereafter, with variability in visit dates across individuals. The primary study outcome CD4 cell count, a well known independent predictor of AIDS/death (Kaufmann et al., 1998), was measured on the same visit schedule. We investigate the effect of treatment switching on longitudinal square root CD4 cell count, and also investigate the association of the timing of treatment switching on square root CD4 cell count.

In addition to the above investigations, we also applied the methods to a distinct objective that arose from the Data Safety Monitoring Board's independent review of the study data in September 1996. Following this review, all subjects were offered randomization to the ZDV+ddI or ZDV+ddI+NVP arms with six months of additional follow-up. Of the 227 subjects who remained on ZDV treatment without the T215Y/F mutation, 137 were still taking ZDV at the time of the interim review and were randomized to ZDV+ddI ($n=69$) or ZDV+ddI+NVP ($n=68$); the remaining 90 subjects went off treatment prior to the interim review. As such, ACTG 244 investigated two treatment randomization strategies, the first comparing ZDV vs. ZDV+ddI vs. ZDV+ddI+NVP in subjects who acquired the T215Y/F mutation, and the second comparing ZDV+ddI vs. ZDV+ddI+NVP in subjects taking ZDV who did not have the T215Y/F mutation. We analyze the effects of treatment switching separately for these two investigations in Section 7.1 and Section 7.2, respectively. The main reason for the separate analyses rather than a single combined analysis is that the study populations for inference are fundamentally different, given the large impact of the T215Y/F drug resistance mutation on CD4 cell count. An additional reason is that the time to develop the T215Y/F mutation introduces informative censoring, and thus our analysis of the second randomization strategy that only includes subjects who did not develop the mutation avoids making a false assumption of non-informative censoring. A final reason for conducting two separate analyses is that the assigned treatments are different in the first and second randomizations.

7.1 Analysis of the effects of switching treatments after drug-resistant virus was detected (First randomization)

First, we examine the effects of switching treatments following detection of the T215Y/F mutation. Let $Y(t)$ be the square root of CD4 count at t years since study entry, Z_1 be Gender (1 if Female; 0 if Male), Z_2 be Age in years at study entry, Z_3 and Z_4 be dummy variables coding race ($Z_3 = 1$ if white and 0 otherwise, $Z_4 = 1$ if black and 0 otherwise). Let S_1 be the time from study entry until the first randomization based on occurrence of the T215Y/F mutation (the treatment switching time), where we set $S_1 = 3$ years, a number longer than the study duration, for subjects who did not experience the mutation. Then $U_1(t) = t - S_1$ is the time elapsed from the T215Y/F mutation-based treatment randomization. Let $T_{A1}(t) = 1$ if $t > S_1$ and randomized to ZDV and 0 otherwise, $T_{A2}(t) = 1$ if $t > S_1$ and randomized to ZDV+ddI and 0 otherwise, and $T_{A3}(t) = 1$ if $t > S_1$ and randomized to ZDV

+ddI+NVP and 0 otherwise. All three indicators are zero prior to detection of the mutation. This analysis includes all $n = 284$ enrolled subjects dispensed ZDV monotherapy. The eight subjects who were off treatment prior to the first randomization as well as the 90 subjects who were off treatment prior to the interim review are censored at the time of drop-off. In addition, Section 7.1 focuses on the treatment comparison in subjects who acquired the T215Y/F mutation; the time of the second randomization is treated as the censoring time for the 137 subjects who did not develop the mutation and were randomized at the interim review.

The analysis is conducted using the following model:

$$Y_i(t) = \alpha_0(t) + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_3 Z_{3i} + \beta_4 Z_{4i} + \gamma_1(U_{1i}(t), \theta_1) T_{A1i}(t) + \gamma_2(U_{1i}(t), \theta_2) T_{A2i}(t) + \gamma_3(U_{1i}(t), \theta_3) T_{A3i}(t) + \varepsilon_i(t), \quad (11)$$

for $t \in [0, \tau]$ where $\tau = 2.5$ years. We assume that $\gamma_k(u, \theta_k)$, $k = 1, 2, 3$, are the second order polynomial functions. Let $\gamma_1(u, \theta_1) = \theta_{10} + \theta_{11}u + \theta_{12}u^2$, $\gamma_2(u, \theta_2) = \theta_{20} + \theta_{21}u + \theta_{22}u^2$, and $\gamma_3(u, \theta_3) = \theta_{30} + \theta_{31}u + \theta_{32}u^2$, where $\theta_1 = (\theta_{10}, \theta_{11}, \theta_{12})$, $\theta_2 = (\theta_{20}, \theta_{21}, \theta_{22})$ and $\theta_3 = (\theta_{30}, \theta_{31}, \theta_{32})$. The 3-fold cross-validation bandwidth selection yields $h = 0.41$ while the bandwidth formula $h = C\hat{\sigma}_{TT}^{-1/3}$ yields $h = 0.41$ for $C = 4$ and $h = 0.51$ for $C = 5$. Our analysis uses $h = 0.41$; the results using $h = 0.51$ are almost the same.

The width of the range of the observed values for $U_{1i}(t)$, $t \in [0, 2.5]$, is 2.25. The test statistics T_1 and T_2 calculated with $u \in [0, 2]$ yield p -values of 0.138 and 0.228, respectively, suggesting that the quadratic functions $\gamma_k(u, \theta_k)$ for $k = 1, 2, 3$ fit well to the data. The tests also indicate inadequacy of the linear functions for $\gamma_k(u, \theta_k)$, $k = 1, 2, 3$, with p -values of 0.016 for T_1 and 0.036 for T_2 .

Using the quadratic functions, the estimates of parameters are presented in the first block of Table 3. The estimations of $\alpha_0(t)$, $\gamma_1(u, \theta_1)$, $\gamma_2(u, \theta_2)$ and $\gamma_3(u, \theta_3)$ are presented in Figure 2, along with their 95% pointwise confidence intervals. The estimation of $\gamma_k(u, \theta_k)$, $k = 1, 2, 3$, are plotted on $[0, 2]$.

The results show that CD4 cell counts are significantly higher for older individuals with p -value of 0.029, females tend to have lower CD4 cell counts with p -value of 0.068, and race is not a significant factor. It appears that the downward trend in CD4 cell counts is more apparent in continuing with the monotherapy ZDV (Figure 2(b)) than switching to the combination therapies (Figure 2(c) and (d)). This analysis points to positive benefits of switching to the combination therapies as compared to continuing with ZDV monotherapy ZDV even after drug-resistant virus was detected.

7.2 Analysis of the effects of switching treatments before drug-resistant virus was detected (Second randomization)

After independent review of the study data, all subjects were offered randomization to ZDV +ddI or ZDV+ddI+NVP with six months of additional follow-up. In this section, we examine the effects of this switching of treatments before the T215Y/F mutation was

detected. This analysis excludes subjects who developed the T215Y/F mutation to allow answering the study question for the sub-population without the mutation and because the time to develop T215Y/F likely introduces dependent censoring. The 90 subjects who were off treatment prior to the interim review without having the T215Y/F mutation are censored at the time of drop-off. Let S_2 be the time of the second randomization after the interim review and $U_2(t) = t - S_2$. We define $T_{B2}(t) = 1$ if $t > S_2$ and randomized to ZDV+ddI and 0 otherwise, $T_{B3}(t) = 1$ if $t > S_2$ and randomized to ZDV+ddI+NVP and 0 otherwise. $T_{B2}(t) = 0$ and $T_{B3}(t) = 0$ indicate a subject is on ZDV at time t before the interim review.

The data are analyzed using the following model:

$$Y_i(t) = \alpha_0(t) + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_3 Z_{3i} + \beta_4 Z_{4i} + \gamma_2(U_{2i}(t), \theta_2) T_{B2i}(t) + \gamma_3(U_{2i}(t), \theta_3) T_{B3i}(t) + \varepsilon_i(t), \quad (12)$$

for $t \in [0, 2.5]$. Similar to the previous analysis, we use the second order polynomial functions for $\gamma_2(u, \theta_2)$ and $\gamma_3(u, \theta_3)$. The range of the observed values for $U_{2i}(t)$, $t \in [0, 2.5]$, is $[0, 0.70]$. The p-values of the test statistics T_1 and T_2 with $\alpha = [0, 0.7]$ are 0.254 and 0.144, respectively, indicating that there is no significant departure from the hypothesized parametric quadratic functions for $\gamma_2(u, \theta_2)$ and $\gamma_3(u, \theta_3)$.

The results of parameter estimation are given in the second block of Table 3. The estimations of $\alpha_0(t)$, $\gamma_2(u, \theta_2)$ and $\gamma_3(u, \theta_3)$ are presented in Figure 3 along with 95% pointwise confidence intervals. Figure 3 shows CD4 cell counts rise significantly for subjects who switch to ZDV+ddI or ZDV+ddI+NVP compared to ZDV. The estimated switching-treatment effects and 95% confidence intervals are above zero, suggesting that switching to ZDV+ddI or ZDV+ddI+NVP improves CD4 counts for patients who had not yet developed the T215Y/F drug resistance mutation.

The ACTG 244 study was previously analyzed by the standard linear regression statistical methods in an unpublished manuscript. The previous analysis was inefficient because the longitudinal nature of the data were not considered. A brief description of these methods are summarized in Web Appendix D. To compare our data analysis with some benchmark methods for longitudinal data analysis, we also analyzed the data using the SAS procedure Proc Glimmix for generalized linear mixed models (GLMM) under some comparable models which are presented in Web Appendix D. These results are in line with the results by using our methods. The analysis using Proc Glimmix yields somewhat narrower confidence intervals, possibly due to the parametric specification of the baseline function $\alpha_0(t)$ and more efficient weight function selection since the methods used completely parametric models. There are no existing methods for testing the functional forms of covariate effects even under the GLMM.

8. Concluding remarks

This article is motivated by investigating the exposure-varying effects of adaptive treatment randomizations on longitudinal biomarkers, illustrated by the ACTG 244 AIDS clinical trial. We developed estimation and hypothesis testing procedures for a generalized semiparametric

varying-coefficient model for longitudinal data with a general link function. The choice of weight process is an important and complicated issue. We conducted some investigation into the two-stage estimation procedure for choosing the weight function within the framework of the marginal approach. The simulation results presented in Appendix C of the Web-based Supplementary Material show that the two-stage estimation procedure can be adopted to improve efficiency, where the first stage estimator is based on the unit weight and the second stage estimator is obtained with the weight estimated based on the first stage estimator. This article considers parametric forms $\gamma(U_i(t), \theta)$ of the covariate-varying effects for $\gamma(U_i(t))$. Nonparametric modeling of covariate-varying effects would provide greater flexibility when sufficient data are available. The theoretical development for estimating the nonparametric component $\gamma(u)$ is, however, significantly different from that considered in this paper because the nonparametric functions $\alpha(t)$ and $\gamma(u)$ have different domains, yet the smoothing for $\alpha(t)$ and $\gamma(u)$ cannot be totally separated because $\gamma(u)$ is a function of the covariate process $U_i(t)$ that may vary with t . This merits future research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the Editor, the Associate Editor and two referees for their constructive comments and suggestions that greatly improved this article. This research was partially supported by NIAID NIH award number R37AI054165, and the research of Yanqing Sun was partially supported by National Science Foundation grant DMS-1208978, DMS-1513072 and the Reassignment of Duties fund provided by the University of North Carolina at Charlotte. The authors thank the AIDS Clinical Trials Group for providing the ACTG 244 data, in particular Ronald Bosch and Justin Ritz for preparing the data set, reviewing the manuscript, and helpful discussions. We also wish to thank the ACTG 244 study participants and study team, including the study chairs Douglas L. Mayers & Thomas C. Merigan. The project described was supported by Award Numbers U01 A038855, AI038858, AI068634 and AI068636 from the National Institute of Allergy and Infectious Diseases and supported by National Institute of Mental Health (NIMH), National Institute of Dental and Craniofacial Research (NIDCR). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Allergy and Infectious Diseases or the National Institutes of Health.

References

- Fan, J., Gijbels, I. Local Polynomial Modelling and Its Applications. Chapman and Hall; London: 1996.
- Fan J, Huang T, Li R. Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of the American Statistical Association*. 2007; 102(478):632–641. [PubMed: 19707537]
- Fan J, Li R. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association*. 2004; 99(467):710–723.
- Japour AJ, Welles S, D'Aquila RT, Johnson VA, Richman DD, Coombs RW, Reichelderfer PS, Kahn JO, Crumacker CS, Kuritzkes DR. Prevalence and clinical significance of zidovudine resistance mutations in human immunodeficiency virus isolated from patients after long-term zidovudine treatment. *Journal of Infectious Diseases*. 1995; 171(5):1172–1179. [PubMed: 7538548]
- Jones MC, Marron JS, Park BU. A simple root n bandwidth selector. *The Annals of Statistics*. 1991; 19(4):1919–1932.
- Kaufmann D, Pantaleo G, Sudre P, Telenti A. CD4-cell count in HIV-1-infected individuals remaining viraemic with highly active antiretroviral therapy (HAART). *Lancet*. 1998; 351(9104):723–724. [PubMed: 9504519]

- Larder BA, Kellam P, Kemp SD. Zidovudine resistance predicted by direct detection of mutations in DNA from HIV-infected lymphocytes. *AIDS*. 1991; 5(2):137–144. [PubMed: 1709563]
- Lin DY, Wei L-J, Ying Z. Checking the cox model with cumulative sums of martingale-based residuals. *Biometrika*. 1993; 80(3):557–572.
- Mellors JW, Munoz A, Giorgi JV, Margolick JB, Tassoni CJ, Gupta P, Kingsley LA, Todd JA, Saah AJ, Detels R, et al. Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. *Annals of Internal Medicine*. 1997; 126(12):946–954. [PubMed: 9182471]
- Qu A, Li R. Quadratic inference functions for varying-coefficient models with longitudinal data. *Biometrics*. 2006; 62(2):379–391. [PubMed: 16918902]
- Rice JA, Silverman BW. Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society Series B (Methodological)*. 1991; 10(6): 233–243.
- Sun Y, Qian X, Shou Q, Gilbert PB. Analysis of two-phase sampling data with semiparametric additive hazards models. *Lifetime Data Analysis*. 2016; doi: 10.1007/s10985-016-9363-2
- Sun Y, Sun L, Zhou J. Profile local linear estimation of generalized semiparametric regression model for longitudinal data. *Lifetime Data Analysis*. 2013; 19:317–349. [PubMed: 23471814]
- Tian L, Zucker D, Wei LJ. On the Cox model with time-varying regression coefficients. *Journal of the American Statistical Association*. 2005; 100:172–183.
- Zhou H, Wang C-Y. Failure time regression with continuous covariates measured with error. *Journal of the Royal Statistical Society: Series B*. 2000; 62(4):657–665.

Figure 1. Plots of bias, SEE, ESE and CP of $\hat{\alpha}_0(t)$ and $\hat{\alpha}_1(t)$ under three link functions with $n=400$, $h = 0.54$ and unit weight $W_\lambda(t) = 1$ based on 1000 simulations. The figures in the right panel are for $\alpha_0(t) = 0.2\sqrt{t}$, and the figures in the left panel are for $\alpha_1(t) = 0.1\sin(t)$.

Figure 2. Estimated effects of adaptive treatment randomizations *after* the T215Y/F mutation based on the ACTG 244 data using unit weight and $h = 0.41$. Figure (a) shows the estimated baseline function $\hat{\alpha}_0(t)$ with 95% pointwise confidence intervals; (b), (c) and (d) show the point and 95% confidence interval estimates of $\gamma_k(u)$, $k = 1, 2, 3$, respectively, under model (11).

Figure 3. Estimated effects of adaptive treatment randomizations *before* the T215Y/F mutation based on the ACTG 244 data using unit weight and $h = 0.41$. Figure (a) shows the estimated baseline function $\hat{\alpha}_0(t)$ with 95% pointwise confidence intervals; (b) and (c) show the point and 95% confidence interval estimates of $\gamma_k(u)$, $k = 1, 2$, respectively, under model (12).

Summary of Bias, SEE, ESE and CP for β , θ_1 and θ_2 for three different link functions and sample sizes using unit weight and bandwidth $h = 4\sigma_{\epsilon}n^{-1/3}$ which yields $h = 0.68, 0.54$ and 0.47 for $n = 200, n = 400$ and $n = 600$, respectively, based on 1000 simulations.

n	$\beta=1$				$\theta_1=1$				$\theta_2=.5$			
	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
Identity link												
200	.0073	.2924	.2638	.912	.0173	.1755	.1675	.925	.0259	.1949	.1958	.930
400	.0011	.1911	.1863	.941	.0086	.1174	.1172	.947	.0230	.1446	.1329	.943
600	.0023	.1607	.1524	.932	.0040	.0966	.0941	.945	.0098	.1009	.1025	.941
Logarithm link												
200	-.0018	.2034	.1856	.934	.0067	.1406	.1426	.956	.0167	.1306	.1299	.955
400	-.0019	.2062	.1861	.931	.0060	.1421	.1429	.955	.0144	.1281	.1293	.956
600	-.0018	.1118	.1069	.945	.0066	.0793	.0798	.947	.0029	.0698	.0685	.947
Logit link												
200	.0030	.1314	.1266	.941	.0159	.1857	.1829	.950	.0095	.1784	.1785	.944
400	-.0069	.1106	.1104	.960	.0127	.1689	.1645	.946	.0185	.1614	.1542	.940
600	-.0022	.0876	.0892	.947	.0107	.1332	.1308	.945	.0078	.1260	.1226	.951

Table 2

Observed sizes and powers of the test statistics T_1 and T_2 for three different link functions using unit weight based on 1000 Gaussian multiplier samples and 1000 simulations. The bandwidth is calculated based on $h = 4\sigma_{\eta t}^{-1/3}$ which yields $h = 0.68, 0.54$ and 0.47 for $n = 200, n = 400$ and $n = 600$, respectively.

n	Size		Power					
	M_0		M_1		M_2		M_3	
	T_1	T_2	T_1	T_2	T_1	T_2	T_1	T_2
Identity link								
200	.059	.042	.341	.368	.682	.714	.999	.987
400	.063	.065	.580	.632	.945	.963	1.00	.999
600	.060	.044	.817	.854	.996	1.00	1.00	1.00
Logarithm link								
200	.053	.056	.532	.662	.912	.970	.992	.996
400	.054	.048	.856	.917	.987	.993	.998	1.00
600	.046	.043	.911	.930	.988	.995	1.00	1.00
Logit link								
200	.045	.051	.126	.175	.267	.285	.965	.960
400	.056	.064	.250	.354	.562	.623	.992	.994
600	.061	.054	.655	.719	.788	.854	.998	1.00

Table 3

Estimated effects of adaptive treatment randomizations based on the ACTG 244 data using $h = 0.41$ and unit weight.

Effect	Parameter	Estimate	Standard deviation	95% Confidence limits	p-value
Treatment effects after the T215Y/F mutation under model (11) (first randomization)					
Gender	β_1	-1.3155	0.7216	-2.7298 0.0988	0.0683
Age	β_2	0.0619	0.0284	0.0063 0.1174	0.0292
Race	β_3	-0.7575	0.7746	-2.2758 0.7607	0.3281
	β_4	-0.8163	0.8565	-2.4950 0.8623	0.3405
$T_{A1}(t)$	θ_{10}	-1.1593	1.0956	-3.3066 0.9881	0.2900
	θ_{11}	4.3785	2.9895	-1.4810 10.2380	0.1430
	θ_{12}	-4.0586	1.4271	-6.8558 -1.2614	0.0045
$T_{A2}(t)$	θ_{20}	0.1222	0.8782	-1.5991 1.8436	0.8893
	θ_{21}	-2.7256	2.0925	-6.8270 1.3757	0.1927
	θ_{22}	1.0751	1.1880	-1.2534 3.4035	0.3655
$T_{A3}(t)$	θ_{30}	-2.2771	0.8704	-3.9831 -0.5710	0.0089
	θ_{31}	4.7834	1.7902	1.2745 8.2922	0.0075
	θ_{32}	-2.5625	0.8431	-4.2149 -0.9101	0.0024
Treatment effects before the T215Y/F mutation under model (12) (second randomization)					
Gender	β_1	-0.6674	0.6396	-1.9211 0.5862	0.2967
Age	β_2	0.0192	0.0250	-0.0299 0.0683	0.4438
Race	β_3	-1.0689	0.7223	-2.4846 0.3469	0.1389
	β_4	-1.7410	0.7750	-3.2600 -0.2221	0.0247
$T_{B2}(t)$	θ_{20}	0.2354	0.5963	-0.9334 1.4042	0.6930
	θ_{21}	3.4616	3.6360	-3.6651 10.5882	0.3411
	θ_{22}	-0.5598	5.5558	-11.4491 10.3295	0.9197
$T_{B3}(t)$	θ_{30}	0.7469	0.5363	-0.3043 1.7981	0.1637
	θ_{31}	5.5801	2.9732	-0.2474 11.4076	0.0605
	θ_{32}	-4.0646	4.4691	-12.8241 4.6949	0.3631