



UW Biostatistics Working Paper Series

11-13-2005

Estimating a Treatment Effect with Repeated Measurements Accounting for Varying Effectiveness Duration

Ying Qing Chen

Fred Hutchinson Cancer Research Center, yqchen@u.washington.edu

Jingrong Yang

University of California at Berkeley, jingrong@berkeley.edu

Su-Chun Cheng

University of California at San Francisco, scheng@biostat.ucsf.edu

Suggested Citation

Chen, Ying Qing; Yang, Jingrong; and Cheng, Su-Chun , "Estimating a Treatment Effect with Repeated Measurements Accounting for Varying Effectiveness Duration" (November 2005). *UW Biostatistics Working Paper Series*. Working Paper 265. <http://biostats.bepress.com/uwbiostat/paper265>

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

Copyright © 2011 by the authors

Estimating a treatment effect with repeated measurements accounting for varying effectiveness duration

BY [Y. Q. CHEN](#)

Program in Biostatistics, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, U.S.A.

yqchen@scharp.org

J. R. YANG

Programs in Biostatistics & Epidemiology, School of Public Health, University of California, Berkeley, California 94720, U.S.A.

jingrong@stat.berkeley.edu

AND S. CHENG

Department of Biostatistics & Epidemiology, University of California, San Francisco, California 94143, U.S.A.

scheng@biostat.ucsf.edu

SUMMARY

To assess treatment efficacy in clinical trials, certain clinical outcomes are repeatedly measured over time for the same subject. These outcomes can be regarded as a function of time. The difference in their mean functions between the treatment arms usually characterises a treatment effect. Due to the potential existence of subject-specific treatment effectiveness lag and saturation times, erosion of treatment effect may occur during the observation period. Instead of using ad hoc parametric or purely nonparametric time-varying coefficients in statistical modeling, we first propose to model the treatment effectiveness durations, which are the varying time intervals between the lag and saturation times. Then some mean response models are used to include such treatment effectiveness durations. Our methodologies are demonstrated by simulations and an application to the dataset of a landmark HIV/AIDS clinical trial of short-course nevirapine against mother-to-child HIV vertical transmission during labour and delivery.

Some key words: Effect erosion; HIV/AIDS clinical trials; Longitudinal outcomes; Mean response model; Semiparametric analysis; Time-varying coefficient; Treatment effectiveness duration.

1 INTRODUCTION

In clinical trials, repeated measurements of the same subject are often collected during a given period of time to assess the efficacy of a new treatment. Assume that (Y_1, Y_2, \dots, Y_m) is the vector of the repeated measurements collected at the time points $T_1 < T_2 < \dots < T_m$, respectively, from the same subject during the time period $[0, C]$, where C is the potential maximum length of the observation time period. These repeated measurements can be considered as observations of an underlying random response curve over time, $\{Y(t); t \geq 0\}$, with $Y(T_j) = Y_j$, $j = 1, 2, \dots, m$. Regression models are the main tool used in statistical literature to model the means of these response curves and to identify a treatment effect. One widely used model for $\{Y(t); t \geq 0\}$ was proposed in [Zeger & Diggle \(1994\)](#) and later generalised in [Lin & Ying \(2001\)](#),

$$E\{Y(t) \mid Z(s); 0 \leq s \leq t\} = \mu(t) + \beta^T Z(t), \quad (1)$$

where $Z(\cdot)$ is the covariate vector of p_1 -dimension, $\beta \in \mathfrak{R}^{p_1}$ is the associated regression parameter and $\mu(\cdot)$ is some baseline function for $Z(t) \equiv 0$, respectively. Here, τ denotes vector transpose. Usually, neither $\mu(t)$ nor the error structure of $\{Y(t), t \geq 0\}$ has to be specified in (1). Such a semiparametric model setup would usually allow for substantial model flexibility when the treatment efficacy parameter β is of major interest.

In model (1), the constant β serves as a concise summary measure of the difference in the mean responses over time, for example, when $Z(t)$ is a treatment indicator. If the assumption of constant β appears valid, a consistent estimator of β can be used to characterise and make inference on the treatment efficacy. This assumption, however, may be undermined in practice by the phenomenon of ‘erosion of regression effect,’ as termed in a recent paper by [O’Quigley & Natarajan \(2005\)](#) on the proportional hazards model of survival outcomes. For the repeated measurements in model (1), similar erosion phenomenon on the constant β may occur when the treatment effectiveness lag time and the saturation time possibly exist.

Treatment effectiveness lag time is the time after which the treatment becomes fully effective. Its existence has long been noted and studied, for example, in [Halperin et al. \(1968\)](#), [Zucker & Lakatos \(1990\)](#) and [Chen et al. \(2002\)](#). Treatment saturation time is the time after which the full treatment effect starts to attenuate or diminish due to reasons such as cumulative buildup of drug resistance ([Eshleman et al., 2001](#); [Komarova & Wodarz, 2005](#); [Wu et al., 2005](#)). The interval between the lag time and the saturation time is thus the treatment effectiveness duration on which the treatment is expected to take full effect.

Because of biological heterogeneity among subjects, the treatment effectiveness durations are not necessarily identical for all subjects. The variation among subjects may lead to the phenomenon of effect erosion, as shown in a later section.

Effect erosion due to varying effectiveness durations is a particular type of departure of the regression parameter from being constant. In the statistical literature, one mathematically convenient approach to handling the effect erosion in model (1) is to allow the regression parameter to be time-varying as $\beta(t)$ in

$$E\{Y(t) \mid Z(s), 0 \leq s \leq t\} = \mu(t) + \beta(t)^T Z(t), \quad (2)$$

similar to that proposed in Hoover et al. (1998). Here, $\beta(\cdot)$ can be specified in some ad hoc parametric form. Or, it can be nonparametrically assumed and estimated by various methods, such as smoothing splines and locally weighted polynomials in Hoover et al. (1998), component-wise smoothing splines in Chiang et al. (2001) and basis function approximation in Huang et al. (2002). The most practical values of flexible nonparametric $\beta(\cdot)$ for model (2) in real applications are in, for example, model adequacy assessment or model-based prediction/forecasting. Although it can be used to explore the time-varying pattern in covariate effect, the time-varying coefficient itself may not result in direct summary of treatment effect or lead to straightforward comparison of treatment effects among the clinical trials.

In the rest of this article, we focus on the treatment effect estimation by taking into account the effect erosion due to the possible varying treatment effectiveness durations. Some joint distributions for the bivariate treatment effectiveness lag time and the effectiveness duration are first proposed. The mean response model (1) of the constant β is subsequently extended by allowing changepoints to accommodate the varying treatment effectiveness durations. Inference procedures including hypothesis testing and interval estimation are studied in §3. Numerical studies are presented in §4. Further model extensions are laid out in §5. Some concluding remarks and discussion are presented in §6.

2 STATISTICAL MODELS

2.1 Models for treatment effectiveness duration

Let U and V be the possible treatment effectiveness lag and saturation times, respectively, where $0 \leq U \leq V$, for a subject in the study. Then the time interval $[U, V]$ is the effectiveness

duration. Denote $W = V - U$ the length of the effectiveness duration. Consider the joint density function of the bivariate vectors of (U, W) , $f_{U,W}(u, w; \theta)$, for the lag time and the effectiveness duration, where θ is the parameter of p_2 -dimension. When U and W are independent, $f_{U,W}(u, w; \theta)$ is simply the product of the marginal density functions, $f_U(u)$ and $f_W(w)$, say, of U and W , respectively. There are a variety of distributions that can be used for $f_U(u)$ and $f_W(w)$, such as the exponential, Weibull or Gamma distributions.

In reality, since U and W reflect a human subject's individual reaction to the treatment, they are often not necessarily independent among subjects. In some instances, when a subject has relatively a shorter metabolic cycle, he or she usually responds to the treatment faster and reaches the full effect sooner. At the same time, the treatment's full effect is maintained for a relatively shorter duration of time. On the other hand, when the metabolic cycle tends to be longer, both the treatment lag time and the effectiveness duration tend to be longer. Thus U and W are positively correlated. The shared frailty models (Oakes, 1989) can be then applied to the bivariate times (U, W) .

Assume that G is the subject's underlying frailty, representing the effect of underlying metabolism on U and W simultaneously. Let its density function be $f_G(g; \alpha)$, where α is the parameter. Conditional on G , the metabolism affects the hazard functions for U and W by $G\lambda_U(t)$ and $G\lambda_W(t)$, respectively. Thus the bivariate survival function of (U, W) is,

$$S_{U,W}(u, w | G = g) = \text{pr}\{U > u, W > w | g\} = \exp[-g\{\Lambda_U(u) + \Lambda_W(w)\}],$$

where $\Lambda_U(t) = \int_0^t \lambda_U(s)ds$ and $\Lambda_W(t) = \int_0^t \lambda_W(s)ds$, respectively. By integrating out the g 's, the marginal bivariate survival function for (U, W) is $S_{U,W}(u, w) = E \exp[-g\{\Lambda_U(u) + \Lambda_W(w)\}]$, which is also the Laplace transform of G 's distribution, $L(\cdot)$, say, at $\Lambda_U(u) + \Lambda_W(w)$. As a result,

$$f_{U,W}(u, w) = \lambda_U(u)\lambda_W(w)L^{(2)}\{\Lambda_U(u) + \Lambda_W(w)\}.$$

With different choices of $f_G(\cdot)$, $\lambda_U(\cdot)$ and $\lambda_W(\cdot)$, $f_{U,W}(\cdot, \cdot)$ embraces a variety of choices of bivariate distributions for (U, W) .

One widely used family is the Clayton-Oakes model, or the Gamma frailty model (Clayton, 1978; Oakes, 1989). In this model, G is assumed to follow the Gamma density function, $f_G(g; \alpha) = (g/\alpha_2)^{\alpha_1-1} \exp(-g/\alpha_2)/\{\alpha_2\Gamma(\alpha_1)\}$, where $\Gamma(\alpha) = \int_0^\infty s^{\alpha-1} \exp(-s)ds$. Therefore, the bivariate density function for (U, W) is thus

$$\alpha_1(1 + \alpha_1)\alpha_2^{-\alpha_1} \lambda_U(u)\lambda_W(w)\{\alpha_2^{-1} + \Lambda_U(u) + \Lambda_W(w)\}^{-\alpha_1-2}.$$

In practice, given the multiplicative form of g on $\lambda_U(\cdot)$ and $\lambda_W(\cdot)$, it is usually sensible to further assume that $\alpha_1 = 1/\alpha_2$, which leads to $EG \equiv 1$.

2.2 Mean response models with varying effectiveness durations

Suppose that the covariate vector of $Z(t) = (Q(t)^T, R(t)^T)^T$ mainly consists of two types of covariates: $Q(t)$ are the covariates that take full effect only on the effectiveness duration, and $R(t)$ are the concomitant variables that are not affected by the effectiveness durations. When the treatment effectiveness duration defined by (U, W) is known, we consider the following model,

$$E\{Y(t) | Z(s), U, W; 0 \leq s \leq t\} = \mu(t) + \beta_Q^T Q(t) I(U \leq t \leq U + W) + \beta_R^T R(t), \quad (3)$$

where $I(\cdot)$ is the indicator function, and $(\beta_Q^T, \beta_R^T)^T = \beta$ are the parameters, respectively. That is, conditional on (U, W) , the mean response of the repeated measurements is $\mu(t) + \beta_Q^T Q(t) + \beta_R^T R(t)$ on the effectiveness duration $[U, U + W]$, and $\mu(t) + \beta_R^T R(t)$, otherwise. Thus the parameter β_Q characterises the subject-specific differences in the mean responses associated with $Q(t)$ on the effectiveness duration. When $Q(t)$ is the treatment indicator, β_Q represents the subject-specific treatment effect.

This model is indeed a changepoint model with two subject-specific changepoints at U and $U + W$, respectively. When $U = 0$ and $W = \infty$, it reduces to the model in (1). Most of the previous methodological research on changepoint models has focused on the detection of location and magnitude of changepoints, such as in Wu et al. (2001). The major interest of the current model (3) still lies in the regression parameter of β 's, while the changepoints and their distributions serve as nuisance to introduce the phenomenon of time-varying erosion on the regression parameter, as observed in the marginalised version of (3),

$$E\{Y(t) | Z(s); 0 \leq s \leq t\} = \mu(t) + \bar{\beta}_Q(t)^T Q(t) + \beta_R^T R(t), \quad (4)$$

with respect to the distribution of (U, W) , where $\bar{\beta}_Q(t) = \beta_Q H(t; \theta)$. Here, $H(t; \theta) = \int_0^t \int_{t-u}^\infty f_{U,W}(u, w; \theta) dw du$ is the 'erosion operator' that modifies the subject-specific constant regression parameter β_Q to the marginal time-varying coefficient $\bar{\beta}_Q(t)$, due to the varying treatment effectiveness durations of (U, W) . In §4, some specific examples are given to demonstrate the generally U-shaped $\bar{\beta}_Q(\cdot)$.

In fact, the erosion operator $H(\cdot)$ has the following properties:

1. $0 \leq H(t) \leq 1$;
2. $\lim_{t \rightarrow 0} H(t) = \lim_{t \rightarrow \infty} H(t) = 0$;
3. $H'(t) = \int_0^t f(u, t-u)du - \int_0^\infty f(t, w)dw$; $H^{(2)}(t) = f(t, 0) + \int_0^t \partial f(u, w)/\partial w|_{w=t-u}du - \int_0^\infty \partial f(u, w)/\partial u|_{u=t}dw$.

The first property mandates that the marginal $|\bar{\beta}_Q(t)|$ should be no larger than the subject-specific $|\beta_Q|$. The second property implies that the treatment efficacy diminishes to null at the beginning of a randomised trial for the short-term and also toward the long-run, which echoes the usual actual observation of regression effect erosion. The third property will allow us to calculate the time point at which $H(t)$ reaches its peak, and the turning points as well when the overall trend of $H(t)$ changes.

It is apparent that the two parameters, namely, the scale parameter of β_Q and the shape parameter of θ , jointly determine a specific $\bar{\beta}_Q(\cdot)$ in the marginalised model (4). When $\mu(\cdot)$ is parametric and the error structure of $Y(\cdot)$ is known, both parameters are identifiable in model (4) using the maximum likelihood method. In general, however, when $\mu(\cdot)$ is unspecified as in Zeger & Diggle (1994) or the error structure is further unknown as in Lin & Ying (2001), the parameters are identifiable only up to the function of $\bar{\beta}_Q(\cdot)$ itself. In the latter semiparametric settings, similar to those in Chen et al. (2002), it can be shown that the parameter of interest, β_Q , is identifiable if and only if the nuisance parameter θ is identifiable. In fact, the parameter θ is only meaningful and identifiable when $\beta_Q \neq 0$, as pointed out by Luo et al. (1997). When $\beta_Q = 0$, it is usually of less scientific interest to consider either the treatment effectiveness lag or the saturation time. Nevertheless, this is the type of problem discussed in Davies (1977, 1987), that is, the nuisance parameter θ is present only under the alternative of $\beta_Q \neq 0$. In the following sections, we will develop inference procedures by taking this complication into consideration.

3 INFERENCES

We first assume in this section that $\beta_Q \neq 0$ as in Luo et al. (1997). Suppose that there are n independent and identically distributed copies of $\{(Y_{ij}, T_{ij}, Z_{ij}, C_i); j = 1, 2, \dots, m_i\}$, $i = 1, 2, \dots, n$, in the collected dataset. Denote $N_i(t) = \sum_{j=1}^{m_i} I(T_{ij} \leq t)$ and assume that $E\{N_i(t)\} = \Omega(t)$ is unspecified. Let $\Delta_i(t) = I(C_i \geq t)$. Conditional on $Z_i(\cdot)$, $(Y_i(\cdot), C_i)$ are assumed to be independent. The true parameters hereinafter are denoted by their respective

counterparts with the subscript ‘*.’ For instance, the true parameters for β_Q and β_R in (3) are β_{Q^*} and β_{R^*} , respectively. Consider the cumulative sum of the residuals for the i th subject’s repeated measurements on $[0, t]$, $t > 0$, $X_i(t; \beta, \theta) = \int_0^t \{Y_i(s) - \nu_i(s; \beta, \theta)\} dN_i(s)$, where $\nu_i(t; \beta, \theta) = \beta_Q^T Q_i(t) H(t; \theta) + \beta_R^T R_i(t)$. Then

$$E\{dX_i(t) \mid Z_i(s), C_i; 0 \leq s \leq t, \beta_*, \theta_*\} = \Delta_i(t) d\Omega_\mu(t). \quad (5)$$

Here, $d\Omega_\mu(t) = \mu(t) d\Omega(t)$.

Let $M_i(t; \beta, \theta, \Omega_\mu) = X_i(t) - \int_0^t \Delta_i(s) d\Omega_\mu(s)$. Then $\{M_i(\cdot; \beta_*, \theta_*, \Omega_{\mu^*})\}$ are the zero-mean stochastic processes. Although the infinite-dimensional functions of $\mu(\cdot)$, $\Omega(\cdot)$ and $\Omega_\mu(\cdot)$ are unknown in (6), an estimator of the Breslow-type can be obtained for $\Omega_\mu(\cdot)$ by solving $\sum_i \int_0^\infty \Delta_i(t) dM_i(t; \beta, \theta, \Omega_\mu) = 0$. That is, $\widehat{\Omega}_\mu(t; \beta, \theta) = \int_0^t \sum_i dX_i(s; \beta, \theta) / \sum_i \Delta_i(s)$. Let $\widehat{M}_i(t; \beta, \theta) = X_i(t; \beta, \theta) - \int_0^t \Delta_i(s) d\widehat{\Omega}_\mu(s; \beta, \theta)$. Therefore, similar to those in Lin & Ying (2001) and Chen & Jewell (2001), the following estimating equations generalise the normal equations of the least-squares for the linear regression models to estimate the parameters in the proposed model (4),

$$\sum_{i=1}^n \int_0^\tau \Delta_i(t) \psi_i(t) d\widehat{M}_i(t; \beta, \theta, \Omega_\mu) = 0, \quad (6)$$

where $\psi_i(t)$ are the known smooth functions of $(p_1 + p_2)$ -dimension and measurable with respect to $\sigma\{Z_i(s), C_i; 0 \leq s \leq t, i = 1, 2, \dots, n\}$. Here, $\tau = \max_i\{C_i\}$. As suggested in Chen & Jewell (2001), $\psi(\cdot)$ can be chosen as $(Z(\cdot)^T, G(\cdot)^T)^T$, where the p_2 -dimensional $G(\cdot)$ does not belong to the linear span of 1 and Z . For example, $G(\cdot)$ is selected to be some nonlinear transformation of $\log Z$ or $Z \cdot Z$. In fact, straightforward algebra further leads to

$$\mathcal{E}(\beta, \theta) = \sum_{i=1}^n \int_0^\tau \Delta_i(t) \{\psi_i(t) - \bar{\psi}(t; \beta, \theta)\} dX_i(t; \beta, \theta) = 0, \quad (7)$$

where $\bar{\psi}(t) = \sum_i \Delta_i(t) \psi_i(t) / \sum_i \Delta_i(t)$. Assume that $\widehat{\beta}$ and $\widehat{\theta}$ are the solutions of β and θ in (7), respectively.

Let $\nu'_i(t)$ be the derivative of $\nu_i(t)$, $i = 1, 2, \dots, n$. Then $-n^{-1} \mathcal{E}'(\beta_*, \theta_*)$ goes to

$$D = E \left[\int_0^\tau \Delta_1(t) \{\psi_1(t) - \bar{\psi}_*(t)\} \nu'_1(t)^T d\Omega(t) \right],$$

where $\bar{\psi}_*(t)$ is the limit of $\bar{\psi}(t)$ almost surely, as $n \rightarrow \infty$. When $f_{u,w}(u, w)$ degenerates to 1 at $(u, w) = (0, \infty)$ and 0 otherwise, and $\psi_i(\cdot)$ are chosen to be $Z_i(\cdot)$, the proposed model

(4) becomes the model (1) and D reduces to the nonsingular matrix of \tilde{D} in Lin & Ying (2001). In general, when the elements in $\psi_i(\cdot)$ are not linearly related, D is nonsingular. Thus as shown in the Appendix, the solutions to $\mathcal{E}(\beta, \theta) = 0$ are strongly consistent under mild conditions as $n \rightarrow \infty$. If the total variation of $\psi_i(\cdot)$, $i = 1, 2, \dots, n$, are bounded, it is then true that

$$n^{-1/2}\mathcal{E}(\beta_*, \theta_*) \simeq n^{-1/2} \sum_{i=1}^n \int_0^\tau \Delta_i(t) \{\psi_i(t) - \bar{\psi}_*(t)\} dM_i(t; \beta_*, \theta_*).$$

By the Central Limit Theorem, it is also shown in the Appendix that $n^{-1/2}\mathcal{E}(\beta_*, \theta_*)$ is asymptotically normal with mean zero and the variance-covariance matrix,

$$\Sigma = E \left[\int_0^\tau \Delta_1(t) \psi(t) \{\psi_1(t) - \bar{\psi}_*(t)\} dM_1(t) \right]^{\otimes 2},$$

where $a^{\otimes 2}$ denotes aa^T . In addition, a Taylor's expansion of $\mathcal{E}(\hat{\beta}, \hat{\theta})$ at (β_*, θ_*) yields that $n^{1/2}(\hat{\beta}^T - \beta_*^T, \hat{\theta}^T - \theta_*^T)^T$ is asymptotically equivalent to $\{-\mathcal{E}'(\beta_*, \theta_*)/n\}^{-1} \cdot n^{-1/2}\mathcal{E}(\beta_*, \theta_*)$. Therefore, $\hat{\beta}$ and $\hat{\theta}$ are consistent, and $n^{1/2}(\hat{\beta}^T - \beta_*^T, \hat{\theta}^T - \theta_*^T)^T \rightarrow N(0, D^{-1}\Sigma D^{-1})$ in distribution in a neighbourhood of (β_*, θ_*) , where D and Σ can be approximated by their empirical counterparts, $\hat{D} = n^{-1} \sum_i \int_0^\tau \Delta_i(t) \{\psi_i(t) - \bar{\psi}(t)\} \nu_i'(t)^T dN_i(t)$, and $\hat{\Sigma} = n^{-1} \sum_i \int_0^\tau \Delta_i(t) \{\psi_i(t) - \bar{\psi}(t)\} d\hat{M}_i(t; \hat{\beta}, \hat{\theta})$, respectively.

To estimate the baseline $\mu(\cdot)$, consider the estimator of $\tilde{\mu}(t) = \bar{Y}(t) - \bar{\nu}(t; \hat{\beta}, \hat{\theta})$, where $\bar{Y}(t) = \sum_{i=1}^n \Delta_i(t) Y_i(t) / \sum_{i=1}^n \Delta_i(t)$ and $\bar{\nu}(t; \beta, \theta) = \sum_{i=1}^n \Delta_i(t) \nu_i(t; \beta, \theta) / \sum_{i=1}^n \Delta_i(t)$, respectively. When the observation times are observed in a continuous time scale, some smoothing technique has to be implemented to obtain a reasonable estimate. For example, the technique by Capra & Müller (1997) can be adapted to estimate $\mu(\cdot)$. Specifically, consider that the time interval $[0, \tau]$ is partitioned into L consecutive equidistant intervals: (t_{l-1}, t_l) , with $t_0 = 0$ and $l = 1, 2, \dots, L \rightarrow \infty$. Assume the smoothing parameter h such that $h \rightarrow 0$ and $n_* h \rightarrow \infty$, as $n_* \rightarrow \infty$, where n_* is the total number of observation time points. Coupled with additional conditions for the consistency in Capra & Müller (1997), a smoothed estimate of $\tilde{\mu}(\cdot)$ is

$$\hat{\mu}(t) = \arg \min_{a_0, a_1} \left[\sum_{l=1}^L K \left(\frac{t - t_l}{h} \right) \{\tilde{\mu}(t_l) - a_0 - a_1(t_l - t)\}^2 \right].$$

Here $K(s) = (1 - s^2)I(|s| \leq 1)$. Other smoothers including higher-order kernel smoothers or local fitting with high-order polynomials can be also used under the necessary conditions of linearity, consistency and consistency with needed rate in Capra & Müller (1997).

When $\beta_Q = 0$, the parameter θ is not identified in the model (4). As in Davies (1977, 1987), θ cannot be estimated under the null hypothesis and, subsequently, traditional large sample theory does not apply. If θ were known, however, we would only need to solve for $\beta = (\beta_Q^T, \beta_R^T)^T$ in the reduced $\mathcal{E}(\beta, \theta) = 0$ of (7), e.g., the first p_1 equations with $\psi(t) = Z(t)$. Denote $S(\beta_R, \theta) = \mathcal{E}(0, \beta_R, \theta) \widehat{\Sigma}^{-1}(\widehat{\beta}, \theta) \mathcal{E}(0, \beta_R, \theta)$. Then at the known θ , $S(\theta) = \inf_{\beta_R} S(\beta_R, \theta)$ is asymptotically χ_q^2 -distribution with $q = \dim\{Q(t)\}$ under $H_0 : \beta_Q = 0$ of no treatment efficacy, and larger $S(\theta)$ would lead to more evidence against H_0 as in Wei et al. (1990). Since θ is usually unknown, we consider the one-sided test statistic against $H_1 : \beta_Q > 0$ as

$$S = \sup_{\theta \in [\theta_l, \theta_u]} \{S(\theta)\},$$

where θ_l and θ_u are the lower and upper limits of the possible values of θ , respectively. Following Davies (1987), for some $s \geq 0$, it is calculated that

$$\text{pr}(S \geq s \mid H_0) \leq \text{pr}(\chi_q^2 > s) + \frac{2^{-q/2} e^{-s/2} s^{(q-1)/2}}{\Gamma(q/2)} \int_{\theta_l}^{\theta_u} E \left| \frac{\partial S(\theta)^{1/2}}{\partial \theta} \right| d\theta. \quad (8)$$

In particular, when $Q(\cdot)$ is the treatment indicator of $q = 1$, an approximate significance p -value given the observed $S = S_0$ can thus be calculated as $\text{pr}(\chi_1^2 > S_0) + 2^{-1/2} e^{-S_0/2} / \Gamma(1/2) \cdot M$, where M is the piecewise numerical difference approximation of $\int_{\theta_l}^{\theta_u} E |\partial \widetilde{S}(\theta)^{1/2} / \partial \theta| d\theta$, i.e., $M = \sum_{j=1}^J |\widetilde{S}(\theta_j)^{1/2} - \widetilde{S}(\theta_{j-1})^{1/2}|$. Here, $\theta_l = \theta_0 \leq \theta_1 \leq \dots \leq \theta_J = \theta_u$ are the turning points such that $\partial^2 \{S(\theta)^{1/2}\} / \partial \theta^2 = 0$. For a two-sided test, $S = \sup_{\theta \in [\theta_l, \theta_u]} \{|S(\theta)|\}$ can be used.

For the $\theta \in \Theta \subset \mathcal{R}^{p_2}$ of higher dimension, i.e., $p_2 > 1$, a natural extension of $S = \sup_{\theta \in \Theta} \{S(\theta)\}$ can be used, although its theoretical calculation of significance probability bounds is foreseeably complicated. In practice, we can instead adapt a computer-intensive resampling method by Parzen et al. (1994) for ease of calculation. Specifically, for a given set of n iid standard normal deviates, $\{\epsilon_1, \epsilon_2, \dots, \epsilon_n\}$, say, multiplying the individual terms in (7) as $\widetilde{\mathcal{E}}(\beta, \theta) = \sum_i \epsilon_i \int_0^\infty \Delta_i(t) \{\psi_i(t) - \bar{\psi}(t)\} dX_i(t)$, $n^{-1/2} \widetilde{\mathcal{E}}(\beta, \theta)$ is mean-zero with the variance of $\widehat{\Sigma}$ conditional on the observed data. Thus, $n^{-1/2} \widetilde{\mathcal{E}}(\beta, \theta)$ would have the same limiting distribution as $n^{-1/2} \mathcal{E}(\beta, \theta)$. As a result, we would generate a large number of sets of $\{\epsilon_i\}$'s to calculate the empirical distribution of $\widetilde{S} = \sup_{\theta \in \Theta} \widetilde{S}(\theta) = \sup_{\theta \in \Theta} \inf_{\beta_R} \{\widetilde{\mathcal{E}}(0, \beta_R, \theta) \widehat{\Sigma}^{-1}(\widehat{\beta}, \theta) \widetilde{\mathcal{E}}(0, \beta_R, \theta)\}$. Then an approximate significance p -value is the proportion of $\widetilde{S} \geq S_0$.

4 EXAMPLES

4.1 Distributions of bivariate action duration times

As proposed in §2.1, the bivariate action duration times can be modelled by the Gamma frailty model. To gain some concrete sense about this family of distributions and their ultimate impact on the mean response curve, we examine a few examples. One choice is to use the Weibull forms for $\lambda_U(\cdot)$ and $\lambda_W(\cdot)$, i.e., $\lambda_U \omega t^{\omega-1}$ and $\lambda_W \omega t^{\omega-1}$, respectively, where ω is a parameter. Thus, the bivariate density function and survival function of (U, W) becomes

$$f_{U,W}(u, w; \theta) = \alpha(1 + \alpha)\alpha^{-\alpha} \lambda_U \lambda_W u^\omega w^\omega (\alpha^{-1} + \lambda_U u^\omega + \lambda_W w^\omega)^{-\alpha-2},$$

and $S_{U,W}(u, w; \theta) = \{1 + (\lambda_U u^\omega + \lambda_W w^\omega)/\alpha\}^{-\alpha}$, where $\theta = (\alpha^T, \omega, \lambda_U, \lambda_W)^T$, respectively. This is the generalised Pareto power distribution, also called the bivariate Burr distribution. When $\omega = 1$, the marginal distributions of (U, W) become exponential. The bivariate density and survival functions when $\lambda_U = \lambda_W = \omega = 1$ and $\alpha = 0.5$ are demonstrated in Fig. 1.

[Fig. 1 about here]

The impact of the inclusion of varying effectiveness duration on the mean response model is reflected in the shape of the function of $H(t)$, as demonstrated in model (4). The function of $H(t)$ under the mentioned distributions of (U, W) are plotted in Fig. 2 under three $\lambda_W = 0.5, 1.0$ and 1.5 , which represent the relatively short/long length of the effectiveness durations. It is not surprising to see that all of the curves appear to be tied down toward $t = 0$ and $t = \infty$ with a peak in the middle. This mimics the phenomenon of effect erosion that is often observed, which is for the treatment to take effect gradually, reach peak efficacy and then dampen as time goes on. More interestingly, as λ_W increases, the time period of action onset becomes shorter, and the curves appear to have uniformly lower efficacy, i.e., $H(t; \lambda_W = 1.5) \leq H(t; \lambda_W = 1.0) \leq H(t; \lambda_W = 0.5)$.

[Fig. 2 about here]

4.2 Simulations

Moderate simulations are conducted mainly to demonstrate the validity of the estimation procedures. According to our models, there are three steps to simulating the data sets (1)

Varying effectiveness times (u_i, w_i) : these bivariate times are simulated following the Gamma frailty model. The ultimate density function used for the bivariate times is $0.75\sqrt{2}uw(u + w + 2)^{-2.5}$. (2) Observation times $(t_{i1}, t_{i2}, \dots, t_{i,m_i})$. The observation times are simulated according to a random effect Poisson process with intensity rate following Gamma $(1, 0.5)$. The total time period of observation following uniform distribution with mean of 20 yields about 11 observation times per subject. (3) Repeated responses $(y(t_{i1}), y(t_{i2}), \dots, y(t_{i,m_i}))$: the repeated responses are simulated according to the following model:

$$y_{ij}(t_{ij}) = \mu(t_{ij}) + \beta_Q QI(u_i < t_{ij} < u_i + w_i) + \beta_R R(t_{ij}) + \varepsilon(t_{ij}).$$

Here Q is the treatment indicator of a Bernoulli random variable with success probability of 50%, $R(t)$ are standard normal, $\varepsilon(t)$ is a Gaussian process with $\text{cov}\{\varepsilon(s), \varepsilon(t)\} = \exp(-|s-t|)$ and $\mu(t) = t^{1/2}$ and $\sin(2\pi t)$, respectively. The true values of (β_R, β_Q) are $(1, 0)$, $(-1, 0)$, $(1, 1)$ and $(-1, 1)$, respectively. The simulation results are summarised in Table 1. For each entry in the table, 1,000 replicates are simulated to estimate the bias and empirical coverage probability. The bias is defined as the difference between the sample mean of the estimates over the 1,000 replicated data sets and its true value. The empirical coverage probability is the percentage of Wald-type 95% confidence intervals that include the true parameters. It is evident that the estimators are virtually unbiased and the nominal confidence intervals carry reasonable coverages.

[Table 1 about here]

4.3 A real randomised clinical trial

Between November 1997 and April 1999, a landmark randomised trial, [HIVNET 012](#), was conducted to assess the safety and efficacy of single-dose nevirapine versus zidovudine for the prevention of mother-to-child HIV-1 vertical transmission among pregnant women in less-developed countries (http://www.hptn.org/research_studies/hivnet012.asp). In this trial, a total of 626 HIV-1 infected pregnant women in Uganda were recruited and randomly assigned to either nevirapine or zidovudine at more than 36 weeks' gestation. Complete medical histories and physical examinations of all participants were collected before their entry to the study, on enrolment, at delivery, at discharge from hospital, and at 7 days and 6 weeks after delivery. According to the trial protocol, the primary efficacy endpoint was the HIV-1 infection of neonates and HIV-1-free survival rates. By the end of this trial, the primary efficacy analysis showed that the HIV-1 transmission risk in the zidovudine and

nevirapine groups was 10.4% and 8.2% at birth ($p > 0.3$); 21.3% and 11.9% by age 6-8 weeks ($p < 0.003$); and 25.1% and 13.1% by age 14-16 weeks ($p < 0.001$), respectively, which suggests that single-dose nevirapine could significantly lower the vertical HIV-1 transmission risk in less-developed countries significantly (Guay et al., 1999).

One of the important secondary objectives of this trial specified in the protocol is to determine the relationship of maternal plasma RNA levels at the delivery with the rate of perinatal transmission. Therefore, it is critical to first focus on how the administration of nevirapine would affect the maternal viral load. In fact, quantitative plasma HIV-1 RNA measurements were taken repeatedly before entry, at delivery, and at 7 days and 6 weeks after delivery, respectively. These measurements can be used to compare the immunologic disease progression over time.

Unlike the control drug, zidovudine, nevirapine is a non-nucleoside benzodiazepine derivative. It does not have to be taken in by the human cell and metabolized to its active form and, thus, it generally takes effect against intracellular and extracellular virus immediately after administration (Zhang et al., 1996). However, the administration of nevirapine in labor is not soon enough to cause a decrease in viral load in labor, and it is difficult to determine precisely at what time the nevirapine takes the immediate effect. On the other hand, the nevirapine was administered only once in order to maintain certain plasma drug concentration for up to seven days. Preliminary studies showed that the potent antiviral effect usually persists for one to two weeks, followed by a rapid development of viral resistance to nevirapine and an increase in plasma virus. By the fourth week after drug administration, usually all of the subjects would develop viral resistance. In Fig. 3, the maternal plasma viral loads are plotted for both treatment arms. As shown in the figure, their mean functions do not vary dramatically at immediate administration or in the relative long-run after the fourth week. More smoothed lowess curves also show similar patterns.

[Fig. 3 about here]

For demonstration purposes, we let $Q(t)$ be the treatment indicator of nevirapine versus zidovudine and $R(t)$ be the number of days since first dose of drug administration and maternal age, respectively, in the proposed model (3). The estimates of regression parameters and their estimated standard errors are tabulated in Table (2). In contrast, this table also shows the usual generalized estimating equation estimates without considering the varying effectiveness durations. As shown in the table, the use of nevirapine significantly lowers

the HIV RNA viral load in both methods. Specifically, when the varying durations were ignored, the generalised estimating equation estimates of HIV viral load reduction by the nevirapine was merely 0.0434 in \log_{10} , i.e., $10^{0.434} = 2.7$ times that of the zidovudine. The subject-specific reduction, however, shows a much larger magnitude of 1.379 in \log_{10} , i.e., $10^{1.379} = 23.9$ times when the varying treatment effectiveness durations are included in the mean response model. In fact, the latter estimate is quite consistent with the estimated reduction of 1.3 in \log_{10} that was reported in a phase I/II pharmacokinetic study of nevirapine (Musoke et al., 1999).

[Table 2 about here]

5 EXTENSIONS

In this section, we discuss some potential extensions of the proposed methods and the associated estimation procedures. To simplify our presentation, we assume that $\beta_Q \neq 0$. As shown earlier, when $\beta_Q \neq 0$, all the discussed extensions can accommodate the hypothesis testing of treatment efficacy by considering the test statistics similar to S .

5.1 Multiplicative mean response models

In addition to the additive model in (1), there is also a parallel multiplicative model proposed in the literature (Cheng & Wei, 2000),

$$E\{Y_i(t) \mid Z_i(s); 0 \leq s \leq t\} = \mu(t) \exp\{\beta^T Z_i(t)\}, \quad (9)$$

to analyze the repeated measurements. The mean structure of this model is equivalent to those of the additive model when the response curves are properly transformed, for instance, if the $Y(t)$ in model (9) is log-transformed. Nevertheless, the multiplicative model also assumes constant treatment effect and may not be appropriate in the presence of varying effectiveness durations, either. To include varying effectiveness durations, we propose the following model:

$$E\{Y_i(t) \mid Z_i(s), U_i, W_i; 0 \leq s \leq t\} = \mu(t) \exp\{\beta_Q^T Q_i(t) I(U_i \leq t \leq U_i + W_i) + \beta_R^T R_i(t)\}. \quad (10)$$

The marginalised version of this model is thus

$$E\{Y_i(t) \mid Z_i(s); 0 \leq s \leq t\} = \mu(t) \exp\{\beta_R^T R_i(t)\} [\exp\{\beta_Q^T Q_i(t)\} H(t; \theta) + \{1 - H(t; \theta)\}]. \quad (11)$$

Apparently, $\exp\{\beta_Q^T Q_i(t)\}H(t) + \{1 - H(t)\}$, which is the weighted average of $\exp\{\beta_Q^T Q_i(t)\}$ and 1, would approach 1 as t goes to 0 or ∞ . When $Q_i(\cdot)$ is the treatment indicator, this property reflects the erosion phenomenon in the observed response curves. Unlike the additive model (4), however, the marginalised multiplicative model does not maintain the linear structure on β_Q , which may add complexity in estimation.

To estimate the parameters (β, θ) in model (10), consider $X_i(t) = \int_0^t Y_i(s)dN_i(s)$. Since

$$E\{dX_i(t)|Z_i(s), C_i; 0 \leq s \leq t, \beta_*, \theta_*\} = \Delta_i(t) \exp\{\rho_i(t; \beta_*, \theta_*)\}d\Omega_\mu(t),$$

where $\rho_i(t; \beta, \theta) = \beta_R^T R_i(t) + \log[\exp\{\beta_Q^T Q_i(t)\}H(t; \theta) + \{1 - H(t; \theta)\}]$, then $M_{\rho,i}(t) = X_i(t) - \int_0^t \Delta_i(s) \exp\{\rho_i(s)\}d\Omega_\mu(s)$ are the zero-mean stochastic processes. The following estimating equations can thus be used

$$\mathcal{E}_\rho(\beta, \theta) = \sum_{i=1}^n \int_0^\tau \Delta_i(t) \{\psi_i(t) - \bar{\psi}_\rho(t; \beta, \theta)\} dX_i(t) = 0,$$

where $\bar{\psi}_\rho(t; \beta, \theta) = \sum_{i=1}^n \Delta_i(t) \exp\{\rho_i(t; \beta, \theta)\}\psi_i(t) / \sum_{i=1}^n \Delta_i(t) \exp\{\rho_i(t; \beta, \theta)\}$. Again, denote $(\hat{\beta}, \hat{\theta})$ as the solutions to $\mathcal{E}_\rho(\beta, \theta) = 0$. Then the similar techniques proposed earlier lead to the large-sample properties of consistency as well as asymptotic normality, $n^{1/2}(\hat{\beta}^T - \beta_*^T, \hat{\theta}^T - \theta_*^T)^T \rightarrow N(0, D_\rho^{-1} \Sigma_\rho D_\rho^{-1})$ in distribution, where $D_\rho = -E \int_0^\tau \Delta_1(t) \exp\{\rho_1(t)\} \bar{\psi}'_\rho(t) d\Omega_\mu(t)$ and $\Sigma_\rho = E[\int_0^\tau \Delta_1(t) \{\psi_i(t) - \bar{\psi}_\rho(t; \beta, \theta)\} dM_{\rho,1}(t)]^{\otimes 2}$, respectively. Here, D_ρ and Σ_ρ can be estimated by their empirical counterparts respectively.

5.2 Covariate-dependent observation times

Usually in a well-designed randomised clinical trial, repeated measurements are supposed to be collected at a pre-determined or fixed set of time points to avoid potential bias or missing values in the data set. In reality, however, measurements may be actually observed at varying sets of time points for different individuals, and may be further affected by the subjects' covariates (Sun & Wei, 2000; Lin & Ying, 2001). In the statistical literature, when the mean functions of the counting processes are different, the following model is usually used,

$$E\{N_i(t) | Z_i(s); 0 \leq s \leq t\} = \eta(t) \exp\{\kappa^T Z_i(t)\}, \quad (12)$$

where κ is a parameter and $\eta(\cdot)$ is an unspecified baseline function, as in Pepe & Cai (1993) and Lawless & Nadeau (1995). Hence, the following estimating equations can be used to

estimate κ ,

$$\mathcal{E}_N(\kappa) = \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t; \kappa)\} dN_i(t) = 0,$$

where $\bar{Z}_i(t; \kappa) = \sum_{i=1}^n \Delta_i(t) \exp\{\kappa^\top Z_i(t)\} Z_i(t) / \sum_{i=1}^n \Delta_i(t) \exp\{\kappa^\top Z_i(t)\}$, by differentiating the log of partial likelihood function of $\sum_i \int_0^\tau \{\kappa^\top Z_i(t) - \log[\sum_k \Delta_k(t) \exp\{\kappa^\top Z_i(t)\}]\} dN_i(t)$ with respect to κ . Let $d\Omega_{\mu, \eta}(t) = \eta(t) d\Omega_\mu(t)$. Since $E\{dX_i(t) \mid Z_i(s), C_i; 0 \leq s \leq t, \beta_*, \theta_*, \kappa_*\} = \Delta_i(t) \exp\{\kappa_*^\top Z_i(t)\} d\Omega_{\mu, \eta}(t)$, the following estimating equations can be similarly established for $(\beta^\top, \theta^\top, \kappa^\top)^\top$ as in (7),

$$\mathcal{E}_1(\beta, \theta, \kappa) = \sum_{i=1}^n \int_0^\tau \Delta_i(t) \{\psi_i(t) - \bar{\psi}_1(t)\} dX_i(t) = 0,$$

where $\bar{\psi}_1(t) = \sum_{i=1}^n \Delta_i(t) \exp\{\kappa^\top Z_i(t)\} \psi_i(t) / \sum_{i=1}^n \Delta_i(t) \exp\{\kappa^\top Z_i(t)\}$. Denote $(\hat{\beta}, \hat{\theta}, \hat{\kappa})$ as the solutions such that $\mathcal{E}_N(\hat{\kappa}) = \mathcal{E}_1(\hat{\beta}, \hat{\theta}, \hat{\kappa}) = 0$. Then following the arguments in Sun & Wei (2000), it is true that they are consist and have the asymptotic normality as,

$$n^{1/2} \begin{pmatrix} \hat{\beta} - \beta_* \\ \hat{\theta} - \theta_* \end{pmatrix} \simeq n^{1/2} (D_1^{-1}(\beta_*, \theta_*, \kappa_*), -D_N^{-1}(\kappa_*)) \begin{pmatrix} \mathcal{E}_1(\beta_*, \theta_*, \kappa_*) \\ \mathcal{E}_N(\kappa_*) \end{pmatrix}, \quad (13)$$

where $D_1 = -\lim_{n \rightarrow \infty} n^{-1} \mathcal{E}'_1$ and $D_N = -\lim_{n \rightarrow \infty} n^{-1} \mathcal{E}'_N$. Hence, by the normal approximation of $(\mathcal{E}_1, \mathcal{E}_N)$ as shown the Appendix, the asymptotic variance of (13) can be estimated by $(\hat{D}_1^{-1}, -\hat{D}_N^{-1}) \hat{\Sigma}_1 (\hat{D}_1^{-1}, -\hat{D}_N^{-1})^\top$, where \hat{D}_N , \hat{D}_1 and $\hat{\Sigma}_1$ are their respective empirical estimates.

5.3 Isotonic regression of mean response models

In either the additive model (1) or the multiplicative model (9), the mean of the baseline response curves is assumed to be arbitrarily unspecified. In some randomised trials, for instance, the subjects' baseline mean response curves may have special monotonic shapes. They can be either monotonically increasing or decreasing, depending upon the underlying measurement progression being persistently ameliorating or deteriorating. Therefore the marginal model (4) is extended as in

$$E\{Y_i(t) \mid Z_i(s); 0 \leq s \leq t\} = \mu(t) + \beta_Q^\top Q_i(t) H(t; \theta) + \beta_R^\top R_i(t), \quad (14)$$

with $\mu(\cdot) \in \mathcal{M}$, where \mathcal{M} is the set of all the monotonically nonincreasing functions, or monotonically nondecreasing functions, denoted by \mathcal{M}^- or \mathcal{M}^+ , respectively. Without loss of generality, we focus on \mathcal{M}^- in this section.

When there is no covariate information included in (14), the regression model reduces to a simple isotonic estimation problem. That is, we need to find $\mu(\cdot) \in \mathcal{M}^-$ such that

$$\mu^-(\cdot) = \arg \min_{\mu \in \mathcal{M}^-} \sum_{i=1}^n \|Y_i - \mu\|^2,$$

with the norm $\|\cdot\|$ defined as in Rice & Silverman (1991). Thus the computational algorithms, such as the most widely used Pooled Adjacent Violators Algorithm or the Minimum Lower Set Algorithm, can be used (Robertson et al., 1988). When the covariate information is included as proposed in the model, we can adapt the back-fitting algorithm as in Zeger & Diggle (1994) to obtain the final estimates of the baseline function $\mu(\cdot)$ and the parameters.

Algorithm.

1. Consider $(\widehat{\beta}_{[k]}, \widehat{\theta}_{[k]})$ are obtained in the k th iterative step, $k = 1, 2, \dots$, where $\widehat{\beta}_{[0]} = \widehat{\theta}_{[0]} = 0$. Use one of the aforementioned algorithms to compute $\mu_{[k+1]}^-(\cdot) \in \mathcal{M}$ such that

$$\mu_{[k+1]}^-(\cdot) = \arg \min_{\mu \in \mathcal{M}^-} \sum_{i=1}^n \|Y_{[k],i} - \mu\|^2,$$

where $Y_{[k],i}(t) = Y_i(t) - \{\beta_{Q,[k]}^T Q_i(t)H(t; \theta_{[k]}) + \beta_{R,[k]}^T R_i(t)\}$;

2. Given $\mu_{[k+1]}^-(\cdot)$, obtain $(\widehat{\beta}_{[k+1]}, \widehat{\theta}_{[k+1]})$ by minimizing

$$\sum_{i=1}^n \int_0^\tau \psi_i(t) [Y_i(t) - \{\mu_{[k+1]}^-(t) + \beta_Q^T Q_i(t)H(t; \theta) + \beta_R^T R_i(t)\}]^2 dN_i(t).$$

In fact, the proposed isotonic regression model belongs to a more general additive isotonic model (Bacchetti, 1989),

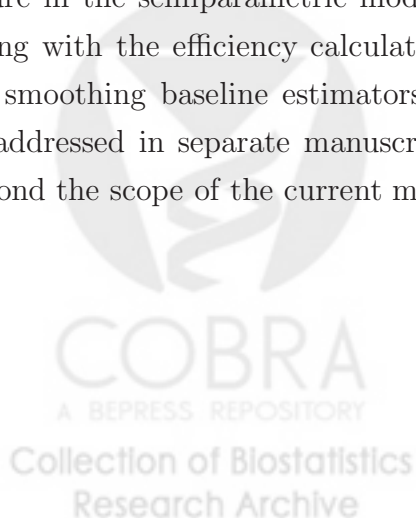
$$E\{Y_i(t) \mid Z_i(s); 0 \leq s \leq t\} = \sum_{l=1}^P \mu_l(t) + \beta_Q^T Q_i(t)H(t; \theta) + \beta_R^T R_i(t),$$

where $(\mu_1, \mu_2, \dots, \mu_P)$ are the P -dimensional isotonic function vectors. When there is no covariate information involved, the backfitting algorithm by Hastie & Tibshirani (1990) can be used with the Pooled Adjacent Violators Algorithm to individual μ_i iteratively. When the covariate information is included, it is straightforward to further extend the above algorithm for the estimation in this model. To avoid complicated variance calculation of the estimators, computer-intensive methods, such as the bootstrap method (Efron & Tibshirani, 1994), can be used.

6 DISCUSSION

As shown earlier, the marginalised model (4) is in fact a type of time-varying coefficient model. Mathematically, the introduction of varying treatment effectiveness duration is equivalent to most of the smoothing techniques applied in the nonparametric estimation approaches: the distributional assumption on the unobserved varying effectiveness durations and its marginalisation essentially smoothes the differences in the mean responses. Compared with a flexible nonparametric time-varying coefficient, the proposed model may offer more specific interpretation of the parameters. Like the time-varying coefficient model, the proposed model also serves as a post hoc tool to analyze repeated measurements, although its estimation on EU and EW may serve future designs of clinical trials to account for such potential treatment effectiveness lag and saturation times. Nevertheless, the proposed methods may be a useful secondary analysis tool to explore and study treatment efficacy when treatment effectiveness durations potentially exist.

The estimation approaches in this article follow the counting processes formulation in analysis of repeated measurements by [Cheng & Wei \(2000\)](#), [Sun & Wei \(2000\)](#), [Lin & Ying \(2001\)](#) and others. This formulation is simple and does not require complicated smoothing, with room for significant improvement in efficiency. In [Lin & Ying \(2001\)](#), an estimate of the baseline function was introduced to improve the efficiency by minimizing the variance of the proposed estimating equations. It is still ad hoc and unknown whether or not the efficiency reaches the semiparametric efficiency bound. Although there are other approaches that do not need smoothing and may have more efficient estimation, for instance, the difference-based method by [Yatchew \(1997\)](#) for the partial linear models with less loss of efficiency, more future in the semiparametric model efficiency framework of [Bickel et al. \(1993\)](#) is needed. Along with the efficiency calculation, the technical development of asymptotic theory for the smoothing baseline estimators in §2 and the isotonic regression algorithms in §3 will be addressed in separate manuscripts, given the interest in development of these theories beyond the scope of the current manuscript.



APPENDIX A: ASYMPTOTICS

A.1. Weak Convergence of $n^{-1/2}\mathcal{E}(\cdot; \beta_*, \theta_*)$

Our proof follows an extension of Appendix 2 in [Cheng & Wei \(2000\)](#). Denote $\mathcal{B}(t) = \sum_{i=1}^n \int_0^t \Delta_i(s)\Phi(s)dM_i(s)$ and $\mathcal{B}_\varphi(t) = \sum_{i=1}^n \int_0^t \Delta_i(s)\Phi(s)\varphi(s)dM_i(s)$. Then $\mathcal{E}(\beta_*, \theta_*) = \mathcal{B}_\varphi(\tau) - \int_0^\tau \bar{\varphi}(t)d\mathcal{B}(t)$. For any $t > 0$, $\mathcal{B}(t)$ and $\mathcal{B}_\varphi(t)$ are the sums of independent and identically distributed zero-mean terms. By the Central Limit Theorem, $n^{-1/2}(\mathcal{B}(t), \mathcal{B}_\varphi(t))$ converges in distribution to a zero-mean Gaussian process, $(\mathcal{W}(t), \mathcal{W}_\varphi(t))$, say.

Assume that $\varphi_i(\cdot)$, $i = 1, 2, \dots, n$, are of bounded variation. Moreover, without loss of generality, $\varphi_i(\cdot)$ are assumed to be non-negative. Then the individual terms of $\mathcal{B}(\cdot)$ and $\mathcal{B}_\varphi(\cdot)$ can be written as sums of monotone functions in t and hence “manageable.” Thus $n^{-1/2}(\mathcal{B}(t), \mathcal{B}_\varphi(t))$ converges weakly to $(\mathcal{W}, \mathcal{W}_\varphi)$, as $n \rightarrow \infty$ ([Pollard, 1990](#), p. 38 and p.53). By the strong embedding theorem in [Shorack & Wellner \(1986, p. 47\)](#), there exists an induced probability space such that $(n^{-1/2}\mathcal{B}(t), n^{-1/2}\mathcal{B}_\varphi(t), n^{-1} \sum_{i=1}^n \Delta_i(t), n^{-1} \sum_{i=1}^n \Delta_i(t)\varphi_i(t))$ converges almost surely. By the Lemma 8.2.3 in [Chow & Teicher \(1988, p.265\)](#) coupled with Helly’s theorem in [Serfling \(1980, p.352\)](#), it is true that

$$n^{-1/2} \int_0^t \frac{n}{\sum_{i=1}^n \Delta_i(s)} d\mathcal{B}(s) \rightarrow \int_0^t \frac{1}{E\Delta_1(s)} d\mathcal{W}(s) \text{ and } n^{-1/2} \int_0^t \bar{\varphi}(s) d\mathcal{B}(s) \rightarrow \int_0^t \bar{\varphi}_*(s) d\mathcal{W}(s)$$

almost surely and uniformly in t . The weak convergence of $n^{-1/2}\mathcal{E}(\beta_*, \theta_*)$ thus follows in the original probability space, due to their convergence almost surely to $\mathcal{W}_\varphi(\tau) - \int_0^\tau \bar{\varphi}_*(s)d\mathcal{W}(s)$ in the induced probability. The calculation of the variance-covariance matrix of Σ is straightforward.

A.2. Asymptotic variance of $n^{-1/2}(\mathcal{E}_1(\beta_*, \theta_*, \kappa_*)^T, \mathcal{E}_N(\kappa_*)^T)^T$

The asymptotic normality of the joint distribution of $n^{-1/2}(\mathcal{E}_1(\beta_*, \theta_*, \kappa_*)^T, \mathcal{E}_N(\kappa_*)^T)^T$ can be similarly established following the arguments in [Lin & Wei \(1989\)](#) and [Sun & Wei \(2000\)](#). To calculate its associated asymptotic variance, it is noted that

$$n^{-1/2}\mathcal{E}_N(\kappa_*) = n^{-1/2} \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t; \kappa)\} dM_{N,i}(t),$$

where $M_{N,i}(t) = N_i(t) - \int_0^t \Delta_i(s) \exp\{\kappa^T Z_i(s)\} d\Omega(s)$. Let $e_i = E \int_0^\tau \{Z_i(t) - \bar{Z}(t; \kappa)\} dM_{N,i}(t)$ and its empirical estimates as \hat{e}_i , respectively. Thus the variance-covariance matrix of

$n^{-1/2}(\mathcal{E}_1(\beta_*, \theta_*, \kappa_*)^\top, \mathcal{E}_N(\kappa_*)^\top)^\top$ can be approximated by

$$\widehat{\Sigma}_1 = \begin{pmatrix} n^{-1} \sum_{i=1}^n \widehat{\epsilon}_i \widehat{\epsilon}_i^\top & n^{-1} \sum_{i=1}^n \widehat{\epsilon}_i \widehat{e}_i^\top \\ n^{-1} \sum_{i=1}^n \widehat{e}_i \widehat{\epsilon}_i^\top & n^{-1} \sum_{i=1}^n \widehat{e}_i \widehat{e}_i^\top \end{pmatrix},$$

where $\widehat{\epsilon}_i = \int_0^\tau \Delta_i(t) \Phi(t) \{\varphi_i(t) - \bar{\varphi}_1(t)\} \exp\{\widehat{\kappa}^\top Z_i(t)\} d\widehat{\Omega}_{\mu, \eta}(t)$ and

$$\widehat{\Omega}_{\mu, \eta}(t) = \int_0^t \frac{\sum_{i=1}^n dX_i(s)}{\sum_{i=1}^n \Delta_i(s) \exp\{\widehat{\kappa}^\top Z_i(s)\}}.$$

REFERENCES

- BACCHETTI, P. (1989). Additive isotonic models. *J. Amer. Statist. Asso.* **84**, 289-294.
- BICKEL, P. J., KLASSESEN, C. A. J., RITOV, Y. & WELLNER (1993). *Efficiency and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press.
- CAPRA, W. B. & MÜLLER, H. G. (1997). An accelerated time model for response curves. *J. Amer. Statist. Asso.* **92**, 72-83.
- CHEN, Y. Q. & JEWELL, N. P. (2001). On a general class of hazards regression models. *Biometrika* **88**, 687-702.
- CHEN, Y. Q., ROHDE, C. A. & WANG, M.-C. (2002). Additive hazards models with latent treatment effectiveness lag time. *Biometrika* **89**, 917-931.
- CHENG, S. C. & L. J. WEI (2000). Inferences for a semiparametric model with panel data. *Biometrika* **87**, 89-97.
- CHIANG, C. T., RICE, J. A. & WU, C. O. (2001). Smoothing spline estimation for varying coefficients models with repeatedly measured dependent variables. *J. Amer. Statist. Asso.* **96**, 605-619.
- CHOW, Y. S. & TEICHER, H. (1988). *Probability Theory: Independence, Interchangeability, Martingales, 2nd Ed.* New York: Springer.
- CLAYTON, D. G. (1978). A model for association in bivariate life tables and its application in epidemiologic studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141-151.

- DAVIES, R. B (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64**, 247-254.
- DAVIES, R. B (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74**, 33-43.
- EFRON, B. & TIBSHIRANI, R. J. (1994). *An Introduction to Bootstrapping*. New York: Chapman & Hall.
- ESHLEMAN, S. H., MRACNA, M., GUAY, L. A., DESEYVE, M., CUNNINGHAM, S., MIROCHNICK, M., MUSOKE, P., FLEMING, T., FOWLER, M. G., MOFENSON, L. M., MMIRO, F. & JACKSON, J. B. (2001). Selection and fading of resistance mutations in women and infants receiving nevirapine to prevent HIV-1 vertical transmission (HIVNET 012). *AIDS* **15**, 1951-1957.
- GUAY, L. A., MUSOKE, P., FLEMING, T., BAGENDA, D., ALLEN, M., NAKABIITO, C., SHERMAN, J., BAKAKI, P., DUCAR, C., DESEYVE, M., EMEL, L., MIROCHNICK, M., FOWLER, M. G., MOFENSON, L., MIOTTI, P., DRANSFIELD, K., BRAY, D., MMIRO, F. & JACKSON, J. B. (1999). Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of HIV-1 in Kampala, Uganda: HIV 012 randomised trial. *Lancet* **354**, 705-802.
- HASTIE, T. & TIBSHIRANI, R. (1990). *Generalized Additive Models*. London: Chapman & Hall.
- HALPERIN, M., ROGOT, E., GURIAN, J. & EDERER, F. (1968). Sample sizes for medical trials with special reference to long-term therapy. *J. Chronic Dis.* **21**, 13-24.
- HOOVER, D. R., RICE, J. A., WU, C. O. & YANG, L. P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85**, 809-822.
- HUANG, J. H., WU, C. O. & ZHOU, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89**, 111-128.
- KOMAROVA, N. L. & WODARZ, D. (2005). Drug resistance in cancer: principles of emergence and prevention. *Proc. Nat. Acad. Sci.* **102**, 9714-9719.

- LAWLESS, J. F. & NADEAU, C. (1995). Some simple robust methods for the analysis of recurrent event data. *Technometrics* **37**, 158-168.
- LIN, D. Y. & WEI, L. J. (1989). The robust inference for the Cox proportional hazards model. *J. Amer. Statist. Asso.* **84**, 1074-1078.
- LIN, D. Y. & YING, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data (with discussion). *J. Amer. Statist. Asso.* **96**, 103-126.
- LUO, X., TURNBULL, B. W. & CLARK, L. C. (1997). Likelihood ratio tests for a changepoint with survival data. *Biometrika* **84**, 555-565.
- MUSOKE, P., GUAY, L. A., BAGENDA, D., MIROCHNICK, M., NAKABIITO, C., FLEMING, T., ELLIOTT, T., HORTON, S., DRANSFIELD, K., PAV, J. W., MURARKA, A., ALLEN, M., FOWLER, M. G., MOFENSON, L., HOM, D., MMIRO, F. & JACKSON, J. B. (1999). A phase I/II study of the safety and pharmacokinetics of nevirapine in HIV-1-infected pregnant Uganda women and their neonates (HIVNET 006). *AIDS* **13**, 479-486.
- OAKES, D. R. (1989). Bivariate survival models induced by frailties. *J. Amer. Statist. Asso.* **84**, 487-493.
- O'QUIGLEY, J. & NATARAJIAN (2005). Erosion of regression effect in a survival study. *Biometrics* **60**, 344-351.
- POLLARD, D. (1990). *Empirical Processes: Theory and Applications*. Hayward, California: Institute of Mathematical Statistics.
- PEPE, M. S. & CAI, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *J. Amer. Statist. Asso.* **88**, 811-820.
- RICE, J. A. & SILVERMAN, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Royal Statist. Soc. Ser. B* **53**, 233-243.
- ROBERTSON, T., WRIGHT, F. T. & DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. New York: Wiley.
- SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

- SHORACK, G. R. & WELLNER, J. A. (1986). *Empirical Process with Applications to Statistics*. New York: Wiley.
- SUN, J. & WEI, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *J. Royal Statist. Soc. Ser. B* **62**, 293-302.
- WEI, L. J., YING, Z. & LIN, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845-51.
- WU, H., HUANG, Y., ACOSTA, E. P., ROSENKRANZ, S. L., KURITZKES, D. R., ERON, J. J., PERELSON, A. S. & GERBER, J. G. (2005). Modeling long-term HIV dynamics and antiretroviral response: effects of drug potency, pharmacokinetics, adherence and drug resistance. *J. AIDS* **39**, 272-283.
- WU, W. B., WOODROOFE, M. & MENTZ, G. (2001). Isotonic regression: another look at the changepoint problem. *Biometrika* **88**, 793-804.
- YATCHEW, A. (1997). An elementary estimator of the partial linear model. *Economics Letters* **57**, 135-143.
- ZEGER, S. L. & DIGGLE, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* **50**, 689-699.
- ZHANG, H., DORNADULA, G., WU, Y., HAVLIR, D., RICHMAN, D. D. & POMERANTZ, R. J. (1996). Kinetic analysis of intravirion reverse transcription in the blood plasma of human immunodeficiency virus type 1-infected individuals: direct assessment of resistance to reverse transcriptase inhibitors in vivo. *J. Virol.* **70**, 628-634.
- ZUCKER, D. M. & LAKATOS, E. (1990). Weighted log rank type statistics for comparing survival curves when there is a time lag in effectiveness of treatment. *Biometrika* **77**, 853-864.

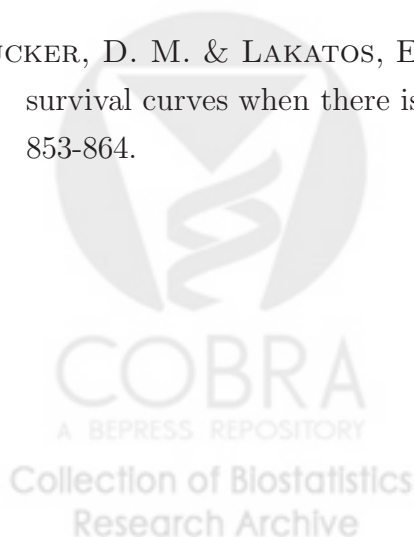


Fig. 1: Density function and survival function of the bivariate action onset

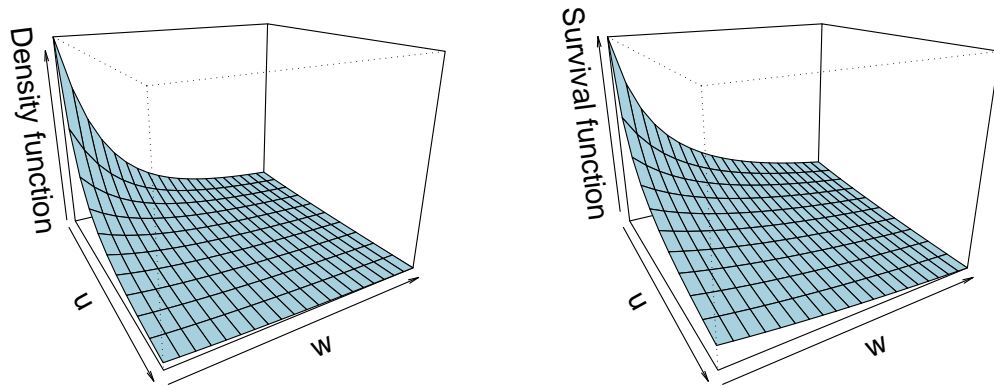


Fig. 2: The functions of $H(t)$ in model (4)

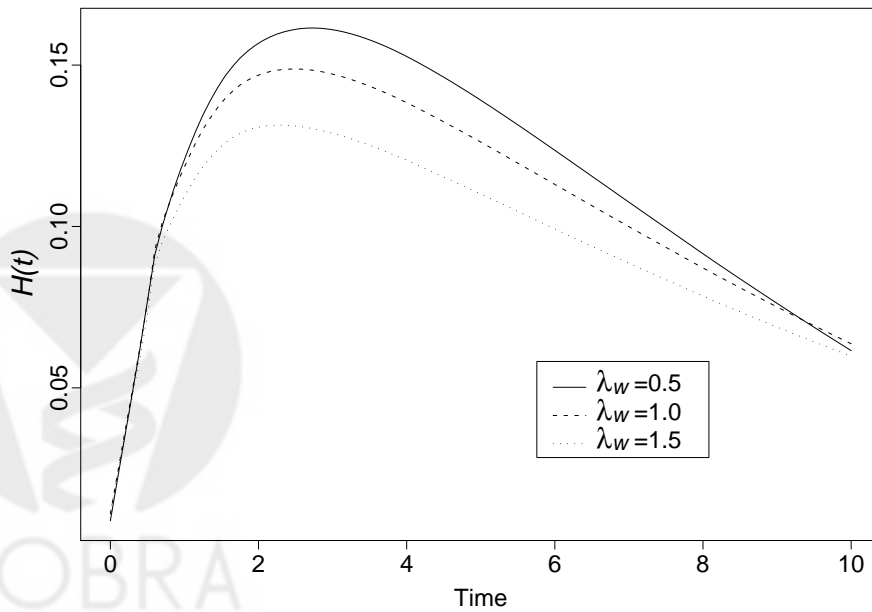


Fig. 3: Maternal HIV plasma viral load and their mean functions in HIVNET 012 trial: smoothed lines are loess curves

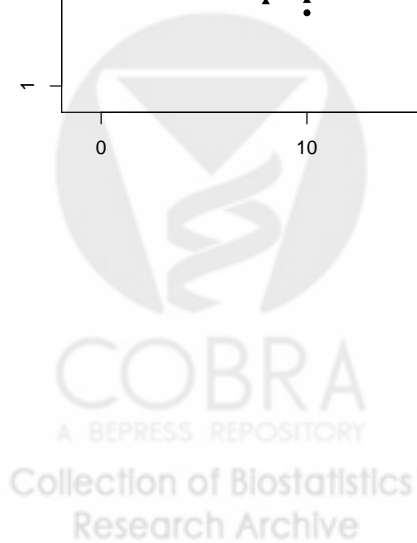
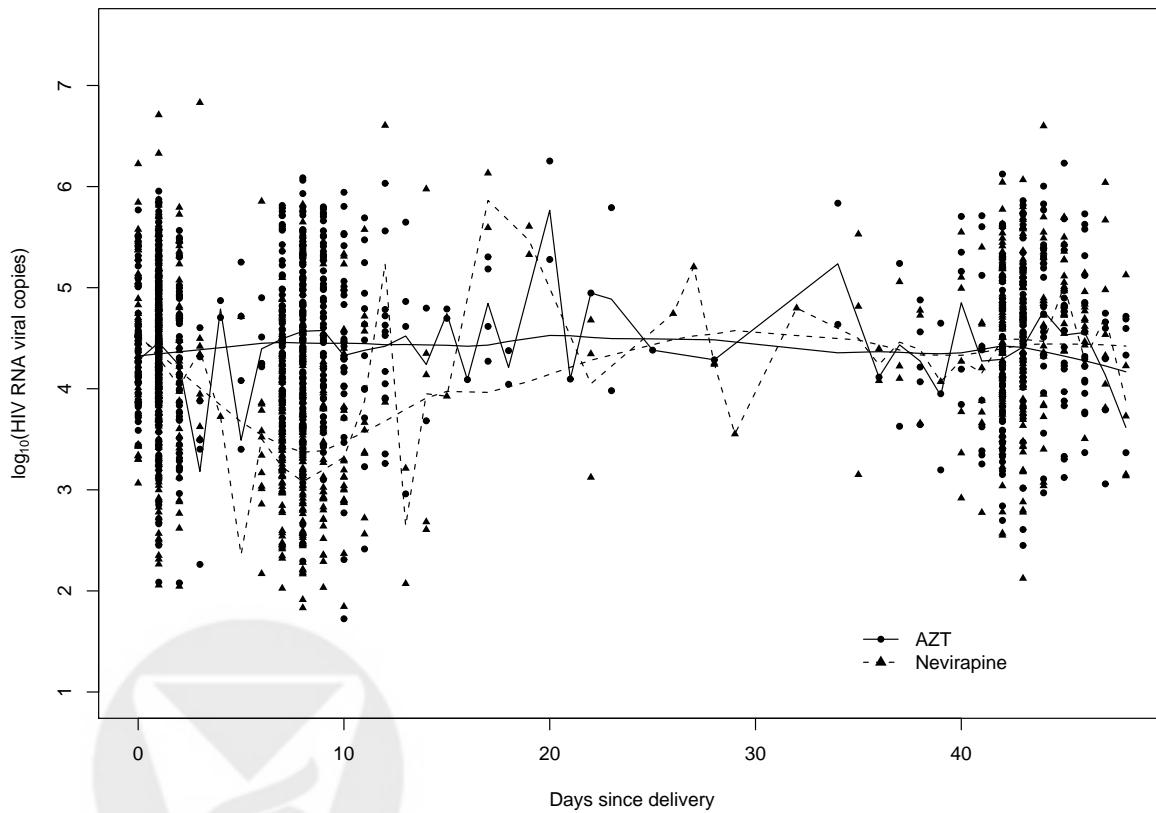


Table 1: Summary of simulation results. Each entry is the estimated bias with 95% empirical coverage probabilities in brackets

n	$\mu(t)$	$(\beta_{Q^*}, \beta_{R^*}) = (1, 0)$		$(\beta_{Q^*}, \beta_{R^*}) = (-1, 0)$		$(\beta_{Q^*}, \beta_{R^*}) = (1, 1)$		$(\beta_{Q^*}, \beta_{R^*}) = (-1, 1)$	
		β_Q	β_R	β_Q	β_R	β_Q	β_R	β_Q	β_R
50	\sqrt{t}	0.003	-0.005	0.011	0.008	-0.008	0.015	-0.017	0.013
		(0.944)	(0.971)	(0.960)	(0.930)	(0.938)	(0.951)	(0.959)	(0.961)
50	$\sin(2\pi t)$	-0.004	-0.003	0.010	-0.002	0.014	0.001	-0.003	-0.007
		(0.929)	(0.938)	(0.949)	(0.941)	(0.937)	(0.939)	(0.967)	(0.935)
100	\sqrt{t}	0.002	-0.003	-0.005	0.004	-0.006	0.011	0.013	0.002
		(0.966)	(0.958)	(0.971)	(0.957)	(0.919)	(0.947)	(0.944)	(0.958)
100	$\sin(2\pi t)$	0.014	0.023	-0.004	0.004	0.007	-0.002	0.006	-0.009
		(0.952)	(0.951)	(0.955)	(0.952)	(0.947)	(0.961)	(0.977)	(0.936)
200	\sqrt{t}	0.011	-0.009	-0.006	-0.016	0.005	0.006	0.006	0.004
		(0.968)	(0.969)	(0.931)	(0.958)	(0.955)	(0.976)	(0.975)	(0.966)
200	$\sin(2\pi t)$	0.006	0.002	-0.016	-0.014	0.012	-0.004	0.006	0.005
		(0.972)	(0.941)	(0.971)	(0.973)	(0.936)	(0.958)	(0.953)	(0.968)

Table 2: Parameter estimates in model (4) with/without varying effectiveness onset considered: Cov., covariates; Est., parameter estimates; s.e., standard errors; CI, confidence interval.

Cov.	Without varying onset			With varying onset		
	Est.	s.e.	95% CI	Estimate	s.e.	95% CI
Nevirapine vs. Zidovudine	-0.434	0.070	(-0.572,-0.296)	-1.379	0.083	(-1.542,-1.216)
Days since first dose	0.008	0.001	(0.006,0.009)	0.012	0.056	(-0.098,0.122)
Age	0.009	0.008	(-0.070,0.025)	-0.001	0.009	(-0.019,0.017)

