A Varying-Coefficient Model for the Evaluation of Time-Varying Concomitant Intervention Effects in Longitudinal Studies

Colin O. Wu*, Xin Tian

Office of Biostatistics Research

National Heart, Lung and Blood Institute, Bethesda, MD 20892

and

Heejung Bang

Division of Biostatistics and Epidemiology

Department of Public Health

Weill Medical College of Cornell University, NY 10021

*email: wuc@nhlbi.nih.gov, *Phone*: 301-435-0440, *Fax*: 301-480-1862

Summary

Concomitant interventions are often introduced during a longitudinal clinical trial to patients who respond undesirably to the pre-specified treatments. In addition to the main objective of evaluating the pre-specified treatment effects, an important secondary objective in such a trial is to evaluate whether a concomitant intervention could change a patient's response over time. Because the initiation of a concomitant intervention may depend on the patient's general trend of pre-intervention outcomes, regression approaches that treat the presence of the intervention as a time-dependent covariate may lead to biased estimates for the intervention effects. Borrowing the techniques of Follmann and Wu (1995) for modeling informative missing data, we propose a varying-coefficient mixed-effects model to evaluate the patient's longitudinal outcome trends before and after the patient's starting time of the intervention. By allowing the random coefficients to be correlated with the patient's starting time of the intervention, our model leads to less biased estimates of the intervention effects. Nonparametric estimation and inferences of the coefficient curves and intervention effects are developed using B-splines. Our methods are demonstrated through a longitudinal clinical trial in depression and heart disease and a simulation study.

Key words: Change-Point Models; Concomitant Intervention; Longitudinal Study; Polynomial Splines; Shared Parameter Model; Varying-Coefficient Model.

1 Introduction

A main objective of longitudinal analysis in clinical trials is to evaluate the effects of covariates of interest on the time-trends of the outcome variables. The treatment effects are usually modeled using a time-invariant categorical covariate, while the other covariates can be either time-invariant or time-dependent. Recent advances in longitudinal analysis have led to a wide range of regression methods. For parametric models, estimation and inference procedures, such as maximum likelihood estimation, the restricted maximum likelihood (REML) estimation and the generalized estimating equations (GEEs), can be found [1-4]. For models involving nonparametric components, smoothing methods, such as local polynomials and splines, are often used [5-9].

The above regression approaches generally lead to satisfactory results when the subjects are properly randomized so that the treatments and the covariates are not subject to "selection bias". In many longitudinal studies, however, concomitant interventions are initiated, usually due to ethical reasons, to patients who exhibit less satisfactory trends in their medical outcomes. This phenomenon bears some resemblance to longitudinal studies with informative missing data, where patients with undesirable outcome trends tend to drop out early from the study, except that in our situation the outcomes of these patients are continuously observed under the additional concomitant treatment. In a randomized longitudinal clinical trial with pre-specified treatments, patients who have taken a concomitant intervention in addition to their assigned treatments may generally have different disease pathology from those who do not need the intervention. Thus, in addition to the primary goal of evaluating the effect of the main treatment, an important objective is to evaluate the effects of the additional concomitant intervention(s) among eligible patients.

Our motivating example is the Enhancing Recovery in Coronary Heart Disease (EN-RICHD), which is a randomized clinical trial that evaluated the efficacy of a psychosocial treatment versus usual cardiological care on survival and depression severity in 2,481 pa-

tients with depression and/or low perceived social support after acute myocardial infraction. Depression severity was measured by Hamilton Rating Scale for Depression (HRSD) and Beck Depression Inventory (BDI), where higher HRSD and BDI scores indicate worsened depression. In addition to the randomized treatments, patients with high baseline depression scores and/or nondecreasing BDI trends were eligible for pharmacotherapy with antidepressant. Antidepressants were also prescribed at the requests of the patients or their primary care physicians. Details of the study design, objectives and major findings of the trial have been described [10,11]. Taylor et al. [12] compared the survival rates for death and cardiovascular morbidity and mortality among 1,834 depressed patients in this trial and found that the use of selective serotonin reuptake inhibitors seemed to reduce subsequent cardiovascular morbidity and mortality. Bang and Robins [13] also analyzed the same data but a cross-sectional component only. However, the question of whether these antidepressant medications had added benefits for lowering the BDI scores of patients in the psychosocial treatment arm who had undergone this concomitant intervention during the trial was not addressed. To answer this question, we treat the patient's starting time of pharmacotherapy as a subject-specific "change-point" and evaluate the effects of pharmacotherapy on the patient's BDI scores over time using different models before and after the "change-point". We attempt to model the relationships between the subject's parameters for BDI trends and his/her change-point time using a B-spline nonparametric approach. It is important to note that in the ENRICHD trial pharmacotherapy was not simply initiated based on certain common HRSD or BDI threshold values, which may require a different methodology.

Methodologically, Murphy, van der Laan and Robins [14] studied estimation and causal inferences for the mean responses to dynamic treatment regimens that were tailored to subjects' individual needs. Their designs involve a set of treatment intervals specified by selected time points and a pre-specified sequential randomization rule that assigns subjects to different treatment levels. The main difference between our data structure and their dy-

namic regimens is that the concomitant interventions considered in this paper are often not assigned based on specific treatment rules, so that their estimating equations may not be directly applied to the current context. Also Wall, Dai and Eberly [15] examined the impact of a misspecified time-varying covariate when they analyzed the effect of (nonrandomized) alcoholism treatment on medical utilization in the GEE framework. They implicitly suggested that a change-point model was a way to go but did not provide a solution.

We propose in this paper a nonparametric approach for estimating the effects of a concomitant intervention in longitudinal clinical trials. The proposed methodology is based on a varying-coefficient mixed-effects model and a B-spline least-squares estimation method. A key question that has not been previously well-understood in the literature is why a naive linear mixed-effects model (e.g., Verbeke and Molenberghs [3], Sec. 3.3) could lead to biased estimates when a concomitant intervention is present. Using the general framework of the change-point shared-parameter model (Section 2.3), we show that the mixed-effects models without properly incorporating the joint distributions of the random parameters and the starting time of the concomitant intervention are misspecified models. On the other hand, our varying-coefficient mixed-effects model is a flexible nonparametric version of the change-point shared-parameter model which can adequately incorporate the concomitant intervention starting time when the joint distribution of the random parameters and the concomitant intervention starting time is completely unknown. Because our model may include both parametric and nonparametric components, our B-spline estimation method is more flexible than the usual local smoothing methods, such as kernel or local polynomial methods, in the sense that it can be naturally adapted to both parametric and nonparametric situations [8, 9]. Although certain components of our approach, such as B-splines, have been used for other longitudinal settings in the literature [5-9], the systematic modeling and estimation procedure proposed in this paper fills a gap for obtaining unbiased estimates in longitudinal clinical trials with the presence of a concomitant intervention.

Our modeling approach, however, shares some similarities with the shared parameter

model that addresses the informative missingness [16,17]. Follmann and Wu [17] used this approach to link a mixed-effects model for the response variable and a marginal model for the characteristics of missing data, such as time to drop-out. Their conditional model resembles our varying-coefficient model with only the observations before the change-points. In contrast, our model also includes the response curves after the change-points. Unlike the classical change-point problems where all the subjects have the same change-point and its location is the unknown parameter to be estimated, individual starting times of the concomitant intervention are observed in our data [18,19].

We describe our regression models and its biological interpretations in Section 2. Subsequently, we propose a class of nonparametric estimation and inference procedures based on B-splines in Sections 3 and 4, apply these methods to the ENRICHD data in Section 5, and present a simulation study in Section 6. Finally, we discuss some potential extensions of our methods in Section 7.

2   Change-Point Mixed-Effects Models

2.1   Data Structure

We consider a study with n randomly selected subjects. For the ith subject, n_i is the number of visits, $T_{ij} \in [\mathcal{T}_0, \mathcal{T}_1]$ is the *trial time* or *study time* at the jth visit, Y_{ij} is the real-valued outcome measured at T_{ij} and $\mathbf{X}_i = (1, X_{1i}, \ldots, X_{Pi})^T$ is the R^{P+1}-valued time-invariant covariate vector. Here $[\mathcal{T}_0, \mathcal{T}_1]$ is the known time interval for the study. We assume that the study has one concomitant intervention, and each subject has one change-point from non-intervention to intervention within the study period $[\mathcal{T}_0, \mathcal{T}_1]$. Let S_i be the ith subject's *intervention starting time* or *change-point time*, and $\delta_{ij} = 1_{[T_{ij} \geq S_i]}$ the intervention indicator at the jth visit. The time from the concomitant intervention to the jth visit is $R_{ij} = T_{ij} - S_i$. The observed data are $\{(T_{ij}, Y_{ij}, \mathbf{X}_i, S_i); \ 1 \leq j \leq n_i, \ 1 \leq i \leq n\}$.

Because our objective is to investigate the relationships between Y_{ij} and $(T_{ij}, \mathbf{X}_i)$ before

6

and after the change-point S_i, our model does not include subjects who have not received the concomitant intervention within the study period $[\mathcal{T}_0, \mathcal{T}_1]$. In some situations, these subjects may provide useful information for the pre-intervention covariate effects on Y_{ij}. However, models incorporating such subjects require additional assumptions and/or models on the treatment receipt pattern and disease pathology. Hence, discussion on this problem is out of the scope of this paper. When the context is clear, we may interchange intervention with concomitant intervention and denote the random variables by $(T, Y, \mathbf{X}, S, \delta, R)$.

2.2 A Naive Linear Mixed-Effects Model: Review

Intuitively we may evaluate the intervention effects by comparing the response trajectories before and after the change-points. For the ease of notation, we discuss the case without $\mathbf{X}$. Suppose that Y_{ij} are given by $Y_{ij} = a_{0i} + a_{1i}T_{ij} + \epsilon_{ij}$ when $T_{ij} < S_i$ and $(a_{0i} + b_{0i}) + (a_{1i} + b_{1i})T_{ij} + \epsilon_{ij}$ when $T_{ij} \geq S_i$ for some subject-specific parameters $(a_{0i}, a_{1i}, b_{0i}, b_{1i})$ and measurement errors ϵ_{ij}. The individual intervention effects for the ith subject are characterized by (b_{0i}, b_{1i}), and the marginal intervention effects for the population are then characterized by $E(b_{0i}, b_{1i})^T = (\beta_0, \beta_1)^T$. Using the framework of linear mixed-effects models (see Ch.3 in [3]), an intuitive change-point model is

$$
\begin{cases}
Y_{ij} = \mathbf{T}_{ij}^T \mathbf{a}_i + \left(\mathbf{T}_{ij}^*\right)^T \mathbf{b}_i + \epsilon_{ij}, \\
\left(\mathbf{a}_i^T, \mathbf{b}_i^T\right)^T \sim \mathrm{MVN}\left(\left(\alpha^T, \beta^T\right)^T, \Gamma\right) \quad \text{for some unknown } (\alpha, \beta, \Gamma),
\end{cases}
\tag{2.1}
$$

where, for known constants D_1 and D_2, $\mathbf{a}_i = (a_{0i}, \ldots, a_{D_1 i})^T$, $\mathbf{b}_i = (b_{0i}, \ldots, b_{D_2 i})^T$, $\mathbf{T}_{ij} = (1, T_{ij}, \ldots, T_{ij}^{D_1})^T$, $\mathbf{T}_{ij}^* = (\delta_{ij}, \delta_{ij}T_{ij}, \ldots, \delta_{ij}T_{ij}^{D_2})^T$, ϵ_{ij} are mean zero error processes, and $(\mathbf{a}_i^T, \mathbf{b}_i^T)^T$ and $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})^T$ are independent. The intervention effects are characterized by $\mathbf{b}_i$ for the ith subject and $E(\mathbf{b}_i) = \beta$ for the population. A crucial assumption of (2.1) is that $\{\mathbf{a}_i, \mathbf{b}_i\}$ and S_i are independent. Although it appears that the time-varying intervention is incorporated as a covariate by the term involving δ_{ij}, we will demonstrate in Sections 5 and Section 6 that, by ignoring the correlations between $\{\mathbf{a}_i, \mathbf{b}_i\}$ and S_i in the

distribution assumption of $\{\mathbf{a}_i, \mathbf{b}_i\}$, (2.1) is a misspecified model for our data and may lead to biased estimates for the intervention effects.

2.3 The Shared-Parameter Model

To model the initiation of the concomitant intervention, a natural extension of (2.1) is to allow the intervention starting time S_i to be correlated with the pre-intervention random coefficients $\mathbf{a}_i$ or more generally $\{\mathbf{a}_i, \mathbf{b}_i\}$. When the context is clear, we will denote by $\mu_1(\cdot; \mathbf{a}_i)$ and $[\mu_1(\cdot; \mathbf{a}_i) + \mu_2(\cdot; \mathbf{b}_i)]$ the subject-specific response curves before and after the start of the intervention, respectively. We interpret $\mu_2(\cdot; \mathbf{b}_i)$ as the intervention effect. Given $\{T_{ij}, \mathbf{X}_i, S_i\}$, our shared-parameter model is

$$
\begin{cases}
Y_{ij} = \mu_1(T_{ij}, \mathbf{X}_i; \mathbf{a}_i) + \delta_{ij}\, \mu_2(T_{ij}, \mathbf{X}_i, R_{ij}; \mathbf{b}_i) + \epsilon_{ij}, \\
\left(\mathbf{a}_i^T, \mathbf{b}_i^T, S_i\right)^T \sim \text{ Joint Distribution,}
\end{cases}
\tag{2.2}
$$

where ϵ_{ij} are mean zero errors with $cov(\epsilon_{ij_1}, \epsilon_{ij_2}) = \sigma_{ij_1 j_2}$, $\epsilon_{i_1 j_1}$ and $\epsilon_{i_2 j_2}$ are independent if $i_1 \neq i_2$, and, conditioning on $\{\mathbf{a}_i, \mathbf{b}_i\}$, S_i and $\{T_{ij}, \mathbf{X}_i\}$ are independent. In addition, we assume that $\{\mathbf{a}_i, \mathbf{b}_i\}$ and $\{T_{ij}, \mathbf{X}_i\}$ are independent.

Let $Y_i = (Y_{i1}, \ldots, Y_{in_i})^T$, $T_i = (T_{i1}, \ldots, T_{in_i})^T$ and $H(\cdot, \cdot)$ be the joint distribution function of $\{\mathbf{a}_i, \mathbf{b}_i\}$. The joint likelihood of $(Y_i^T, S_i)^T$ given $\{T_i, \mathbf{X}_i\}$ is

$$
f(Y_i, S_i | T_i, \mathbf{X}_i) = \int f(Y_i | T_i, \mathbf{X}_i, S_i, \mathbf{a}_i, \mathbf{b}_i) f(S_i | \mathbf{a}_i, \mathbf{b}_i)\, dH(\mathbf{a}_i, \mathbf{b}_i),
\tag{2.3}
$$

where $f(\cdot | \cdot)$ denotes the conditional density. Because of the extra $f(S_i | \mathbf{a}_i, \mathbf{b}_i)$ in the integrand, (2.3) differs from the usual likelihood functions for the mixed-effects models (see p24 in [3]).

Unlike (2.1), (2.2) is a change-point model with shared parameters $\{\mathbf{a}_i, \mathbf{b}_i\}$ which determine both the response curves of Y_{ij} and the distribution of S_i. The shared parameters approach was proposed for modeling the behaviors of informative missing data [7]. In (2.2), the correlation between S_i and $\mathbf{a}_i$ suggests that the ith subject's intervention starting time is determined by the pre-intervention response curve μ_1, while the correlation between S_i

and $\mathbf{b}_i$ suggests that S_i may also influence the response curve μ_2 that characterizes the intervention effect.

2.4 The Varying-Coefficient Mixed-Effects Model

The approach based on the joint likelihood (2.3) can be computationally complicated and requires some assumptions about the distribution of S_i. In this paper, we consider a simpler method based on the conditional model, which is robust to the distributional assumption of S_i. The conditional distribution can be written as

$$f(Y_i|S_i, T_i, \mathbf{X}_i) = \int f(Y_i|T_i, \mathbf{X}_i, S_i, \mathbf{a}_i, \mathbf{b}_i)dG(\mathbf{a}_i, \mathbf{b}_i|S_i). \tag{2.4}$$

Then we can rewrite (2.2) as a varying-coefficient model using the conditional distribution of $\{\mathbf{a}_i, \mathbf{b}_i\}$ given S_i. When μ_1 and μ_2 are linear functions, let $\mu_1(T_{ij}, \mathbf{X}_i; \mathbf{a}_i) = \mathbf{Z}_{ij}^T \mathbf{a}_i$ for $\mathbf{Z}_{ij} = (Z_{ij0}, \dots, Z_{ijD_1})^T$ generated by $\{(T_{ij}, \mathbf{X}_i);\ 1 \leq j \leq n_i,\ \delta_{ij} = 0\}$, and $\mu_2(T_{ij}, \mathbf{X}_i, S_i; \mathbf{b}_i) = \mathbf{W}_{ij}^T \mathbf{b}_i$ for $\mathbf{W}_{ij} = (W_{ij0}, \dots, W_{ijD_2})^T$ generated by $\{(T_{ij}, \mathbf{X}_i, S_i);\ 1 \leq j \leq n_i,\ \delta_{ij} = 1\}$. Writing $\alpha(S_i) = E(\mathbf{a}_i|S_i)$, $\beta(S_i) = E(\mathbf{b}_i|S_i)$, $\mathbf{a}_i^* = \mathbf{a}_i - \alpha(S_i)$ and $\mathbf{b}_i^* = \mathbf{b}_i - \beta(S_i)$, our varying-coefficient mixed-effects model has the expression

$$\begin{cases} Y_{ij} = \mathbf{Z}_{ij}^T \left[\alpha(S_i) + \mathbf{a}_i^* \right] + \delta_{ij} \mathbf{W}_{ij}^T \left[\beta(S_i) + \mathbf{b}_i^* \right] + \epsilon_{ij}, \\ \left(\mathbf{a}_i^{*T}, \mathbf{b}_i^{*T} \right)^T \Big| S_i \sim G(\cdot|S_i) \end{cases} \tag{2.5}$$

where, for $S_i = s$, $G(\cdot|s)$ is a distribution function with mean zero and covariance matrix $cov[(\mathbf{a}_i^{*T}, \mathbf{b}_i^{*T})^T|s] = \mathbf{C}(s)$. Marginal parameters of interest are $\alpha(s)$ and $\beta(s)$. When $S_i = s$, the mean intervention effect is $\beta(s)$, and $\beta(s) = 0$ for all $s \in (\mathcal{T}_0, \mathcal{T}_1)$ implies that the concomitant intervention had no marginal effect on the response curve.

An obvious choice for $G(\cdot|S_i)$ is the multivariate normal distribution with mean zero and covariance matrix $\mathbf{C} = cov[(\mathbf{a}_i^{*T}, \mathbf{b}_i^{*T})^T|s]$, which, for simplicity, is assumed to be time-invariant. Extensions to time-dependent covariances can be made by modeling $\mathbf{C}(s)$. Since explicit forms of $G(\cdot|S_i)$ are often unknown, modeling $\alpha(s)$ and $\beta(s)$ is often more important than modeling $\mathbf{C}(s)$. In linear models, we have $\alpha(s; \gamma) = (\alpha_0(s; \gamma_0), \dots, \alpha_{D_1}(s; \gamma_{D_1}))^T$,

9

$$\beta(s;\tau) = (\beta_0(s;\tau_0), \ldots, \beta_{D_2}(s;\tau_{D_2}))^T,$$

$$\alpha_d(s;\gamma) = \sum_{l=0}^{L_d} \gamma_{dl}\mathcal{T}_{dl}(s) \quad \text{and} \quad \beta_d(s;\tau) = \sum_{m=0}^{M_d} \tau_{dm}\mathcal{T}_{dm}^*(s) \qquad (2.6)$$

where $\{L_d, M_d\}$ are fixed, and $\{\mathcal{T}_{dl}(s), \mathcal{T}_{dm}^*(s)\}$ are known transformations of s. The choice of $\mathcal{T}_{dl}(s) = s^l$ and $\mathcal{T}_{dm}^*(s) = s^m$ leads to polynomials for (2.6).

Extended linear models can be used to approximate $\{\alpha(s), \beta(s)\}$ when their parametric forms are unknown. Let $\{\mathcal{B}_{d_1}(s) = (\mathcal{B}_{d_10}(s), \ldots, \mathcal{B}_{d_1\mathcal{L}_{d_1}}(s))^T; \ 0 \leq d_1 \leq D_1\}$ and $\{\mathcal{B}_{d_2}^*(s) = (\mathcal{B}_{d_20}^*(s), \ldots, \mathcal{B}_{d_2\mathcal{M}_{d_2}}^*(s))^T; \ 0 \leq d_2 \leq D_2\}$ be some pre-specified basis functions. Then $\alpha(s)$ and $\beta(s)$ can be approximated by

$$\alpha_d(s;\gamma) \approx \sum_{l=0}^{\mathcal{L}_d} \gamma_{dl}\mathcal{B}_{dl}(s) \quad \text{and} \quad \beta_d(s;\tau) \approx \sum_{m=0}^{\mathcal{M}_d} \tau_{dm}\mathcal{B}_{dm}^*(s), \qquad (2.7)$$

where $\mathcal{L}_d$ and $\mathcal{M}_d$ may tend to infinity as $n \to \infty$. Popular basis choices include truncated polynomial bases, Fourier bases or B-splines. In this paper, we restrict our attention to B-splines with fixed knot sequences because of their superior numerical stability. The smoothing parameters $\{\mathcal{L}_d, \mathcal{M}_d\}$ may be chosen subjectively or by a variable selection procedure, such as cross-validation and information criteria [8,9,20]. An alternative smoothing approach is to approximate $\{\alpha(s), \beta(s)\}$ by smoothing splines [21,22]. Because the explicit expressions and statistical properties of smoothing spline estimators are generally different from B-splines, we do not discuss this class of estimators in this paper.

3   Estimation Methods

3.1   Likelihood-Based Estimation

If (2.3) has an explicit parametric expressions, the parameters can be in principle estimated by maximizing the log-likelihood $\sum_{i=1}^{n} \log f(Y_i, S_i | T_i, \mathbf{X}_i)$. Suppose that (2.5) and (2.6) are satisfied, $G(\cdot|s)$ is Gaussian, and $(\epsilon_{i1}, \ldots, \epsilon_{in_i})^T \sim N(0, \Gamma_i)$. We can estimate $\{\gamma, \tau\}$ by

maximizing the partial likelihood

$$L\big(\{\gamma,\tau\}|T_i,\mathbf{X}_i,S_i\big) = \sum_{i=1}^{n} \log\left[\int f(Y_i|T_i,\mathbf{X}_i,S_i,\mathbf{a}_i,\mathbf{b}_i)\,dG(\mathbf{a}_i,\mathbf{b}_i|S_i)\right]. \qquad (3.1)$$

Let $\mathcal{W}_i$ be the matrix whose jth row is $(\mathbf{Z}_{ij}^T, \delta_{ij}\mathbf{W}_{ij}^T)$, $\mathcal{T}_d(s) = (\mathcal{T}_{d0}(s),\ldots,\mathcal{T}_{dL_d}(s))^T$, $\mathcal{T}_d^*(s) = (\mathcal{T}_{d0}^*(s),\ldots,\mathcal{T}_{dM_d}^*(s))^T$, $\mathcal{T}(s) = diag\{\mathcal{T}_0^T(s),\ldots,\mathcal{T}_{D_1}^T(s),\mathcal{T}_0^{*T}(s),\ldots,\mathcal{T}_{D_2}^{*T}(s)\}$, $\mathcal{T}(S_i) = \mathcal{T}_i$, and $\mathbf{V}_i$ be the covariance matrix of

$$e_{ij} = \mathbf{Z}_{ij}^T\mathbf{a}_i^* + \delta_{ij}\mathbf{W}_{ij}^T\mathbf{b}_i^* + \epsilon_{ij}, \qquad 1 \le j \le n_i. \qquad (3.2)$$

The matrix representation for (2.7) is $(\alpha^T(s;\gamma), \beta^T(s;\tau))^T = \mathcal{T}(s)(\gamma^T, \tau^T)^T$, where $\gamma_d = (\gamma_{d0},\ldots,\gamma_{dL_d})^T$, $\gamma = (\gamma_0^T,\ldots,\gamma_{D_1}^T)^T$, $\tau_d = (\tau_{d0},\ldots,\tau_{dM_d})^T$ and $\tau = (\tau_0^T,\ldots,\tau_{D_2}^T)^T$. When $\mathbf{V}_i$ are known, maximizing (3.1) leads to

$$\begin{pmatrix} \widehat{\gamma}_{ML}(\mathcal{T}) \\ \widehat{\tau}_{ML}(\mathcal{T}) \end{pmatrix} = \left\{\sum_{i=1}^{n}[\mathcal{W}_i\mathcal{T}_i]^T\,\mathbf{V}_i^{-1}\,[\mathcal{W}_i\mathcal{T}_i]\right\}^{-1}\left\{\sum_{i=1}^{n}[\mathcal{W}_i\mathcal{T}_i]^T\,\mathbf{V}_i^{-1}Y_i\right\} \qquad (3.3)$$

provided that $\sum_{i=1}^{n}[(\mathcal{W}_i\mathcal{T}_i)^T\mathbf{V}_i^{-1}(\mathcal{W}_i\mathcal{T}_i)]$ is nonsingular. When $\mathbf{V}_i$ are unknown but can be consistently estimated by a non-singular $\widehat{\mathbf{V}}_i$, we can estimate $\{\gamma,\tau\}$ by $\{\widetilde{\gamma}_{ML}(\mathcal{T}), \widetilde{\tau}_{ML}(\mathcal{T})\}$ which are given by (3.3) with $\mathbf{V}_i$ substituted by $\widehat{\mathbf{V}}_i$.

Substituting $\{\mathcal{T}_{dl}(s), \mathcal{T}_{dm}^*(s)\}$ in (3.3) with the basis functions $\{\mathcal{B}_{dl}(s), \mathcal{B}_{dm}^*(s)\}$, we can compute $\{\widehat{\gamma}_{ML}(\mathcal{B}), \widehat{\tau}_{ML}(\mathcal{B})\}$. Likelihood-based nonparametric estimators of $\{\alpha(s), \beta(s)\}$ under (2.7) and known $\mathbf{V}_i$ are

$$\left(\widehat{\alpha}_{ML}^T(s;\mathcal{B})), \widehat{\beta}_{ML}^T(s;\mathcal{B})\right)^T = \mathcal{B}(s)\left(\widehat{\gamma}_{ML}^T(\mathcal{B}), \widehat{\tau}_{ML}^T(\mathcal{B})\right)^T,$$

where $\mathcal{B}_i = \mathcal{B}(S_i)$, $\mathcal{B}(s)$ is defined similarly to $\mathcal{T}(s)$ with $\{\mathcal{T}_{dl}(s), \mathcal{T}_{dm}^*(s)\}$ replaced by $\{\mathcal{B}_{dl}(s), \mathcal{B}_{dm}^*(s)\}$. Nonparametric estimators computed with $\widehat{\mathbf{V}}_i$ used in (3.3) are

$$\left(\widetilde{\alpha}_{ML}^T(s;\mathcal{B}), \widetilde{\beta}_{ML}^T(s;\mathcal{B})\right)^T = \mathcal{B}(s)\left(\widetilde{\gamma}_{ML}^T(\mathcal{B}), \widetilde{\tau}_{ML}^T(\mathcal{B})\right)^T.$$

3.2 Least-Squares Based Estimation

Likelihood-based estimates of $\{\alpha(s), \beta(s)\}$ can not be computed when the explicit forms of $G(\cdot|S_i)$ and the distribution of ϵ_{ij} are unknown. In such situations, a practical approach is

11

to first parameterize $\{\alpha(s), \beta(s)\}$ by certain parametric models $\{\alpha(s; \gamma), \beta(s; \tau)\}$ and then derive the weighted least-squares estimators $\{\widehat{\gamma}_{LS}, \widehat{\tau}_{LS}\}$ which minimize

$$
\begin{aligned}
\ell(\gamma, \tau) \;=\; & \sum_{i=1}^{n} \Big\{ \big[Y_i - \big(\mathbf{Z}_i^T \alpha(S_i; \gamma) + (\delta \mathbf{W})_i^T \beta(S_i; \tau) \big) \big]^T \\
& \times \Lambda_i \big[Y_i - \big(\mathbf{Z}_i^T \alpha(S_i; \gamma) + (\delta \mathbf{W})_i^T \beta(S_i; \tau) \big) \big] \Big\},
\end{aligned}
\tag{3.4}
$$

where $\mathbf{Z}_i = (\mathbf{Z}_{i1}, \ldots, \mathbf{Z}_{in_i})^T$, $(\delta \mathbf{W})_i = (\delta_{i1} \mathbf{W}_{i1}, \ldots, \delta_{in_i} \mathbf{W}_{in_i})^T$, and Λ_i are some pre-specified symmetric nonsingular $n_i \times n_i$ weight matrices. The weighted least-squares estimators for (2.6) are

$$
\left(\begin{array}{c} \widehat{\gamma}_{LS}(\mathcal{T}) \\ \widehat{\tau}_{LS}(\mathcal{T}) \end{array} \right) = \left\{ \sum_{i=1}^{n} [\mathcal{W}_i \mathcal{T}_i]^T \Lambda_i [\mathcal{W}_i \mathcal{T}_i] \right\}^{-1} \left\{ \sum_{i=1}^{n} [\mathcal{W}_i \mathcal{T}_i]^T \Lambda_i Y_i \right\},
\tag{3.5}
$$

where $\sum_{i=1}^{n} [\mathcal{W}_i \mathcal{T}_i]^T \Lambda_i [\mathcal{W}_i \mathcal{T}_i]$ is nonsingular, and the jth row of $\mathcal{W}_i$ is $(\mathbf{Z}_{ij}^T, \delta_{ij} \mathbf{W}_{ij}^T)$. Substituting the basis approximations (2.7) in (3.4), the least-squares based nonparametric estimators of $\{\alpha(s), \beta(s)\}$ are

$$
(\widetilde{\alpha}_{LS}^T(s; \mathcal{B}), \widetilde{\beta}_{LS}^T(s; \mathcal{B}))^T = \mathcal{B}(s) \left(\widetilde{\gamma}_{LS}^T(\mathcal{B}), \widetilde{\tau}_{LS}^T(\mathcal{B}) \right)^T,
\tag{3.6}
$$

where $\{\widetilde{\gamma}_{LS}(\mathcal{B}), \widetilde{\tau}_{LS}(\mathcal{B})\}$ are given in (3.5) with $\mathcal{T}(s)$ replaced by $\mathcal{B}(s)$. Consistency and the rates of convergence for (3.6) can be derived [9].

Clearly, (3.5) and (3.6) are the same as the likelihood-based estimators when $\Lambda_i = \mathbf{V}_i^{-1}$ and normality assumptions hold. In practice, $\mathbf{V}_i$ are usually unknown and often difficult to estimate, so that subjective choices for Λ_i are used. Guidance on this choice is also available [8,9,23].

3.3   Estimation of the Covariances

The covariance structure $\mathbf{V}_i$ defined in Section 3.1 can be modeled in a number of ways. By the definition of e_{ij} in (3.2), the (j_1, j_2)th component of $\mathbf{V}_i$ is

$$
V_{i,j_1,j_2} = E(e_{ij_1} e_{ij_2}) = \rho_{i,j_1,j_2}(\mathbf{A}, \mathbf{B}, \mathbf{C}) + \sigma_{i,j_1,j_2},
\tag{3.7}
$$

12

where $\mathbf{A} = E(\mathbf{a}_i^* \mathbf{a}_i^{*T})$, $\mathbf{B} = E(\mathbf{b}_i^* \mathbf{b}_i^{*T})$, $\mathbf{C} = E(\mathbf{a}_i^* \mathbf{b}_i^{*T})$, $\sigma_{i,j_1,j_2} = E(\epsilon_{ij_1} \epsilon_{ij_2})$ and

$$\rho_{i,j_1,j_2}(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \mathbf{Z}_{ij_1}^T \mathbf{A} \mathbf{Z}_{ij_2} + \mathbf{Z}_{ij_1}^T \mathbf{C} \left(\delta_{ij_2} \mathbf{W}_{ij_2} \right) + \left(\delta_{ij_1} \mathbf{W}_{ij_1}^T \right) \mathbf{C} \mathbf{Z}_{ij_2}$$
$$+ \left(\delta_{ij_1} \mathbf{W}_{ij_1}^T \right) \mathbf{B} \left(\delta_{ij_2} \mathbf{W}_{ij_2} \right).$$

For the special case that ϵ_{ij} are independent measurement errors such that $\sigma_{i,j_1,j_2} = 0$ if $j_1 \neq j_2$ and σ^2 if $j_1 = j_2$, $\mathbf{V}_i$ adopts the parametric model $\mathbf{V}_i(\mathbf{A}, \mathbf{B}, \mathbf{C}, \sigma^2)$ with $V_{i,j_1,j_2} = \rho_{i,j_1,j_2}(\mathbf{A}, \mathbf{B}, \mathbf{C})$ if $j_1 \neq j_2$ and $\rho_{i,j,j}(\mathbf{A}, \mathbf{B}, \mathbf{C}) + \sigma^2$ if $j_1 = j_2 = j$. Other structures for $\mathbf{V}_i$ can be formulated by modeling σ_{i,j_1,j_2} [4].

For the general case of ϵ_{ij} having unknown correlation structures, σ_{i,j_1,j_2} is an non-parametric component in (3.7), hence can be either directly estimated or approximated by a parametric model. Under a different regression model, a local smoothing technique was suggested but can be computationally intensive [24]. To ease the computational burden, a consistent covariance estimator can be constructed by B-spline approximations [9]. σ_{i,j_1,j_2} can be approximated via B-spline by $\sigma_{i,j_1,j_2}(\mathbf{u}, \mathbf{v}) = \sum_{k=1}^{K_1} \sum_{l=1}^{K_1} u_{kl} \mathcal{B}_k(T_{ij_1}) \mathcal{B}_l(T_{ij_2})$ if $j_1 \neq j_2$, and $\sum_{k=1}^{K_2} v_k \mathcal{B}_k(T_{ij})$ if $j_1 = j_2 = j$, where $\{\mathcal{B}_k\}$ is a spline basis with a fixed knot sequence, $\mathbf{u} = \{u_{kl} = u_{lk}; \; k, l = 1, \ldots, K_1\}$ and $\mathbf{v} = \{v_k; k = 1, \ldots, K_2\}$. Substituting $\sigma_{i,j_1,j_2}(\mathbf{u}, \mathbf{v})$ into (3.7), $\mathbf{V}_i$ is approximated by $\mathbf{V}_i(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{u}, \mathbf{v})$ such that

$$V_{i,j_1,j_2} = \begin{cases} \rho_{i,j_1,j_2}(\mathbf{A}, \mathbf{B}, \mathbf{C}) + \sum_{k=1}^{K_1} \sum_{l=1}^{K_1} u_{kl} \mathcal{B}_k(T_{ij_1}) \mathcal{B}_l(T_{ij_2}), & \text{if } j_1 \neq j_2; \\ \rho_{i,j_1,j_2}(\mathbf{A}, \mathbf{B}, \mathbf{C}) + \sum_{k=1}^{K_2} v_k \mathcal{B}_k(T_{ij}), & \text{if } j_1 = j_2 = j. \end{cases}$$

Once an approximate parametric model for $\mathbf{V}_i$ is established, estimation of $\mathbf{V}_i$ can be achieved by least squares. Let $\widehat{e}_{ij} = Y_{ij} - [\mathbf{Z}_{ij}^T \widehat{\alpha}(S_i) + \delta_{ij} \mathbf{W}_{ij}^T \widehat{\beta}(S_i)]$ be the residual of Y_{ij} computed based on some consistent estimators $\widehat{\alpha}(s)$ and $\widehat{\beta}(s)$. If $V_{i,j_1,j_2} = \rho_{i,j_1,j_2}(\mathbf{A}, \mathbf{B}, \mathbf{C}) + \sigma_{i,j_1,j_2}(\mathbf{u}, \mathbf{v})$, we can estimate $\mathbf{V}_i$ by $\mathbf{V}_i(\widehat{\mathbf{A}}, \widehat{\mathbf{B}}, \widehat{C}, \widehat{\mathbf{u}}, \widehat{\mathbf{v}})$ where $(\widehat{\mathbf{A}}, \widehat{\mathbf{B}}, \widehat{C}, \widehat{\mathbf{u}}, \widehat{\mathbf{v}})$ minimizes

$$\sum_{i=1}^{n} \sum_{j_1, j_2 = 1, j_1 < j_2}^{n_i} \left\{ \widehat{e}_{ij_1} \widehat{e}_{ij_2} - \left[\rho_{i,j_1,j_2}(\mathbf{A}, \mathbf{B}, \mathbf{C}) + \sum_k \sum_l u_{kl} \mathcal{B}_k(T_{ij_1}) \mathcal{B}_l(T_{ij_2}) \right] \right\}^2$$

subject to $u_{kl} = u_{lk}$ when $j_1 \neq j_2$, and

$$\sum_{i=1}^{n} \sum_{j=1}^{n_i} \left\{ \widehat{e}_{ij}^2 - \left[\rho_{i,j,j}(\mathbf{A}, \mathbf{B}, \mathbf{C}) + \sum_k v_k \mathcal{B}_k(T_{ij}) \right] \right\}^2,$$

when $j_1 = j_2 = j$.

Here, $\mathbf{V}_i(\widehat{\mathbf{A}}, \widehat{\mathbf{B}}, \widehat{C}, \widehat{\mathbf{u}}, \widehat{\mathbf{v}})$ needs not to be positive definite for a finite sample, although, by consistency, it is asymptotically positive definite [9]. The problem of imposing finite sample positive definiteness to the spline estimators of $\mathbf{V}_i$ deserves substantial further investigation. The adequacy of $\mathbf{V}_i(\widehat{\mathbf{A}}, \widehat{\mathbf{B}}, \widehat{C}, \widehat{\mathbf{u}}, \widehat{\mathbf{v}})$ depends on the choices of knots and the degrees of the splines. Although it is possible to develop data-driven knots using cross-validation or the generalized cross-validation, statistical properties of such procedures are currently unknown. Subjective knot choices, such as using a few equal spaced knots, often give satisfactory results in biomedical applications.

4 Inferences

4.1 Inferences for Linear Models

Following the classical inferential framework with linear mixed-effects models, we consider the inferences for the fixed effects of (2.5) and (2.6) given $\{(\mathbf{Z}_i, \delta_{ij}, \mathbf{W}_{ij}, S_i); \ 1 \leq i \leq n, \ 1 \leq j \leq n_i\}$. From (3.5), $E[\widehat{\gamma}_{LS}(\mathcal{T})] = \gamma$, $E[\widehat{\tau}_{LS}(\mathcal{T})] = \tau$ and the covariance of $(\widehat{\gamma}_{LS}^T(\mathcal{T}), \widehat{\tau}_{LS}^T(\mathcal{T}))^T$ is

$$\left[\sum_{i=1}^{n}(\mathcal{W}_i\mathcal{T}_i)^T\Lambda_i(\mathcal{W}_i\mathcal{T}_i)\right]^{-1}\left[\sum_{i=1}^{n}(\mathcal{W}_i\mathcal{T}_i)^T\Lambda_i\mathbf{V}_i\Lambda_i(\mathcal{W}_i\mathcal{T}_i)\right]\left[\sum_{i=1}^{n}(\mathcal{W}_i\mathcal{T}_i)^T\Lambda_i(\mathcal{W}_i\mathcal{T}_i)\right]^{-1}.$$

Let $\widehat{\xi}_{LS}^T = (\widehat{\gamma}_{LS}^T(\mathcal{T}), \widehat{\tau}_{LS}^T(\mathcal{T}))^T$, $\xi = (\gamma^T, \tau^T)^T$, and $\mathcal{L}\widehat{\xi}_{LS}$ and $\mathcal{L}\xi$ be their corresponding linear combinations. Following the central limit theorem (see Sec.1.9.3 in [25]), it can be shown that, when n is sufficiently large, $\mathcal{L}\widehat{\xi}_{LS}$ is asymptotically distributed as $N(\mathcal{L}\xi, Var(\mathcal{L}\widehat{\xi}_{LS}))$, where $Var(\mathcal{L}\widehat{\xi}_{LS})$ can be derived from the covariance matrix of $\widehat{\xi}_{LS}$. Substituting $Var(\mathcal{L}\widehat{\xi}_{LS})$ with a consistent estimate $\widehat{Var}(\mathcal{L}\widehat{\xi}_{LS})$, an approximate $(1 - \alpha)$ confidence interval for $\mathcal{L}\xi$ is

$$\mathcal{L}\widehat{\xi}_{LS} \pm Z_{\alpha/2}\left[\widehat{Var}\left(\mathcal{L}\widehat{\xi}_{LS}\right)\right]^{1/2}, \tag{4.1}$$

where $Z_{\alpha/2}$ is the $[100 \times (1 - \alpha/2)]$th percentile of the standard normal distribution.

Let $\mathbf{L}$ be any given matrix with $rank(\mathbf{L}) = d$. The above asymptotic approximations can be used to test the null hypothesis, H_0: $\mathbf{L}\xi = \mathbf{C}$ for a known constant vector $\mathbf{C}$, versus the alternative, H_A: $\mathbf{L}\xi \neq \mathbf{C}$. An α-level approximate χ^2-test rejects the null hypothesis if

$$\left(\mathbf{L}\widehat{\xi}_{LS} - \mathbf{C}\right)^T \left\{ \mathbf{L}\left[Cov(\widehat{\xi}_{LS})\right]\mathbf{L}^T \right\}^{-1} \left(\mathbf{L}\widehat{\xi}_{LS} - \mathbf{C}\right) \geq \chi^2_{d,\alpha}, \tag{4.2}$$

where $\chi^2_{d,\alpha}$ is the $[100 \times (1-\alpha)]$th upper percentile of the χ^2_d-distribution.

4.2 Bootstrap Confidence Intervals

Nonparametric inferences for the smoothing estimators of $\alpha(s)$ and $\beta(s)$ can be constructed using the "resampling-subject" bootstrap [8,26]. For the construction of bootstrap pointwise confidence intervals (CIs), this procedure generates bootstrap samples $\{(Y_{ij}^b, \mathbf{Z}_{ij}^b, (\delta_{ij}\mathbf{W}_{ij})^b, S_i^b);$ $1 \leq i \leq n,\ 1 \leq j \leq n_i\}$ by sampling the subjects with replacement from the original data and obtains the spline estimators $\widetilde{\alpha}_{LS}^{(b)}(s, \mathcal{B})$ and $\widetilde{\beta}_{LS}^{(b)}(s, \mathcal{B})$. Repeat the procedure multiple times, and let $L_{\alpha/2}(\widetilde{\alpha}_{LS,d}(s, \mathcal{B}))$ and $U_{\alpha/2}(\widetilde{\alpha}_{LS,d}(s, \mathcal{B}))$ be the lower and upper $[100 \times (\alpha/2)]$th percentiles of the bootstrap estimators of $\alpha_d(s)$. An approximate $(1-\alpha)$ pointwise CI for $\alpha_d(s)$ is $[L_{\alpha/2}(\widetilde{\alpha}_{LS,d}(s, \mathcal{B})),\ U_{\alpha/2}(\widetilde{\alpha}_{LS,d}(s, \mathcal{B}))]$. Confidence intervals for $\beta_d(s)$ and other parameters can be constructed similarly. Moreover, variances of the parameters can be estimated by the sample variances of bootstrap estimates.

5 Application to Pharmacotherapy in the ENRICHD Study

As described in Section 1, our objective is to evaluate the additional effects of pharmacotherapy (antidepressants) on the trends of depression (measured by BDI scores) for patients who received pharmacotherapy during the six-month psychosocial treatment period. Because pharmacotherapy was only designed as a concomitant intervention in this trial, the starting time of pharmacotherapy was decided by the patients or their physicians. Unfortunately, since patients in the usual care arm did not have accurate pharmacotherapy starting time and repeated BDI scores recorded within the first six-month period, the effects

of pharmacotherapy could not be properly analyzed for these patients (refer to [11] for more details). Ninety one patients (total 1,446 observations) in the psychosocial treatment arm received pharmacotherapy as a concomitant intervention during this period and had clear records of their pharmacotherapy starting time. Among them, 43 started pharmacotherapy at baseline and 48 started pharmacotherapy between 7 and 172 days. The number of visits for these patients ranges from 5 to 36 and has the median of 16. Patients who did not have proper records of antidepressant use were excluded.

For the ith patient, Y_{ij}, T_{ij}, S_i, $R_{ij} = T_{ij} - S_i$ and $\delta_{ij} = 1_{[T_{ij} \geq S_i]}$ are the BDI score, trial time (in months), starting time of pharmacotherapy, time from initiation of pharmacotherapy, and pharmacotherapy indicator, respectively, at the jth visit. Our preliminary examination of the data revealed that the BDI scores over T_{ij} could be approximated by a linear model (results not shown). An intuitive model is the following special case of (2.1),

$$Y_{ij} = a_{0i} + a_{1i}T_{ij} + b_{0i}\delta_{ij} + b_{1i}\delta_{ij}R_{ij} + \epsilon_{ij}, \tag{5.1}$$

where $E(a_{0i}, a_{1i}, b_{0i}, b_{1i})^T = (\alpha_0, \alpha_1, \beta_0, \beta_1)^T$ and, when $\delta_{ij} = 1$ and $R_{ij} = r$, $(\beta_0 + \beta_1 r)$ describes the mean pharmacotherapy effect at r months since the start of pharmacotherapy. Clearly, (5.1) ignores the correlation between S_i and the pre-pharmacotherapy depression trends. To evaluate whether (5.1) leads to potential bias, we next considered the following special case of (2.5),

$$Y_{ij} = \alpha_0(S_i) + \alpha_1(S_i)T_{ij} + \beta_0\delta_{ij} + \beta_1\delta_{ij}R_{ij} + e_{ij}, \tag{5.2}$$

where $e_{ij} = a_{i0}^* + a_{i1}^*T_{ij} + b_{i0}^*\delta_{ij} + b_{i1}^*\delta_{ij}R_{ij} + \epsilon_{ij}$, $\alpha_0(S_i) = \gamma_{00} + \gamma_{01}S_i$ and $\alpha_1(S_i) = \gamma_{10} + \gamma_{11}S_i$. In (5.2), the mean pre-pharmacotherapy BDI trend is associated with S_i through intercept $\alpha_0(S_i)$ and slope $\alpha_1(S_i)$. At r months after the start of pharmacotherapy, the mean pharmacotherapy effect is $\beta_0 + \beta_1 r$, where a negative value for $\beta_0 + \beta_1 r$ corresponds to a beneficial effect for reducing depression. To reduce model complexity, we assume in (5.2) that $\beta_0(S_i) \equiv \beta_0$ and $\beta_1(S_i) \equiv \beta_1$ in the sense that the effects of pharmacotherapy only depend on how long the antidepressant has been used, but not on when it was started.

Table 1 summarizes the parameter estimates and their corresponding standard errors, 95% CIs and p-values obtained by the REML procedure with unstructured correlations. The negative estimates for (β_0, β_1) under (5.1) and (5.2) suggest that the beneficial effect of pharmacotherapy for this patient population is detected by both models. However, a slightly stronger depression lowering effect is exhibited by (5.2). The 95% CI for γ_{01} suggests a negative correlation of S_i with baseline BDI scores, so that patients with higher baseline BDI scores tend to start pharmacotherapy sooner.

6 Simulation

Following the general framework (2.3), we consider a simulation design that resembles the data structure of the ENRICHD trial. Each simulated sample contains $n = 200$ subjects. Each subject has 30 "scheduled visits" at time points $(T_{i1}, \ldots, T_{i,30}) = (0, 0.2 + e_1, \ldots, 5.8 + e_{29})$, where $\{e_l\}$ are independently generated from uniform $U(-0.2, 0.2)$ distribution, but each scheduled visit has 40% probability skipped. This leads to unequal numbers of repeated measurements among the subjects with n_i being the number of repeated measurements for the ith subject. The random parameters $(\mathbf{a}_i^T, \mathbf{b}_i^T)^T = (a_{0i}, a_{1i}, b_{0i}, b_{1i})^T$ are generated from the multivariate normal distribution with mean $(25, 0, -4, -2)^T$ and covariance matrix $cov(a_{0i}, a_{1i}, b_{0i}, b_{1i}) = diag(6.25, 1, 1, 1)$. For each $\{\mathbf{a}_i, \mathbf{b}_i\}$, we generate two different change-point times: (a) $S_i \sim N(10 - 0.3\, a_{0i},\ 0.16)$; and (b) $S_i \sim N(1 + 4\sin[(a_{0i} - 4)/9],\ 0.09)$. For each given $\{T_{ij}, S_i, \mathbf{a}_i, \mathbf{b}_i\}$, Y_{ij} is generated from $N(a_{0i} + a_{1i}T_{ij} + b_{0i}\delta_{ij} + b_{1i}\delta_{ij}R_{ij},\ 4)$. When $\{S_i;\ i = 1, \ldots, 200\}$ are generated from (a), direct calculation based on conditional normal distributions shows that the marginal model of Y_{ij} is (5.2) with $\gamma_{00} = 31.49$, $\gamma_{01} = -2.60$, $\gamma_{10} = \gamma_{11} = 0$, $\tau_0 = -4$ and $\tau_1 = -2$. When $\{S_i;\ i = 1, \ldots, 200\}$ are generated from (b), we assume that the parametric form of $\alpha_0(S_i)$ is unknown and the marginal model of Y_{ij} is

$$Y_{ij} = \alpha_0(S_i) + \alpha_1 T_{ij} + \beta_0 \delta_{ij} + \beta_1 \delta_{ij} R_{ij} + e_{ij}, \tag{6.1}$$

where $e_{ij} = a_{i0}^* + a_{i1}^* T_{ij} + b_{i0}^* \delta_{ij} + b_{i1}^* \delta_{ij} R_{ij}$, $\beta_0 = -4$, $\beta_1 = -2$ and $\alpha_0(S_i)$ is not a linear function.

The simulation was repeated 2,000 times. For samples with S_i generated from (a), we first ignored the correlation between a_{0i} and S_i and estimated $(\alpha_0, \alpha_1, \beta_0, \beta_1)$ using (5.1) with unstructured correlations, and then fitted the data to (5.2) with $\alpha_1(S_i) \equiv \alpha_1$ and estimated $(\gamma_{00}, \gamma_{01}, \alpha_1, \beta_0, \beta_1)$, all by REML. Table 2 summarizes the averages of the estimates, their standard errors, and root mean-squared errors as well as the empirical coverage probabilities of the 95% asymptotic CIs. The bias for the estimation of β_0 in (5.1) can be seen from the large average root mean-squared errors and the low coverage probabilities compared with the estimates obtained using (5.2).

For samples with S_i generated from (b), we approximated $\alpha_0(s)$, $s \in [0, 6]$, using the quadratic B-spline with 2 equally spaced interior knots (see Ch.5.2 in [27]), and estimated $(\alpha_0(s), \alpha_1, \beta_0, \beta_1)$ under (6.1) using (3.5) and (3.6) with the $\Lambda_i = I_{n_i \times n_i}$ weight. We computed the 95% bootstrap CIs for $(\alpha_1, \beta_0, \beta_1)$ and the 95% pointwise bootstrap CIs for $\alpha_0(s)$ at 60 equally spaced values of $s \in [0, 6]$ using the percentile procedures with $B = 300$ bootstrap replications. For comparison, we also fitted the data to (5.1), which assumes that $\alpha_0(s) \equiv \alpha_0$, and estimated $(\alpha_0, \alpha_1, \beta_0, \beta_1)$ using the same procedure of (3.5). Figure 1(a) shows the spline-estimated coefficient curve $\alpha_0(s)$ and the 95% pointwise bootstrap CIs obtained from a randomly selected simulated sample and Figure 1(b) displays the empirical coverage probabilities of the 95% pointwise bootstrap CIs, where true $\alpha_0(s)$ is numerically calculated. Table 3 presents the same set of summary statistics used in Table 2 for $(\alpha_1, \beta_0, \beta_1)$ under both (5.1) and (6.1). Standard errors and CI coverage probabilities based on either the bootstrap procedure (Section 4.2) or the least squares and normal approximation procedure (Sections 3.3 and 4.1) are compared, and the performances of the two procedures are similar under each model. The large root mean-squared errors and poor coverage probabilities for the estimates obtained under (5.1) suggest that ignoring the association between a_{0i} and S_i may lead to erroneous conclusions in the present situation.

18

7 Discussion

Our proposed methodology is focused on concomitant interventions in longitudinal clinical trials and such interventions commonly appear in other settings. For example, subjects in an epidemiological study may take antihypertensive medication during the study when their blood pressure levels either exhibit some undesirable trends or stay in an intolerable range. Crucial in dealing with this type of data is to model the intervention selection mechanism as realistic as possible. In our pharmacotherapy example of the ENRICHD trial, there was only a vague guideline for the initiation of pharmacotherapy, so that it appeared reasonable to model the intervention selection mechanism through some shared-parameters. We focus on the varying-coefficient model mainly because it has a simple and clear biological interpretation for this example, its assumptions seem to be realistic for this type of trials, and the nonparametric B-spline method can be easily implemented.

We also find that there are several possible extensions worthy of further investigation. First, our data structure allows for only a single intervention with one change-point per subject. Generally, subjects in longitudinal studies may have single or multiple concomitant interventions which can be turned on or off at different time points. In such situations, more general shared-parameter models may be needed to accommodate the possibility of multiple interventions and/or multiple change-points. Second, our model relies on linear functions to describe the trends before and after the intervention. It can be generalized to models with nonlinear response curves. Finally, we use the classical frequentist's framework for the B-spline methods. In a different context, Fahrmeir and Lang [28] demonstrated a promising Bayesian inference procedure for generalized additive mixed models based on Markov random field priors. Analogous approaches for our model and estimators may lead to useful confidence regions and model diagnostic procedures.

References

[1] Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38**:963–974.

[2] Davidian M, Giltinan DM. *Nonlinear Models for Repeated Measurement Data.* London; New York: Chapman Hall, 1995.

[3] Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data.* New York: Springer, 2000.

[4] Diggle PJ, Heagerty P, Liang K-Y, Zeger SL. *Analysis of Longitudinal Data.*, 2nd ed. Oxford: Oxford University Press, 2002.

[5] Fan J, Zhang JT. Functional linear models for longitudinal data. *Journal of the Royal Statistical Society. Ser. B* 2000; **62**:303–322.

[6] Lin X, Carroll RJ. Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association* 2000; **95**:520–534.

[7] Lin X, Carroll, RJ. Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association* 2001; **96**:1045–1056.

[8] Huang JZ, Wu CO, Zhou L. Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* 2002; **89**:111–128.

[9] Huang JZ, Wu CO, Zhou L. Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica* 2004; **14**:763–788.

[10] The ENRICHD Investigators. Enhancing recovery in coronary heart disease patients (ENRICHD): Study intervention rationale and design. *Psychosomatic Medicine* 2001; **63**:747–755.

[11] The ENRICHD Investigators. Enhancing recovery in coronary heart disease patients (ENRICHD): The effects of treating depression and low perceived social support on clinical events after myocardial infarction. *Journal of the American Medical Association* 2003; **289**:3106–3116.

[12] Taylor CB, Youngblood ME, Catellier D, Veith RC, Carney RM, Burg MM, Kaufmann P, Shuster J, Mellman T, Blumenthal JA, Krishnan R, Jaffe AS. Effects of antidepressant medication on morbidity and mortality in depressed patients after myocardial infarction. *Archives of General Psychiatry* 2005; **62**:792–798.

[13] Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics* 2005; **61**:962–972.

[14] Murphy SA, van der Laan MJ, Robins JM. Marginal mean models for dynamic regimes. *Journal of the American Statistical Association* 2001; **96**:1410–1423.

[15] Wall MM, Dai Y, Eberly LE. GEE estimation of a misspecified time-varying covariate: an example with the effect of alcoholism treatment on medical utilization. *Statistics in Medicine* 2005; **24**:925–939.

[16] Wu MC, Carroll R. Estimation and comparison fo changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* 1988; **44**:175–188.

[17] Follmann D, Wu M. An approximate generalized linear model with random effects for informative missing data. *Biometrics* 1995; **51**:151–168

[18] Naumova EN, Must A, Laird NM. Tutorial in Biostatistics: Evaluating the impact of critical periods in longitudinal studies of growth using piecewise mixed effects models. *International Journal of Epidemiology* 2001; **30**:1332–1341.

[19] Bang H, Mazumdar M, Spence, JD. Tutorial in Biostatistics: Analyzing associations between total plasma Homocysteine and B vitamins using optimal categorization and segmented regression. *Neuroepidemiology* 2006; **27**:188–200.

[20] Rice JA, Wu CO. Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* 2001;**57**:253–259.

[21] Lin X, Zhang D. Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society. Ser. B* 1999; **61**:381–400.

[22] Chiang CT, Rice JA, Wu CO. Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *Journal of the American Statistical Association* 2001; **96**:605-619.

[23] Welsh AH, Lin X, Carroll RJ. Marginal longitudinal nonparametric regression: Locality and efficiency of spline and kernel methods. *Journal of the American Statistical Association* 2002; **97**:482-493.

[24] Diggle PJ, Verbyla AP. Nonparametric estimation of covariance structure in longitudinal data. *Biometrics* 1998; **54**:401–415.

[25] Serfling RJ. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York 1980.

[26] Hoover DR, Rice JA, Wu CO, Yang LP. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* 1998; **85**:809–822.

[27] Hastie TJ, Tibshirani RJ, Friedman J. *The Elements of Statistical Learning; Data Mining, Inference, and Prediction.* New York: Springer 2001.

[28] Fahrmeir L, Lang S. Bayesian inference for generalized additive mixed models based on Markov random field priors. *Applied Statistics* 2001; **50**:201–220.

Table 1. The ENRICHD Data Analysis

Model	Parameter	Estimate	SE	95% CI	p-value
(5.1)	α_0	23.380	1.107	(21.167, 25.594)	<0.0001
	α_1	-0.619	0.479	(-1.577, 0.339)	0.199
	β_0	-3.410	0.994	(-5.399, -1.422)	0.0013
	β_1	-1.584	0.521	(-2.626, -0.542)	0.0039
(5.2)	γ_{00}	25.670	1.431	(22.808, 28.533)	<0.0001
	γ_{01}	-1.389	0.586	(-2.562, -0.216)	0.0180
	γ_{10}	-0.278	0.822	(-1.922, 1.366)	0.736
	γ_{11}	0.078	0.174	(-0.272, 0.426)	0.654
	β_0	-4.302	1.041	(-6.385, -2.220)	0.0001
	β_1	-2.062	0.773	(-3.608, -0.516)	0.0105

Parameter estimates and their standard errors (SE), 95% confidence intervals (CIs) and p-values were obtained by restricted maximum likelihood with unstructured correlations for models (5.1) and (5.2).

Table 2. Simulation results for (a) $S_i \sim N(10 - 0.3\, a_{0i},\ 0.16)$

Model	Parameter	Estimate	SE	$\sqrt{\text{MSE}}$	CP
(5.1)	$\alpha_0 = 25$	25.018	0.202	0.211	0.940
	$\alpha_1 = 0$	-0.058	0.100	0.123	0.897
	$\beta_0 = -4$	-3.805	0.153	0.253	0.745
	$\beta_1 = -2$	-1.971	0.115	0.123	0.936
(5.2)	$\gamma_{00} = 31.488$	31.480	0.350	0.359	0.942
	$\gamma_{01} = -2.595$	-2.592	0.131	0.134	0.943
	$\alpha_1 = 0$	0.001	0.099	0.099	0.951
	$\beta_0 = -4$	-4.004	0.152	0.154	0.946
	$\beta_1 = -2$	-1.998	0.114	0.114	0.955

Estimate, SE and $\sqrt{\text{MSE}}$ denote the averages of estimates, standard errors and square root of the mean squared errors and CP represents the estimated coverage probability of the 95% confidence intervals, computed from 2,000 simulated samples. Parameter estimates and SEs were obtained by restricted maximum likelihood with unstructured correlations.

Table 3. Simulation results for (b) $S_i \sim N(1 + 4\sin[(a_{0i} - 4)/9], 0.09)$

Model	Parameter	Estimate	SE (SE*)	$\sqrt{\text{MSE}}$	CP (CP*)
(5.1)	$\alpha_1 = 0$	-0.862	0.150 (0.145)	0.874	0 (0)
	$\beta_0 = -4$	-2.404	0.340 (0.329)	1.632	0.004 (0.007)
	$\beta_1 = -2$	-0.237	0.372 (0.390)	1.804	0.005 (0.012)
(6.1)	$\alpha_1 = 0$	-0.001	0.097 (0.101)	0.095	0.951 (0.956)
	$\beta_0 = -4$	-3.999	0.264 (0.255)	0.267	0.942 (0.940)
	$\beta_1 = -2$	-1.991	0.258 (0.247)	0.259	0.944 (0.921)

Estimate, SE and $\sqrt{\text{MSE}}$ denote the averages of estimates, bootstrap standard errors and square root of the mean squared errors and CP represents the estimated coverage probability of the 95% bootstrap confidence intervals, computed from 2,000 simulated samples. SE* and CP* denote the standard errors and coverage probability of CIs obtained by the procedures in Sections 3.3 and 4.1. Parameter estimates were obtained using (3.5) with $\Lambda_i = I_{n_i \times n_i}$.
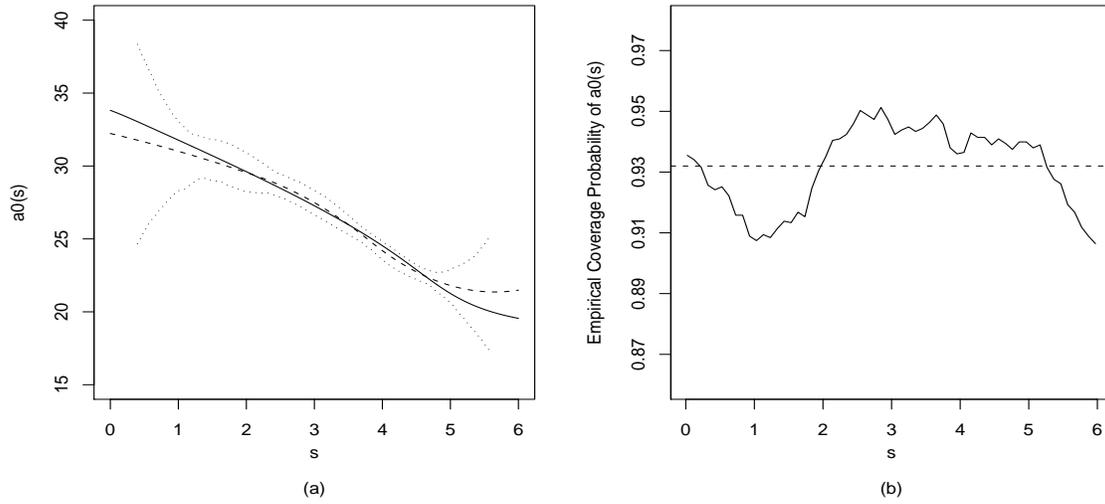
Figure 1: *(a) True curve $\alpha_0(s)$ (solid line), spline estimated curve $\widetilde{\alpha}_0(s)$ (dash line) and pointwise 95% bootstrap percentile confidence intervals (dotted lines) obtained from a randomly selected simulated sample. (b) Empirical coverage probability of pointwise 95% bootstrap confidence intervals for $\alpha_0(s)$ (solid line) and their sample mean (dash line).*