A shared parameter model for the estimation of longitudinal concomitant intervention effects

COLIN O. WU*, XIN TIAN, WENHUA JIANG
Office of Biostatistics Research,
National Heart, Lung and Blood Institute, Bethesda, MD 20892, USA
wuc@nhlbi.nih.gov

SUMMARY
We investigate a change-point approach for modeling and estimating the regression effects caused by a concomitant intervention in a longitudinal study. Since a concomitant intervention is often introduced when a patient’s health status exhibits undesirable trends, statistical models without properly incorporating the intervention and its starting time may lead to biased estimates of the intervention effects. We propose a shared parameter change-point model to evaluate the pre- and postintervention time trends of the response and develop a likelihood-based method for estimating the intervention effects and other parameters. Application and statistical properties of our method are demonstrated through a longitudinal clinical trial in depression and heart disease and a simulation study.

Keywords: Change-point model; Concomitant intervention; Likelihood; Longitudinal study; Shared parameter model.

1. INTRODUCTION
Change-point models are often used for evaluating the intervention effects in longitudinal clinical trials or epidemiological studies. By treating the starting time of an intervention as a change-point time, the longitudinal trends of the response variable in these models are described using different regression models before and after the intervention. Estimation and inferences can be obtained using well-known procedures, such as the maximum likelihood estimation (MLE), restricted maximum likelihood estimation (REMLE), and generalized estimation equations (e.g. Verbeke and Molenberghs, 2000; Diggle and others, 2002). This approach works well when the intervention is randomly assigned. When an intervention is not randomly assigned, treating the intervention as a time-dependent covariate in a change-point model may lead to model misspecification and biased estimation of the intervention effects (Wall and others, 2005).

Concomitant interventions are generally not randomly assigned in biomedical studies. The Enhancing Recovery in Coronary Heart Disease (ENRICHD) Study is a typical example which involves a concomitant intervention in addition to the randomly assigned treatment regimens. This is a randomized clinical trial for evaluating the efficacy of a 6-month psychosocial treatment (PT) versus the usual cardiovascular care (UC) on survival, hospitalization, myocardial infarction, and depression severity in 2481 patients with depression and/or low perceived social support after acute myocardial infarction (ENRICHD, 2001, 2003). Depression severity was measured by Hamilton Rating Scale for Depression (HRSD) at the baseline and

*To whom correspondence should be addressed.
6-month visits and Beck Depression Inventory (BDI) at regularly scheduled visits during the treatment and follow-up periods, where higher HRSD and BDI scores indicate worsened depression. BDI scores for patients in the PT arm were repeatedly measured at weekly visits during the treatment and 4 yearly follow-up visits, while BDI scores for patients in the UC arm were only measured at baseline, the 6-month visit and the yearly follow-up visits. Patients in both treatment arms were eligible for pharmacotherapy with antidepressants as a concomitant intervention if they had high baseline BDI scores or nondecreasing BDI trends after enrollment. In addition, antidepressants were prescribed at the requests of the patients or their primary care physicians. Taylor and others (2005) investigated the effects of antidepressants on cardiovascular morbidity and mortality among 1834 depressed ENRICHD patients and found that pharmacotherapy improved survival for this patient population. Their results, however, did not address the question of whether pharmacotherapy was beneficial for lowering patients’ depression severity. In an analysis using a subsample of 91 ENRICHD patients in the PT arm who have received pharmacotherapy within the treatment period, Wu and others (2008) showed that the naïve mixed-effects models were misspecified when a concomitant intervention was present and proposed a varying-coefficient mixed-effects model to reduce the potential bias associated with the estimated pharmacotherapy effects. A main drawback of Wu and others (2008) is the potential loss of information because their regression model cannot be applied to patients who have already received a concomitant intervention at baseline or have not received any concomitant intervention during the study.

We propose in this paper a comprehensive regression method for evaluating the concomitant intervention effects which is capable to incorporate information from all the patients in a longitudinal study. Based on the framework of shared parameter models (Follmann and Wu, 1995), we describe the covariate effects on the response variable through a change-point mixed-effects model and incorporate the random coefficients and the intervention starting time (change-point time) through a series of joint distributions. Patients who have received a concomitant intervention at baseline or have not received any concomitant intervention during the study are treated as censored. A likelihood-based method is established for statistical estimation and inferences and its computation is implemented through a 2-stage iteration procedure. Our application to the ENRICHD pharmacotherapy data and simulation results suggest that the proposed method leads to adequate estimates when a concomitant intervention is present, while the naïve mixed-effects model is likely misspecified under such situations.

We describe our shared parameter model and its interpretations in Section 2, develop the likelihood-based estimation and inference procedures in Section 3, and present the application to the ENRICHD pharmacotherapy data and simulation results in Sections 4 and 5, respectively. In Section 6, we discuss some potential extensions of our method. Further theoretical and technical results can be found in the supplementary material available at Biostatistics online.

2. Change-point shared parameter models

2.1 Data structure

Let \( T_0 \) and \( T_1 \) be the beginning and ending times of a study, and \( n \) the total number of randomly selected subjects. The \( i \)th subject has \( n_i \) visits and observations \((T_{ij}, Y_{ij}, X_i)\) at the \( j \)th visit, where \( T_{ij} \), the study time, is the time elapsed from the beginning of the study to the \( j \)th visit, \( X_i \) is a time-invariant covariate vector, and \( Y_{ij} \) is the real-valued outcome variable. For simplicity, we assume in this paper that the study involves only one concomitant intervention, and each subject can change from “without intervention” to “intervention” only once during the study. Extension to multiple concomitant interventions and multiple intervention changes involves more notation and computational complexity, hence, is not discussed in the current manuscript. When the context is clear, we may simply refer a “concomitant intervention” as an intervention. Define \( S_i \) to be the intervention starting time or change-point time, and \( \lambda_{ij} = 0 \) or 1, if \( T_{ij} \leq S_i \) or \( T_{ij} > S_i \), respectively, to be the intervention indicator for the \( i \)th subject. Since not every
subject changes from without intervention to intervention during the study, the $i$th subject’s change-point time is observed if $T_{i1} \leq S_i \leq T_{in}$. If $S_i < T_{i1}$ or $S_i > T_{in}$, the subject’s change-point time is left or right censored, respectively. The indicator variable for censoring $\delta_i^{(c)}$ is defined by $\delta_i^{(c)} = 0$ if $T_{i1} \leq S_i \leq T_{in}$, 1 if $S_i > T_{in}$, and 2 if $S_i < T_{i1}$. The observed change-point times are $\{S_i^{(c)} = (S_i^{(c)}, \delta_i^{(c)}); i = 1, \ldots, n\}$, where $S_i^{(c)} = S_i$ if $\delta_i^{(c)} = 0$, $T_{in}$ if $\delta_i^{(c)} = 1$, and $T_{i1}$ if $\delta_i^{(c)} = 2$.

### 2.2 Model formulation and rationale

The change-point model assumes that $Y_{ij}$ follows different trajectories before and after an intervention. The $i$th subject’s trajectory before the intervention is $\mu_0(T_{ij}, X_i; a_i)$, which is parameterized by the subject-specific parameter $a_i^T = (a_{i1}, \ldots, a_{id_0})^T, d_0 \geq 1$. The change of the trajectory after the intervention is $\mu_1(T_{ij}, X_i, R_{ij}; b_i)$, which is parameterized by the subject-specific parameter $b_i = (b_{i1}, \ldots, b_{id_1})^T, d_1 \geq 1$, and may depend on the “intervention duration time” $R_{ij} = T_{ij} - S_i$ as well as $T_{ij}$ and $X_i$. Following the usual mixed-effects model framework (Davidian and Giltinan, 1995; Verbeke and Molenberghs, 2000; Diggle and others, 2002), a naïve mixed-effects model for evaluating the pre- and postintervention trajectories is

$$
\begin{align*}
Y_{ij} = & \mu_0(T_{ij}, X_i; a_i) + \lambda_{ij} \mu_1(T_{ij}, X_i, R_{ij}; b_i) + \epsilon_{ij}, \\
(a_i^T, b_i^T)^T & \sim \text{joint distribution } G(\cdot),
\end{align*}
$$

(2.1)

where $(a_i^T, b_i^T)^T$ has unknown mean $(\alpha^T, \beta^T)^T$ and covariance matrix $\Sigma$, $\epsilon_{ij}$ are mean-zero random errors with cov($\epsilon_{ij1}, \epsilon_{ij2}$) = $\sigma_{ij1j2}$, and $\epsilon_{ij1}$ and $\epsilon_{ij2}$ are independent if $i_1 \neq i_2$. The joint distribution of $a_i$ and $b_i$ is usually assumed to be multivariate Gaussian $\mathcal{N}((\alpha^T, \beta^T)^T, \Sigma)$. A positive (or negative) value for $\mu_1(T_{ij}, X_i, R_{ij}; b_i)$ would suggest that the intervention tends to increase (or decrease) the mean of $Y_{ij}$ given $(T_{ij}, X_i, R_{ij})$. Wu and others (2008) showed that (2.1) could be a misspecified model even if $\mu_0(\cdot; a_i), \mu_1(\cdot; b_i), \epsilon_{ij}$, and $G(\cdot)$ were correctly specified because the “self-selectiveness” of the intervention was ignored.

If the change-point time $S_i$ depends on the preintervention trend of $Y_{ij}$, a natural extension of (2.1) is

$$
\begin{align*}
Y_{ij} = & \mu_0(T_{ij}, X_i; a_i) + \lambda_{ij} \mu_1(T_{ij}, X_i, R_{ij}; b_i) + \epsilon_{ij}, \\
a_i & \sim F_{\alpha}(\cdot), \ S_i | a_i \sim F_{\alpha}(\cdot|a_i), \ b_i | a_i \sim F_{\beta}(\cdot|a_i),
\end{align*}
$$

(2.2)

where $F_{\alpha}(\cdot)$ is the cumulative distribution function (CDF) of $a_i$, $F_{\beta}(\cdot|a_i)$ and $F_{\beta}(\cdot|a_i)$ are the conditional CDFs of $S_i$ and $b_i$, respectively, given $a_i$, and $a_i$ and $S_i$ are independent given $a_i$. In contrast to the varying-coefficient model of Wu and others (2008), where the conditional means of $a_i$ and $b_i$ given $S_i$ are used, (2.2) incorporates $S_i$ through $F_{\alpha}(\cdot|a_i)$, which allows $S_i$ to be left or right censored. We assume for simplicity in (2.2) that $b_i$ does not depend on $S_i$, although further generalizations may allow the distribution of $b_i$ to depend on $(S_i, a_i)$. In practice, further specifications of $F_{\alpha}(\cdot), F_{\alpha}(\cdot|a_i)$, and $F_{\beta}(\cdot|a_i)$ are usually needed to balance the computational feasibility and flexibility of the model.

### 2.3 Linear and additive shared parameter models

A useful and mathematically tractable specification for (2.2) is to assume that $S_i$ and $b_i$ depend on $a_i$ only through their conditional means. If the conditional means of $S_i$ and $b_i$ are linear models of $a_i$, a linear shared parameter model for (2.2) is

$$
\begin{align*}
Y_{ij} = & \mu_0(T_{ij}, X_i; a_i) + \lambda_{ij} \mu_1(T_{ij}, X_i, R_{ij}; b_i) + \epsilon_{ij}, \\
a_i = & \alpha + e_i^{(a)}, \ S_i = \gamma^T (1, a_i^T)T + e_i^{(s)}, \ b_i = \beta^T (1, a_i^T)T + e_i^{(b)},
\end{align*}
$$

(2.3)
where \( \alpha = (\alpha_1, \ldots, \alpha_d)^T, \alpha_d \in \mathbb{R}, \beta = (\beta_1^T, \ldots, \beta_d^T)^T, \beta_1 = (\beta_{10}, \ldots, \beta_{d1})^T, \beta_{id} \in \mathbb{R}, \gamma = (\gamma_0, \ldots, \gamma_{d1})^T, \) and \( \{\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{ini})^T, \epsilon_i^{(a)}, \epsilon_i^{(b)}, \epsilon_i^{(c)}\} \) are independent mean-zero random errors with covariance matrices \( \{V_{y}, V_{a_d}, V_{b}, \sigma^2_s\} \), respectively. Here, the unknown parameters are the mean components \( \theta = (\alpha^T, \beta_1^T, \ldots, \beta_d^T, \gamma^T)^T \) and the covariance structures \( \mathbf{V} = \{V_{y}, V_{a}, V_{b}, \sigma^2_s\} \).

When the relationship between \( S_i \) and \( a_i \) is unknown, a nonparametric model is \( S_i = \mu^{(s)}(a_i) + \epsilon_i^{(s)} \), where \( \mu^{(s)}(a_i) \equiv E(S_i|a_i) \) is a smooth function of \( a_i \). Since unstructured estimation of \( \mu^{(s)}(a_i) \) could be difficult when \( a_i \) is a high-dimensional vector, a simple additive model is to replace the relationship between \( S_i \) and \( a_i \) in (2.3) with

\[
S_i = \sum_{d=0}^{d_0} \mu_d^{(s)}(a_{id}) + \epsilon_i^{(s)},
\]

where \( \mu_d^{(s)}(a_{id}) \) are smooth functions of \( a_{id} \). Further generalizations of (2.4) are theoretically possible but at the expense of computational complexity.

### 3. Likelihood-based estimation method

#### 3.1 General formulation and approximate log-likelihood for splines

The parameters in (2.2) can be estimated by a likelihood approach if the distributions can be specified. Let \( Y_i = (Y_{i1}, \ldots, Y_{ini})^T, T_i = (T_{i1}, \ldots, T_{ini})^T, D_i = (T_i, X_i) \), and \( f_y(\cdot), f_b(\cdot), f_s(\cdot), f_a(\cdot) \) be the densities of \( Y_i, b_i, S_i, \) and \( a_i \). The joint density of \( (b_i, S_i, a_i) \) is \( f(b_i, S_i, a_i) = f_b(b_i|a_i) f_s(S_i|a_i) f_a(a_i) \). Integrating over \( a_i \) and \( b_i \), the conditional density of \( (Y_i, S_i) \) given \( D_i = (T_i, X_i) \) is

\[
f_{(y,s)}(Y_i, S_i|D_i) = \int \int f_y(Y_i|D_i, S_i, a_i, b_i) f_b(b_i|a_i) f_s(S_i|a_i) f_a(a_i) da_i db_i.
\]

The conditional density \( f_s(S_i^{(c)}, \delta_i^{(c)}|a_i) \) of \( S_i^{(c)} = (\delta_i^{(c)}, \delta_i^{(c)}) \) given \( a_i \) is \( f_s(S_i|a_i) \) if \( \delta_i^{(c)} = 0 \), \( 1 - F_s(T_{ini}|a_i) \) if \( \delta_i^{(c)} = 1 \), and \( F_s(T_{ini}|a_i) \) if \( \delta_i^{(c)} = 2 \). The log-likelihood function for \( (Y_i, S_i^{(c)}) \) conditioning on \( D_i, i = 1, \ldots, n \), is

\[
L_c = \frac{1}{n} \sum_{i: \delta_i^{(c)} = 0} \log f_{(y,s)}(Y_i, S_i|D_i) + \frac{1}{n} \sum_{l=1,2; \delta_i^{(c)} = l} \log f_{(y,l)}(Y_i|D_i),
\]

where \( f_{(y,s)}(\cdot) \) is given above, \( f_{(y,1)}(\cdot|D_i, a_i) \) and \( f_{(y,2)}(\cdot|D_i, a_i, b_i) \) are the densities of \( Y_i \) given \( \{D_i, a_i, \delta_i^{(c)} = 1\} \) and \( \{D_i, a_i, b_i, \delta_i^{(c)} = 2\} \), respectively, \( f_{(y,1)}(Y_i|D_i) = \int f_{(y,1)}(Y_i|D_i, a_i) (1 - F_s(T_{ini}|a_i)) f_a(a_i) da_i \) and

\[
f_{(y,2)}(Y_i|D_i) = \int \int f_{(y,2)}(Y_i|D_i, a_i, b_i) F_s(T_{ini}|a_i) f_b(b_i|a_i) f_a(a_i) da_i db_i.
\]

When \( f_{(y,s)}(\cdot) \) belongs to a parametric family \( \{f_{(y,s)}(\cdot; \phi|D_i); \phi = (\theta, V)\} \), we can estimate \( \phi = (\theta, V) \) by maximizing the log-likelihood (3.1).

Estimation for the additive model (2.4) can be achieved by an approximate likelihood method. Under some mild smoothness conditions on \( \mu_d^{(s)}(\cdot) \) (Huang and others, 2004), we can approximate \( \mu_d^{(s)}(\cdot) \) by the following B-spline expansion:

\[
\mu_d^{(s)}(a_{id}) \approx \sum_{p=1}^{P_d} \gamma_d^{(p)} B_d^{(p)}(a_{id}) = (\gamma_d^{(p)})^T B_d^{(p)}(a_{id}) \quad \text{for some } P_d,
\]
where \( \{B_p^{(d)}(\cdot) ; 1 \leq p \leq P_d\} \) is a spline basis, \( B^{(d)}(a_i) = (B_1^{(d)}(a_i), \ldots, B_{P_d}^{(d)}(a_i))^T \), and \( \gamma^{(d)} = (\gamma_1^{(d)}, \ldots, \gamma_{P_d}^{(d)})^T \) is a set of real-valued coefficients. It follows from (3.2) that \( S_i \approx \sum_{d=0}^{d_0} (\gamma^{(d)})^T B^{(d)}(a_i) + \epsilon_i^{(s)} \). If we substitute \( \sum_{d=0}^{d_0} \mu^{(d)}(a_i) \) of (2.4) with \( \sum_{d=0}^{d_0} (\gamma^{(d)})^T B^{(d)}(a_i) \), the parameters are \( \theta = (a^T, \beta^T, \ldots, \beta_{P_d}^T, \gamma^T)^T \), \( \gamma = ((\gamma^{(0)})^T, \ldots, (\gamma^{(d_0)})^T)^T \), and \( \mathbb{V} = \{V_y, V_a, V_b, \sigma_s^2\} \). Let \( f_\theta^*(\cdot ; \gamma, \sigma_s | a_i) \) be the conditional density of \( \sum_{d=0}^{d_0} (\gamma^{(d)})^T B^{(d)}(a_i) + \epsilon_i^{(s)} \) given \( a_i \). If the distributions of \( \{\epsilon_{ij}, \epsilon_i^{(a)}, \epsilon_i^{(b)}, \epsilon_i^{(c)}\} \) are parameterized by \( \mathbb{V} \), we can approximate \( f_\theta(S_i | a_i) \) by \( f_\theta^*(S_i ; \gamma, \sigma_s | a_i) \) and obtain the following approximate log-likelihood function for \( (Y_i, S_i^{(c)}) \) given \( D_i \),

\[
L_\theta^*(\phi) = \frac{1}{n} \sum_{i: \delta_i^{(c)} = 0} \log f_{(y,s)}^*(Y_i, S_i ; \phi | D_i) + \frac{1}{n} \sum_{l=1,2} \sum_{i: \delta_i^{(d)} = l} \log f_{(y,l)}^*(Y_i ; \phi | D_i),
\]

where \( \phi = (\theta, \mathbb{V}) \). \( f_{(y,s)}^*(\cdot | D_i) \), \( f_{(y,l)}^*(\cdot | D_i) \), \( k = 1, 2 \), are given in (3.1) with \( f_\theta(S_i | a_i) \) replaced by \( f_\theta^*(S_i ; \gamma, \sigma_s | a_i) \). If \( L_\theta^*(\phi) \) satisfies the regularity conditions for MLEs, the approximate MLE \( \hat{\phi} \) satisfy \( L_\theta^*(\hat{\phi}) = \max_\phi L_\theta^*(\phi) \).

### 3.2 A 2-stage estimation procedure

Since the computation of maximum likelihood (ML) or approximate ML estimators requires a nonlinear maximization over the parameter spaces of \( \theta \) and \( \mathbb{V} \), a global maximization over \( \theta \) and \( \mathbb{V} \) simultaneously could be computationally unfeasible in practice. We propose a 2-stage procedure which combines REMLE with the Newton–Raphson algorithm:

Stage 1. Assume that \( \{\epsilon_{ij}, a_i, b_i, S_i\} \) of (2.3) or (2.4) are independent random variables with covariance matrices \( \mathbb{V} = \{V_y, V_a, V_b, \sigma_s^2\} \), that is, the naïve mixed-effects model (2.1) holds. Compute \( \hat{\mathbb{V}} \) of \( \mathbb{V} \) using the REMLE procedure.

Stage 2. Substitute \( \mathbb{V} \) with \( \hat{\mathbb{V}} \) and maximize \( L_\theta(\theta, \hat{\mathbb{V}}) \) with respect to \( \theta \) using the Newton–Raphson procedure. The maximizer \( \hat{\theta} = \arg \max_\theta L_\theta(\theta, \hat{\mathbb{V}}) \) is the approximate ML estimator for \( \theta \).

By (3.1) and the expressions of \( f_{(y,s)}(Y_i, S_i | D_i) \) and \( f_{(y,l)}(Y_i | D_i) \) for \( k = 1, 2 \), the Newton–Raphson algorithm for maximizing \( L_\theta(\theta, \hat{\mathbb{V}}) \) at Stage 2 involves multidimensional integrations over the functions of \( a_i, b_i, \) and \( S_i \) with respect to the joint distributions of \( a_i \) and \( b_i \). Monte Carlo simulations are used to compute all the necessary quantities, including the log-likelihoods, and their gradients and Hessian matrices. Implementation of the Newton–Raphson algorithm can be computationally costly, since, in each iteration, large Monte Carlo samples are required to compute the gradient and the Hessian matrix. The computational complexity further increases with the number of subjects. If a suitable initial estimator is available, computation of the algorithm can be significantly reduced by a “1-step” Newton–Raphson procedure (Bickel, 1975; Cai and others, 2000). The estimators computed from the REMLE can be used as a natural candidate for the initial estimator \( \hat{\theta}_0 \). Initial estimators of \( \gamma \) can be computed by fitting the regression model \( S_i = y^T (1, \hat{a}_{i\text{pred}}) + \epsilon_i^{(c)} \) using those subjects with \( S_i \) observed (i.e. \( \epsilon_i^{(c)} = 0 \)), where \( \hat{a}_{i\text{pred}} \) is the predicted value for \( a_i \). Further details of this algorithm are provided in Appendix C of the supplementary material available at Biostatistics online.

### 3.3 Inferences

Inferences for the MLE of a parametric model (3.1) may be obtained using their asymptotic distributions. Let \( \hat{\phi} \) be the MLE of \( \phi \). The asymptotic normality of the MLE’s (Serfling, 1980, Chapter 4) implies that, when \( n \) is large, \( \hat{\phi} \) has approximately the multivariate normal distribution \( \mathcal{N}(\phi, \text{Var}(\phi)) \). An approximate
[100 × (1 − α)]th confidence interval (CI) for ℓ(ϕ), a linear combination of ϕ, is ℓ(ϕ) ± Z_{α/2} [Var(ℓ(ϕ))]^{1/2}, where Var(ℓ(ϕ)) is the variance estimator and Z_{α/2} is the [100 × (1 − α/2)]th percentile of the standard normal distribution. For the additive model (2.4), where nonparametric components are present, asymptotic distributions of the approximate MLE in (3.2) have not yet been developed. As a practical alternative, we can generate bootstrap samples by resampling the subjects with replacement and compute the corresponding bootstrap estimators. The estimates obtained from the original sample are natural choices for the bootstrap estimators. The estimates obtained from the original sample are natural choices for the bootstrap approach has been used in Wu and others (2008) among others. We investigate in Section 5 the empirical coverage probabilities (CP) of our CIs in a simulation study.

4. APPLICATION TO THE ENRICHED STUDY

4.1 The ENRICHED pharmacotherapy data

We applied our method to the ENRICHED Study and investigated whether the use of antidepressants as a pharmacotherapy could lower the BDI scores for patients during the PT period. Wu and others (2008) investigated this question by applying their varying-coefficient model to the subsample of 91 depressed patients in the ENRICHED PT arm who had their change-point time S_i observed during the treatment period and concluded that the use of antidepressants was beneficial for lowering the BDI scores for patients who were eligible for pharmacotherapy. Our analysis was based on all 557 depressed patients who had their status of antidepressant use before baseline, 92 started antidepressant during the treatment period, and attended 5 or more PT sessions during the treatment period. In order to have a meaningful interpretation for this study, we excluded 3 types of patients from our analysis: patients in the usual care arm because of the lack of BDI scores observed between the baseline and the end of the treatment period, patients whose pharmacotherapy status and starting dates of antidepressant use were not recorded, and patients who had poor adherence to the required weekly PT sessions, that is, attended less than 5 (approximately 20%) of the sessions. Because antidepressant use for each patient was individually monitored and recorded accurately as possibly by study psychiatrists (Taylor and others, 2005, p. 794), it is reasonable to assume that the missing records on antidepressant starting dates were missing at random. Within our sample, 11 patients used antidepressants before baseline, 92 started antidepressant during the treatment period, and 454 did not use antidepressants before and during the treatment period. The number of visits for these patients ranges from 5 to 36 and has a median of 12.

Let Y_{ij}, T_{ij}, S_{i}^{(c)} , and R_{ij} = T_{ij} − S_{i}^{(c)} be the ith patient’s BDI score, trial time (months), starting time (months) of antidepressant use, and antidepressant duration time (months), respectively, at the jth visit. For all 1 ≤ i ≤ n, T_{i1} = 0, T_{in_i} is the trial time at the last visit, and \lambda_{ij} = 1[S_i < T_{ij}] is the indicator of antidepressants use at T_{ij}. The observed (S_{i}^{(c)}, \delta_{i}^{(c)}) is (S_{i}^{(c)} = S_i, \delta_{i}^{(c)} = 0) if the ith patient used antidepressants within the treatment period, (S_{i}^{(c)} = T_{in_i}, \delta_{i}^{(c)} = 1) if the patient did not use antidepressant within the treatment period, and (S_i = 0, \delta_{i}^{(c)} = 2) if the patient used antidepressants before baseline. When the linear models \mu_0(T_{ij}; a_{i0}, a_{i1}) = a_{i0} + a_{i1} T_{ij} and \mu_1(R_{ij}; b_{i0}, b_{i1}) = b_{i0} + b_{i1} R_{ij} are used, (a_{i0}, a_{i1}) represents the intercept and slope of the ith subject’s BDI trajectory before antidepressant use, and (b_{i0}, b_{i1}) is the intercept and slope of the change of the subject’s BDI trajectory after antidepressant use.

4.2 Fitting the naive mixed-effects model

In order to have a clear and parsimonious interpretation of the overall trend of BDI over the trial time, we examined some preliminary analysis of the data and approximated the relationship between Y_{ij} and
\( T_{ij} \) by a simple linear model. Using the naïve mixed-effects model (2.1), we first estimated the unknown parameters \( \alpha_0, \alpha_1, \beta_{00}, \) and \( \beta_{10} \) of

\[
\begin{align*}
Y_{ij} &= a_{0i} + a_{1i}T_{ij} + \lambda_{ij}(b_{0i} + b_{1i}R_{ij}) + \epsilon_{ij}, \\
(a_{0i}, a_{1i}, b_{0i}, b_{1i})^T &\sim N((\alpha_0, \alpha_1, \beta_{00}, \beta_{10})^T, \Sigma),
\end{align*}
\tag{4.1}
\]

where \((\alpha_0, \alpha_1, \beta_{00}, \beta_{10})^T\) is the unknown mean vector and \(\Sigma\) is the unstructured covariance matrix for the multivariate normal distribution \(N(\cdot, \cdot)\). Intuitively, \(\alpha_0\) and \(\alpha_1\) are the mean intercept and slope for BDI scores before antidepressant use, and \(\beta_{00}\) and \(\beta_{10}\) are the mean intercept and slope for the “correction term” after antidepressant use. Zero value of \(\beta_{00} + \beta_{10}R_{ij}\) indicates ignorable antidepressant effect on the mean BDI scores.

The left panel of Table 1 shows the estimates of \(\alpha_0, \alpha_1, \beta_{00}, \) and \(\beta_{10}, \) and their standard errors (SEs) and 95% CIs computed using REMLE with unstructured covariance matrix \(\Sigma\). The negative estimate \(\hat{\alpha}_1 = -1.887\) suggests that the mean BDI score for these patients tends to decrease over the trial time since the start of the PT sessions, while the positive estimate \(\hat{\beta}_{00} = 3.579\) and its 95% CI seem to suggest that the use of antidepressants increase the patients’ mean BDI scores. As discussed in Section 2.2, the self-selectiveness of the antidepressant use as a concomitant intervention is not considered in (4.1), so that the positive estimate of \(\beta_{00}\) may not reflect the real effect of pharmacotherapy on depression severity.

### 4.3 Fitting the shared parameter model

To account for the possible link between antidepressant starting time and the BDI trend before the use of antidepressants, we examined a series of preliminary plots among \(S_i, (a_{0i}, a_{1i})\) and \((b_{0i}, b_{1i})\) using patients in the PT arm who had started using antidepressants during the treatment period. These preliminary results (see Appendix D.1 of the supplementary material available at Biostatistics online) suggested that the following linear shared parameter model could give a parsimonious approximation to the variables considered:

\[
\begin{align*}
Y_{ij} &= a_{0i} + a_{1i}T_{ij} + \lambda_{ij}(b_{0i} + b_{1i}R_{ij}) + \epsilon_{ij}, \\
(b_{0i}, b_{1i})^T &= (\alpha_0, \alpha_1)^T + \epsilon^{(a)}_i, \\
S_i &= \gamma_0 + \gamma_1a_{1i} + \epsilon^{(s)}_i,
\end{align*}
\tag{4.2}
\]

where \(\epsilon^{(a)}_i\) and \(\epsilon^{(b)}_i = (\epsilon^{(b)}_{i0}, \epsilon^{(b)}_{i1})^T\) are mean-zero bivariate normal random vectors with unstructured covariance matrices \(\Sigma^{(a)}\) and \(\Sigma^{(b)}\), respectively. \(\epsilon^{(s)}_i\) is a mean-zero normal random variable with variance \(\sigma^2_s\), and \(\epsilon^{(a)}_i, \epsilon^{(b)}_i, \) and \(\epsilon^{(s)}_i\) are independent. Similar to (4.1), \((\alpha_0, \alpha_1)\) represents the mean intercept and slope for the BDI trend over the trial time before antidepressant use. Given \((a_{0i}, a_{1i})^T, (\beta_{00}, \beta_{01}, \beta_{02}, \beta_{10}, \beta_{11}, \beta_{12})^T\) specifies the conditional mean structure for the BDI correction term after antidepressant use.

The middle panel of Table 1 shows the estimates of the coefficients in (4.2) computed using the 2-stage ML procedure with 10 Newton–Raphson iterations, and their SE and 95% percentile CIs computed from the bootstrap procedure. Since the 95% CIs for \(\beta_{01}, \beta_{02}, \beta_{10}, \) and \(\beta_{11}\) in (4.2) include zero, we repeated the estimation procedure for the submodel of (4.2) with \(\beta_{01}, \beta_{02}, \beta_{10}, \) and \(\beta_{11}\) set to zero and summarized the estimates, their SEs, and 95% percentile CIs from this submodel in the right panel of Table 1, which gives \(\hat{\beta}_{00} = -6.707\) and \(\hat{\beta}_{12} = -0.216\). Contrary to the interpretation from (4.1), we obtain from this submodel that \(\hat{\beta}_{00} + \hat{\beta}_{12}\hat{a}_1R_{ij} < 0\) since \(0 \leq R_{ij} \leq 6\) and \(\hat{a}_1 = -1.972\), which suggests that pharmacotherapy has on average a beneficial effect for lowering a patient’s depression severity. The
Table 1. Parameter estimates, SEs and 95% CIs computed for the ENRICHD pharmacotherapy data based on the naïve linear mixed-effects model (4.1), the linear shared parameter model (4.2) and the submodel of (4.2) with $\beta_{01}$, $\beta_{02}$, $\beta_{10}$, and $\beta_{11}$ set to zero.

<table>
<thead>
<tr>
<th>Effect parameter</th>
<th>Linear mixed-effects model</th>
<th>Linear shared parameter model</th>
<th>Linear shared parameter submodel</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SE</td>
<td>95% CI</td>
</tr>
<tr>
<td>$a_0$</td>
<td>14.454</td>
<td>0.312</td>
<td>(13.842, 15.066)</td>
</tr>
<tr>
<td>$a_1$</td>
<td>-1.887</td>
<td>0.067</td>
<td>(-2.018, -1.756)</td>
</tr>
<tr>
<td>$\beta_{00}$</td>
<td>3.579</td>
<td>0.825</td>
<td>(1.962, 5.196)</td>
</tr>
<tr>
<td>$\beta_{01}$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$\beta_{02}$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$\beta_{10}$</td>
<td>0.036</td>
<td>0.227</td>
<td>(-0.409, 0.481)</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
95% CIs for $\gamma_1$ obtained from (4.2) and its submodel suggests that (4.1) is likely a misspecified model for this data set. These results are consistent with the analysis of Wu and others (2008).

5. Simulation Study

5.1 Data generation

The simulation study has 2 objectives: (i) to demonstrate the biases of the naïve mixed-effects models; and (ii) to investigate the practical properties of our ML and approximate MLE procedures. In order to resemble a moderate sized longitudinal study with data structures similar to the ENRICHD pharmatherapy data, each of our simulated samples contains $n = 200$ or 300 independent subjects who have unequal number of repeated measurements. For $i = 1, \ldots, n$, the $i$th subject in the sample has a maximum of 30 “scheduled visits” at time points $\{T_{i,1}, \ldots, T_{i,30}\} = [0, 0.2 + e_1, \ldots, 5.8 + e_29]$, where $\{e_i\}$ are independent “distortions” generated from the uniform $U(-0.1, 0.1)$ distribution, but each scheduled visit has 40% probability to be skipped. The remaining observed visiting times are $\{T_{ij}; \; j = 1, \ldots, n_i\}$ with possibly unequal $n_i$ for different subjects. For each $1 \leq i \leq n$, the random parameter $\mathbf{a}_i = (a_{i0}, a_{i1})^T$ is generated by adding $(a_0, a_1)^T = (20, 10)^T$ to a randomly generated $\epsilon^{(a)}_i$ from the zero-mean bivariate normal distribution with covariance matrix diag(2.5, 1), and $\mathbf{b}_i = (b_{i0}, b_{i1})^T$ is generated by adding the conditional mean $(\beta_{00} + \beta_{01} a_{i0}, \beta_{12} a_{i1})^T$ with $(\beta_{00}, \beta_{01}, \beta_{12})^T = (-4, 0.01, 0.02)^T$ to a randomly generated $\epsilon^{(b)}_i$ from the standard bivariate normal distribution. For each $\{a_i^T, b_i^T\}$, we generate $S_i$ from 2 different scenarios: (a) $S_i \sim N(\gamma_1 a_{i0} + \gamma_2 a_{i1}, 1)$ with $\gamma = (\gamma_1, \gamma_2)^T = (1, -1.7)^T$; (b) $S_i \sim N(\mu_0^{(s)}(a_{i0}), 0.04)$, where $\mu_0^{(s)}(a_{i0}) = -0.75(a_{i0} - 18)^2 + 6$ if $a_{i0} < 20$, and $0.75(a_{i0} - 22)^2$ if $a_{i0} > 20$. The purpose of (b) is to evaluate the B-spline estimate when the parametric form of $\mu_0^{(s)}(a_{i0})$ is unknown. For each $\{T_{ij}, S_i, a_i, b_i\}$, we generate $Y_{ij}$ from the subject-specific change-point model

$$Y_{ij} = a_{i0} + a_{i1} T_{ij} + \lambda_{ij}(b_{i0} + b_{i1} R_{ij}) + \epsilon_{ij},$$

(5.1)

where $\lambda_{ij} = 1_{[T_{ij} > S_i]}$, $R_{ij} = \max(T_{ij} - S_i, 0)$, and $\epsilon_{ij} \sim N(0, 1)$. When $S_i < T_{i1}$ or $S_i > T_{in}$, we use censored $S_i$ and obtain $\{Y_{ij}, T_{ij}, \lambda_{ij}, S_i^{(c)}, \delta_i^{(c)}; \; j = 1, \ldots, n_i\}$.

The simulation was separately repeated 200 times under scenarios (a) and (b) with $n = 200$ and 300. Within each simulation, we used $\mu_0(T_{ij}; a_i) = a_{i0} + a_{i1} T_{ij}$ and $\mu_1(T_{ij}, R_{ij}; b_i) = b_{i0} + b_{i1} R_{ij}$, and estimated the parameters $\alpha = (a_{00}, a_{11})^T$ and $\beta = (\beta_{00}, \beta_{01}, \beta_{12})^T$ using the naïve mixed-effects model (2.1) and the shared parameter model (2.2). For estimation under the naïve mixed-effects model (2.1), we ignored the relationship between $a_i$ and $S_i$ and computed the restricted ML estimators $(\hat{\alpha}^T, \hat{\beta}^T)^T$.

5.2 Estimation for linear shared parameter models

For the Gaussian linear shared parameter model (2.3) with $S_i$ generated from (a), we computed the MLE $(\hat{\alpha}^T, \hat{\beta}^T, \hat{\gamma}^T)^T$ under (3.1) using the 2-stage estimation procedure with diagonal estimated covariance matrices and 1-step Newton–Raphson iteration. The REML estimators $\tilde{\alpha}$ and $\tilde{\beta}$ were used as the initial estimators of $\alpha$ and $\beta$ for the Newton–Raphson iteration. For the initial estimator $\hat{\gamma}$ of $\gamma$, we used the subjects with $S_i$ observed (i.e. $\delta_i^{(c)} = 0$), computed the predicted values $(\hat{a}_{i0}^{\text{pred}}, \hat{a}_{i1}^{\text{pred}})^T$ and calculated $\hat{\gamma}$ by the least squares estimate from the linear model $S_i = \gamma_1 a_{i0}^{\text{pred}} + \gamma_2 a_{i1}^{\text{pred}} + \epsilon_i^{(s)}$.

Table 2 summarizes the averages of the estimates and their SEs, the mean squared errors (MSEs), and the empirical CP for the 95% bootstrap CIs based on the simulated samples with $S_i$ generated from (a). In each simulated sample, the SE and CP are computed using $B = 200$ bootstrap samples. The estimates of $\alpha$ and $\beta$ obtained under the naïve mixed-effects model (2.1) have significantly larger biases and MSEs than
Table 2. True parameter values, averages of the parameter estimates and their SEs, the MSEs, and the empirical CPs of the 95% CIs computed over the simulated samples generated from the Gaussian linear shared parameter model (2.3)

<table>
<thead>
<tr>
<th>Model parameter</th>
<th>Naïve linear mixed-effects model</th>
<th>Gaussian linear shared parameter model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SE</td>
</tr>
<tr>
<td>( n = 200 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_0 = 20 )</td>
<td>20.184</td>
<td>0.117</td>
</tr>
<tr>
<td>( \alpha_1 = 10 )</td>
<td>9.747</td>
<td>0.085</td>
</tr>
<tr>
<td>( \beta_{00} = -4 )</td>
<td>-6.462</td>
<td>1.862</td>
</tr>
<tr>
<td>( \beta_{01} = 0.01 )</td>
<td>0.127</td>
<td>0.094</td>
</tr>
<tr>
<td>( \beta_{12} = 0.02 )</td>
<td>0.064</td>
<td>0.014</td>
</tr>
<tr>
<td>( \gamma_1 = 1 )</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( \gamma_2 = -1.7 )</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( n = 300 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_0 = 20 )</td>
<td>20.177</td>
<td>0.095</td>
</tr>
<tr>
<td>( \alpha_1 = 10 )</td>
<td>9.751</td>
<td>0.069</td>
</tr>
<tr>
<td>( \beta_{00} = -4 )</td>
<td>-6.506</td>
<td>1.501</td>
</tr>
<tr>
<td>( \beta_{01} = 0.01 )</td>
<td>0.130</td>
<td>0.075</td>
</tr>
<tr>
<td>( \beta_{12} = 0.02 )</td>
<td>0.063</td>
<td>0.011</td>
</tr>
<tr>
<td>( \gamma_1 = 1 )</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( \gamma_2 = -1.7 )</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

The estimates obtained under the linear shared parameter model (2.3). The 95% bootstrap CIs constructed under (2.1) have low empirical CPs, while the CPs for (2.3) are close to the nominal level of 95%. These results suggest that, because of the potential model misspecification and estimation bias, the naïve mixed-effects models may lead to erroneous conclusions when a concomitant intervention is present, while the estimation bias can be dramatically reduced by a shared parameter model.

5.3 Estimation for additive shared parameter model

For samples with \( S_i \) generated from (b), we used the additive shared parameter model (2.4) with \( S_i = \mu_0^{(s)}(a_{i0}) + \epsilon_i^{(s)} \) and Gaussian errors. We approximated \( \mu_0^{(s)}(a_{i0}) \) using the quadratic B-splines with one interior knot at \( a_{i0} = 20 \) and estimated \( (\alpha^T, \beta^T)^T \) using the approximate likelihood (3.2) and the 2-stage estimation procedure with diagonal estimated covariance matrices and 1-step Newton–Raphson iteration. The estimated curve \( \mu_0^{(s)}(a_{i0}) \) was computed by the B-splines with the estimated spline coefficients. We computed the SEs and the 95% bootstrap CIs for \( (\alpha^T, \beta^T)^T \) using the “resampling-subject” bootstrap procedure with \( B = 200 \) replications. Pointwise 95% CIs for \( \mu_0^{(s)}(a_{i0}) \) were also computed using the same bootstrap procedure.

Table 3 summarizes the averages of the estimates and their SEs, the MSEs and the empirical CPs for the 95% bootstrap CIs based on the 200 simulated samples with \( S_i \) generated from (b). Similar to Table 2, the estimates obtained under the naïve mixed-effects model (2.1) have mostly larger MSEs than the estimates obtained under the additive shared parameter model (2.4), and the CIs obtained under (2.1) have lower empirical CPs than those obtained under (2.4). The left panel of Figure 1 shows the true curve \( \mu_0^{(s)}(a_{i0}) \), and the spline estimated curve for \( \mu_0^{(s)}(a_{i0}) \) together with its 95% bootstrap CIs at 600 equally spaced \( a_{i0} \) points within \([17, 23]\) computed from one randomly selected simulated sample. The right panel of Figure 1 shows the empirical CPs of these pointwise 95% bootstrap CIs computed over all
Table 3. True parameter values, averages of the parameter estimates and their SEs, the MSEs, and the empirical CPs of the 95% CIs computed over the simulated samples generated from the additive shared parameter model (2.4)

<table>
<thead>
<tr>
<th>Model parameter</th>
<th>Naive linear mixed-effects model</th>
<th>Additive shared parameter model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SE</td>
</tr>
<tr>
<td>$n = 200$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_0 = 20$</td>
<td>19.920</td>
<td>0.120</td>
</tr>
<tr>
<td>$a_1 = 10$</td>
<td>9.991</td>
<td>0.092</td>
</tr>
<tr>
<td>$\beta_{00} = -4$</td>
<td>-4.850</td>
<td>1.680</td>
</tr>
<tr>
<td>$\beta_{01} = 0.01$</td>
<td>0.061</td>
<td>0.083</td>
</tr>
<tr>
<td>$\beta_{12} = 0.02$</td>
<td>0.018</td>
<td>0.012</td>
</tr>
<tr>
<td>$n = 300$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_0 = 20$</td>
<td>19.916</td>
<td>0.098</td>
</tr>
<tr>
<td>$a_1 = 10$</td>
<td>9.993</td>
<td>0.074</td>
</tr>
<tr>
<td>$\beta_{00} = -4$</td>
<td>-4.844</td>
<td>1.363</td>
</tr>
<tr>
<td>$\beta_{01} = 0.01$</td>
<td>0.061</td>
<td>0.068</td>
</tr>
<tr>
<td>$\beta_{12} = 0.02$</td>
<td>0.017</td>
<td>0.010</td>
</tr>
</tbody>
</table>

Fig. 1. Left Panel: The thick solid line shows the true curve for $\mu_0^{(s)}(a_{i0})$. The thin solid and dashed lines are the spline estimated curve and the 95% pointwise bootstrap CIs obtained from one simulated sample. Right Panel: The solid line gives the empirical CPs of the 95% bootstrap CIs at 600 equally spaced $a_{i0}$ points within [17, 23] for the spline estimated curve of $\mu_0^{(s)}(a_{i0})$, and the dashed line shows the average of the 600 CPs. Each simulated sample has $n = 200$.

the simulated samples with $n = 200$, and the average of these CPs over the 600 equally spaced $a_{i0}$ points within [17, 23]. The spline estimated curve appears to be reasonably close to the true curve $\mu_0^{(s)}(a_{i0})$, and the CPs of the pointwise 95% bootstrap CIs are generally higher than the nominal level of 95% in the interior of the range of $a_{i0}$.
6. Discussion

Our approach suggests that concomitant interventions should not be treated the same way as a usual time-dependent covariate, and a shared parameter model may reduce the estimation bias by incorporating the self-selectiveness of the intervention. A number of extensions are worthwhile for further investigations. First, a patient may have multiple concomitant interventions in practice, so that a multiple change-point extension of (2.2) would be needed in such situations. Second, a worthy future development is to incorporate missing data and multiple concomitant interventions into a single framework of joint models. Third, asymptotic distributions of our estimators may provide useful insights and should be developed. Finally, since it may not be always clear whether an intervention is a concomitant intervention, model diagnostic methods for evaluating the appropriateness of a shared parameter model would be a valuable tool to be developed.

Supplementary material

Supplementary material is available at http://biostatistics.oxfordjournals.org.

Acknowledgments

We would like to thank Drs Myron Waclawiw and Minjung Kwak, and the reviewers for their thoughtful comments. Conflict of Interest: None declared.

Funding

Financial support of the ENRICHD study was provided by contracts N01-HC-55140, N01-HC-55141, N01-HC-55142, N01-HC-55143, N01-HC-55144, N01-HC-55145, N01-HC-55146, N01-HC-55147, and N01-HC-55148 from the National Heart, Lung and Blood Institute.

References


A shared parameter concomitant intervention


[Received February 27, 2010; revised December 17, 2010; accepted for publication December 18, 2010]