



Varying-coefficient models for longitudinal processes with continuous-time informative dropout

LI SU*

*MRC Biostatistics Unit, Institute of Public Health, University Forvie Site,
Robinson Way, Cambridge CB2 0SR, UK
li.su@mrc-bsu.cam.ac.uk*

JOSEPH W. HOGAN

*Center for Statistical Sciences, Department of Community Health, Box G-S121-7,
Brown University, Providence, RI 02912, USA*

SUMMARY

Dropout is a common occurrence in longitudinal studies. Building upon the pattern-mixture modeling approach within the Bayesian paradigm, we propose a general framework of varying-coefficient models for longitudinal data with informative dropout, where measurement times can be irregular and dropout can occur at any point in continuous time (not just at observation times) together with administrative censoring. Specifically, we assume that the longitudinal outcome process depends on the dropout process through its model parameters. The unconditional distribution of the repeated measures is a mixture over the dropout (administrative censoring) time distribution, and the continuous dropout time distribution with administrative censoring is left completely unspecified. We use Markov chain Monte Carlo to sample from the posterior distribution of the repeated measures given the dropout (administrative censoring) times; Bayesian bootstrapping on the observed dropout (administrative censoring) times is carried out to obtain marginal covariate effects. We illustrate the proposed framework using data from a longitudinal study of depression in HIV-infected women; the strategy for sensitivity analysis on unverifiable assumption is also demonstrated.

Keywords: HIV/AIDS; Missing data; Nonparametric regression; Penalized splines.

1. INTRODUCTION

Many longitudinal studies suffer from dropout, which is termed “informative” if the dropout process depends on the unobserved outcomes even after conditioning on the observed data. To account for informative dropout, a number of model-based approaches have been proposed for the joint modeling of the dropout and longitudinal outcome processes (Little, 1995; Hogan and Laird, 1997b; Kenward and Molenberghs, 1999). These approaches can be generally classified as “selection models” (Wu and Carroll, 1988; Diggle and Kenward, 1994; Follman and Wu, 1995; Ten Have *and others*, 1998), “pattern-mixture

*To whom correspondence should be addressed.

models” (PMMs) (Wu and Bailey, 1989; Little, 1993, 1994; Hogan and Laird, 1997a), or “shared parameter models” (Wulfsohn and Tsiatis, 1997; Henderson *and others*, 2000; Tsiatis and Davidian, 2004). In all these frameworks, inference is based on the assumptions that are not verifiable from the observed data, and tools for evaluating the sensitivity to these assumptions are required (Rotnitzky *and others*, 2001; Verbeke *and others*, 2001; Molenberghs *and others*, 2003; Lee, 2007).

Focusing on the pattern-mixture modeling approach, in this article we develop a general framework of varying-coefficient models (VCMs) (Hastie and Tibshirani, 1993) for longitudinal data, where measurement times may be irregular across individuals and where dropout can occur at any point in continuous time (not just at observation times) and be unobserved due to administrative censoring. Specifically, the conditional distribution of the longitudinal outcomes given the dropout time (or the administrative censoring time) follows a VCM, where the outcome model parameters such as regression coefficients, variance components, and correlation parameters depend on the dropout time (or the administrative censoring time) through unspecified smooth functions. Two separate VCMs are specified to distinguish the administratively censored individuals from those who actually drop out. The full-data distribution is a mixture over the dropout/administrative censoring time distribution, which is left unspecified.

The proposed framework generalizes the pattern-mixture models (Little, 1993, 1994; Fitzmaurice and Laird, 2000), the conditional linear models (CLMs) (Wu and Bailey, 1989), and the class of VCMs developed for continuous outcomes in Hogan *and others* (2004). Specifically, our approach is distinguished from the work in Hogan *and others* (2004) by (a) handling administrative censoring separately from other types of dropout that could be related to the outcome process, (b) allowing all model parameters to depend on the dropout process, and (c) accommodating binary outcomes. In this article, we demonstrate the proposed approach using both continuous and binary longitudinal data with continuous-time dropout. The unspecified smooth functions are modeled by Bayesian penalized splines (Ruppert *and others*, 2003). When the marginal covariate effects on the outcome process are of interest, Rubin’s (1981) Bayesian bootstrap (BB) is used for averaging over the dropout time distribution with administrative censoring. The advantage of building our VCM framework within the Bayesian paradigm is that there is no need to model the continuous dropout time distribution parametrically. With a frequentist approach to model the dropout times nonparametrically, extra simulation by bootstrapping the continuous dropout times is necessary for standard error estimation if the delta method fails (Hogan *and others*, 2004). On the other hand, the BB is naturally merged with the Markov chain Monte Carlo (MCMC) for the outcome process model, and the variability of the observed dropout/administrative censoring times is appropriately taken into account when making inferences on the marginal covariate effects.

The HIV epidemiology research study (HERS) (Smith *and others*, 1997; Ickovics *and others*, 2001) was a longitudinal study of women with, or at high risk for, HIV infection. Twelve core visits (each in a calendar time window) were scheduled for 1310 women, where a variety of clinical, behavioral, and sociologic outcomes were to be recorded approximately every 6 months. If women came to the sites on a date out of the visit window, the visit procedures were not performed. Further, mid-interval visits were added for severely immunosuppressed women (CD4 count < 100). The actual measurement times correspond to assessment dates and vary across participants.

Our interest is in studying the course of depression in the 753 women who had HIV infection at baseline and did not drop out of the study due to HIV-related death before the study end. Depression was measured using the Center for Epidemiologic Studies Depression Scale (CES-D). The CES-D includes 20 questions related to mood, each of which can take a value from 0 (symptom rarely present) to 3 (symptom almost always present); scores therefore range from 0 to 60. Because the distribution of CES-D for a general population can be very skewed, in practice transformations or nonparametric methods need to be applied (Radloff, 1977). In HIV research, a score of 16 or greater for CES-D is frequently used as a cutoff for clinical depression (Ickovics *and others*, 2001; Cook *and others*, 2004; Leserman, 2008). This can avoid the potential nonnormality problem for continuous CES-D data and can be useful

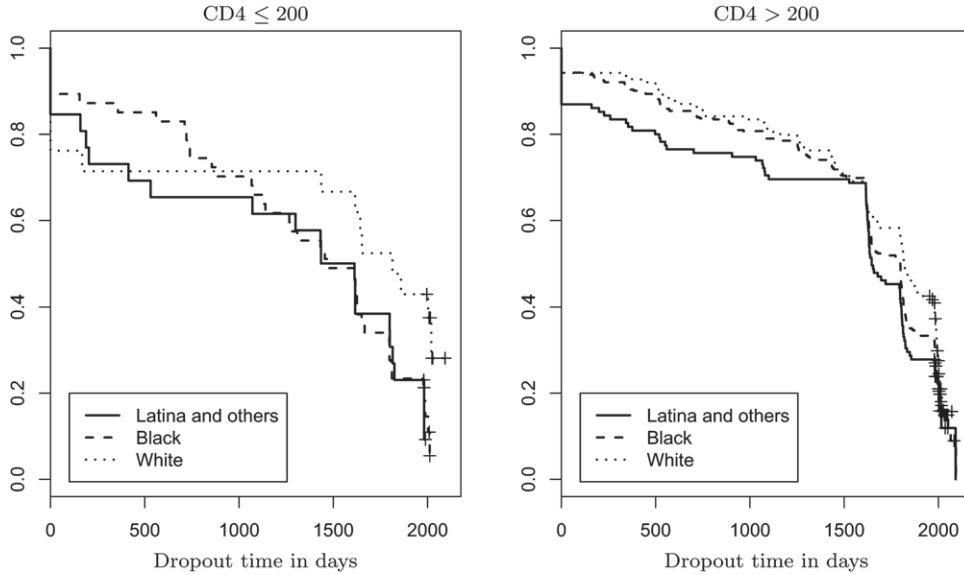


Fig. 1. Kaplan–Meier curves of the dropout time by the race and baseline CD4 count groups in the HERS application. Cross represents right-censored dropout time due to administrative reasons.

for screening depression cases in study populations. As in the original analysis of the HERS CES-D data presented in Ickovics *and others* (2001), our proposed VCMs were originally motivated by the analysis of the dichotomized HERS CES-D data. However, we will also illustrate our framework using continuous HERS CES-D data and compare the results from the 2 analyses.

A challenge with the analysis of these HERS depression data is that dropout could be related to the disease progression and associated depressive symptoms. Figure 1 presents the Kaplan–Meier curves for the dropout time by race and baseline CD4 count. Only 173 women finished the 12 scheduled visits, and their dropout times are treated as administratively censored. We distinguish these women from those who dropped out prematurely due to reasons other than HIV-related death.

The remainder of this article is organized as follows. The proposed modeling framework is described in Section 2. Estimation procedures are detailed in Section 3. In Section 4, we apply our methods to the HERS depression data. Conclusions and discussion follow in Section 5.

2. VCMs FOR INFORMATIVE DROPOUT

Suppose that the data come from N individuals, and for the i th ($i = 1, \dots, N$) individual, there is an outcome process $\{Y_i(t)\}$, where t ($t \geq 0$) is the time since enrollment. Correspondingly, there is a p -dimensional covariate process $\{\mathbf{x}_i(t)\}$ associated with $\{Y_i(t)\}$. In the absence of dropout, the conditional distribution of the variable $Y_i(t)$ given $\mathbf{x}_i(t)$ can be described by a model F with parameters $\boldsymbol{\theta}$,

$$\{Y_i(t) \mid \mathbf{x}_i(t)\} \sim F\{\boldsymbol{\theta}; \mathbf{x}_i(t)\}.$$

For example, if $Y_i(t)$ is continuous, we might assume that $F\{\boldsymbol{\theta}; \mathbf{x}_i(t)\}$ is a Gaussian process with mean function $\mu_i^M(t) = \mathbf{x}_i(t)\boldsymbol{\beta}$ and variance–covariance function $\text{cov}\{Y_i(s), Y_i(t) \mid \mathbf{x}_i(s), \mathbf{x}_i(t)\} = V_i(s, t)$ ($s \leq t$), where $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients. Parametric forms can be used for $V_i(s, t)$, for instance, $V_i(s, t) = \sigma^2 \exp(-\gamma |t - s|)$. In this case, $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2, \gamma)^T$.

If $\{Y_i(t)\}$ is a binary process, we might assume that $F\{\boldsymbol{\theta}; \mathbf{x}_i(t)\}$ follows a marginalized transition model (MTM) (Heagerty, 2002), with marginal mean $g\{\mu_i^M(t)\} = \mathbf{x}_i(t)\boldsymbol{\beta}$, where $g(\cdot)$ is a link function. The serial dependence is modeled by the conditional mean of $Y_i(t)$ given its history $\mathcal{H}_i^y(t-)$ before t and the covariate process history $\mathcal{H}_i^x(t)$, that is, $\mu_i^C(t) = E\{Y_i(t) \mid \mathcal{H}_i^y(t-), \mathcal{H}_i^x(t); \boldsymbol{\phi}\}$. Here, $\boldsymbol{\phi}$ is the dependence parameter vector and $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\phi}^T)^T$.

For the i th individual, let T_i denote the administrative censoring time (or the scheduled study end) and let D_i denote the dropout time. The observed data consist of the total follow-up time, $U_i = \min(D_i, T_i)$, and the indicator for dropout, $\delta_i = I(D_i < T_i)$. In other words, for individuals who are administratively censored (or finish the study), D_i is right censored and $\delta_i = 0$; for individuals who drop out prematurely, D_i is observed. At the time points t_{i1}, \dots, t_{in_i} ($t_{in_i} \leq U_i$), we also observe the outcome measurements $\mathbf{Y}_i = \{Y_i(t_{i1}), \dots, Y_i(t_{in_i})\}^T$.

When $D_i \geq T_i$ for all i , it is not necessary to consider the dropout process while modeling the outcome process $\{Y_i(t)\}$. Otherwise, the dropout process is potentially informative. To deal with this situation, we assume that the full data comprise $\{Y_i(t), \mathbf{x}_i(t), U_i, \delta_i\}$ and factor the joint distribution as

$$f(\mathbf{y} \mid \mathbf{x}, u, \delta) f(u, \delta \mid \mathbf{x}).$$

To induce the dependence between \mathbf{y} and (u, δ) in the first factor, we assume that

$$\{Y_i(t) \mid \mathbf{x}_i(t), u_i, \delta_i\} \sim F\{\boldsymbol{\theta}(u, \delta); \mathbf{x}_i(t)\},$$

where F is an appropriate outcome model and

$$\boldsymbol{\theta}(u, \delta) = \begin{cases} \boldsymbol{\theta}_1(u) & \text{if } \delta = 1, \\ \boldsymbol{\theta}_0(u) & \text{if } \delta = 0. \end{cases}$$

Here, $\boldsymbol{\theta}_1(\cdot)$ and $\boldsymbol{\theta}_0(\cdot)$ are the vectors of functions for the dropout time D_i and the administrative censoring time T_i . Therefore, the administratively censored individuals are distinguished from those who drop out by allowing them to have different model parameters for the outcome process. The second factor $f(u, \delta \mid \mathbf{x})$ can be specified using any distribution for event times, where the dependence on \mathbf{x} can be checked using standard event time regression analysis methods. In the HERS application, the covariates are discrete and we allow $f(u, \delta \mid \mathbf{x})$ to be completely unspecified within the levels of \mathbf{x} .

Different assumptions can be made for the form of $\boldsymbol{\theta}(u, \delta)$. For example, if $\boldsymbol{\theta}(u, \delta)$ are constant in u , then the dropout process is ignorable and methods for modeling $Y_i(t)$ given $\mathbf{x}_i(t)$ can be used without explicitly considering (U_i, δ_i) . When the dropout/administrative censoring times and the values of $\boldsymbol{\theta}(u, \delta)$ are discrete, we have a PMM (Little, 1993; Fitzmaurice and Laird, 2000). When the dropout/administrative censoring times are continuous and $\boldsymbol{\theta}(u, \delta)$ are polynomial functions, we have the CLM (Wu and Bailey, 1989). The VCMs by Hogan *and others* (2004) generalized the CLM for continuous outcome data by allowing the mean parameters to be unspecified smooth functions. Unlike in Hogan *and others* (2004), our approach handles the administrative censoring differently from other outcome-related dropout. For example, in the HERS analysis reported in Section 4, we assume that $\boldsymbol{\theta}_1(u)$ are unspecified smooth functions for the dropped-out individuals and $\boldsymbol{\theta}_0(u)$ are constants for the administratively censored individuals. Furthermore, we extend the work in Hogan *and others* (2004) by allowing all model parameters to depend on u and accommodating binary outcomes.

In a linear mixed model (LMM) for the HERS depression data introduced in Section 1, ‘‘missingness at random’’ (MAR) is assumed such that the conditional distribution of missing CES-D scores given the observed ones for those who remained in the study at u is the same as the corresponding conditional distribution for those who left the study at u (Molenberghs *and others*, 1998), that is,

$$f\{y(t_j) \mid y(t_1), \dots, y(t_{j-1}), \mathbf{x}, t_{j-1} \leq u < t_j\} = f\{y(t_j) \mid y(t_1), \dots, y(t_{j-1}), \mathbf{x}, t_j \leq u\}.$$

For our VCM, if the marginal distribution of $\{Y_i(t) \mid \mathbf{x}_i(t)\}$ is of interest, we assume that conditional on u and the covariates, the outcome distribution after u can be characterized by the same parameters in the distribution for the observed data. For example, the time trend of the CES-D score estimated from the observed data can be extrapolated for the missing CES-D scores beyond u up to the study end.

Neither the assumption in the LMM nor the one in the VCM can be verified from the observed data. One advantage of the pattern-mixture modeling approach is that the extrapolation of the missing data is transparent, which makes the substantive critique and empirical sensitivity analysis relatively straightforward (Little and Wang, 1996; Daniels and Hogan, 2000; Rotnitzky *and others*, 2001). For example, in the HERS analysis reported in Section 4, we can assume a different time slope for the CES-D scores beyond u . The sensitivity parameters would be the difference between the time slopes before u and beyond u , which cannot be identified by the observed data. Then, we can recompute the quantities of interest (such as the marginal CES-D profiles) to check their sensitivity to the nonidentifiable parameters. Because the unidentifiable part of the model is distinguished from those identifiable from the observed data, in the VCM the inferences based on the observed data remain the same regardless of the sensitivity parameters.

2.1 A model for continuous longitudinal data

Hogan *and others* (2004) developed varying-coefficient LMMs for continuous longitudinal data, where the mean parameters were allowed to depend on the dropout time, but the variance components were constants. In addition, they did not distinguish administrative censoring from other outcome-related dropouts. We generalize their model by allowing variance-component parameters to vary by the dropout/administrative censoring times.

Recall that for the i th subject, \mathbf{Y}_i is an $n_i \times 1$ continuous outcome vector, $U_i = u$ is the observed dropout/administrative censoring time, and $\delta_i = 0, 1$ is the indicator for dropout. Let $\mathbf{x}_i = \{\mathbf{x}_i(t_{i1})^T, \dots, \mathbf{x}_i(t_{in_i})^T\}^T$ be the $n_i \times p$ exogenous covariate matrix associated with the fixed effects and $\mathbf{z}_i = \{\mathbf{z}_i(t_{i1})^T, \dots, \mathbf{z}_i(t_{in_i})^T\}^T$ be an $n_i \times q$ covariate matrix associated with the random effects. Conditional on (U_i, δ_i) , we assume that

$$\begin{aligned} (\mathbf{Y}_i \mid \mathbf{x}_i, \mathbf{z}_i, \mathbf{b}_i, U_i, \delta_i) &\sim N[\mathbf{x}_i \boldsymbol{\beta}_{\delta_i}(u) + \mathbf{z}_i \mathbf{b}_i, \mathbf{R}_i \{\boldsymbol{\phi}_{\delta_i}(u)\}], \\ (\mathbf{b}_i \mid U_i, \delta_i) &\sim N[\mathbf{0}, \mathbf{G}_i \{\boldsymbol{\phi}_{\delta_i}(u)\}], \end{aligned} \quad (2.1)$$

where $\boldsymbol{\beta}_1(u)$ and $\boldsymbol{\beta}_0(u)$ are $2 \times p \times 1$ vectors of unknown regression coefficient functions and $\boldsymbol{\phi}_1(u)$ and $\boldsymbol{\phi}_0(u)$ are the vectors of unknown variance-component functions.

We use a Cholesky decomposition for modeling the variance components as the functions of u (Daniels and Zhao, 2003). Other formulations using multivariate normal distributions are possible; this one is chosen for convenience. Details are given in the supplementary material available at <http://www.biostatistics.oxfordjournals.org>.

In the HERS analysis reported in Section 4.1, we assume that $\boldsymbol{\beta}_1(u)$ and $\boldsymbol{\phi}_1(u)$ are unspecified smooth functions that are modeled by penalized splines. Because the administrative censoring times in these data are similar, we assume that $\boldsymbol{\beta}_0(u)$ and $\boldsymbol{\phi}_0(u)$ are constant functions. In practice, when study participants have staggered entry and the administratively censored individuals are a heterogeneous group with respect to the outcome distribution, we could also allow $\boldsymbol{\beta}_0(u)$ and $\boldsymbol{\phi}_0(u)$ to be unspecified smooth functions. Note that with variance components varying by the dropout/administrative censoring times, we need to pay attention to the effective number of parameters that are incorporated in the VCM (Spiegelhalter *and others*, 2002). If the results for variance-component functions suggest particular parametric forms, we could reduce the model complexity accordingly.

2.2 A model for binary longitudinal data

To illustrate the VCM for binary longitudinal data, we build on MTMs (Heagerty, 2002), where the mean and serial dependence structures, and their dependence on the dropout process, are separately specified.

Specifically, let $\mu_{ij}^M(u) = E\{\mathbf{Y}_i(t_{ij}) \mid \mathbf{x}_i(t_{ij}), U_i = u, \delta_i\}$ ($j = 1, \dots, n_i$) and

$$g\{\mu_{ij}^M(u)\} = \mathbf{x}_i(t_{ij})\boldsymbol{\beta}_{\delta_i}(u), \quad (2.2)$$

where $g(\cdot)$ is a link function, $\mathbf{x}_i(t_{ij})$ is a $1 \times p$ covariate vector, and $\boldsymbol{\beta}_1(u)$ and $\boldsymbol{\beta}_0(u)$ are $2 \times p \times 1$ vectors of unknown regression coefficient functions.

Serial dependence between the outcomes within individuals follows an r th-order Markov model; that is,

$$\begin{aligned} \mu_{ij}^C(u) &= E\{Y_i(t_{ij}) \mid U_i = u, \delta_i, \mathbf{x}_i(t_{ij}), Y_i(t_{i,j-1}), \dots, Y_i(t_{i,1})\} \\ &= E\{Y_i(t_{ij}) \mid U_i = u, \delta_i, \mathbf{x}_i(t_{ij}), Y_i(t_{i,j-1}), \dots, Y_i(t_{i,j-r})\}. \end{aligned}$$

The dependence structure is modeled via

$$\text{logit}\{\mu_{ij}^C(u)\} = \Delta_{ij} + \sum_{l=1}^r \gamma_{ijl,\delta_i}(u) \cdot y_{j-l}, \quad (2.3)$$

although in principle any valid link function can be used (Heagerty, 2002). Note that, for simplicity, the dependence of Δ_{ij} and $\gamma_{ijl,\delta_i}(u)$ on $\mathbf{x}_i(t_{ij})$ is suppressed for now. The log-odds ratios $\gamma_{ijl,\delta_i}(u)$ measure the dependence between $Y_i(t_{ij})$ and $Y_i(t_{i,j-1}), \dots, Y_i(t_{i,j-r})$ among those with $U_i = u$ and $\delta_i = \delta$; the intercept Δ_{ij} is determined such that the mean structure in (2.2) and the Markov dependence structure in (2.3) are simultaneously satisfied (Azzalini, 1994; Heagerty, 2002).

We further assume that the serial dependence $\gamma_{ijl,\delta_i}(u)$ can be modeled via

$$\gamma_{ijl,\delta_i}(u) = \mathbf{z}_{i,l}(t_{ij})\boldsymbol{\alpha}_{l,\delta_i}(u), \quad l = 1, \dots, r, \quad (2.4)$$

where $\mathbf{z}_{i,l}(t_{ij})$ is a subset of the covariates $\mathbf{x}_i(t_{ij})$, while $\boldsymbol{\alpha}_{l,1}(u)$ and $\boldsymbol{\alpha}_{l,0}(u)$ are $2 \times d_l \times 1$ ($l = 1, \dots, r$) vectors of unknown functions of u . For example, if $\gamma_{ijl,1}(u) = \alpha_{l1,1}(u) + \alpha_{l2,1}(u) \cdot Z_i$, where Z_i is a treatment group indicator, individuals for whom $Z_i = 1$ are allowed to have different serial dependence compared with individuals for whom $Z_i = 0$, given that they drop out at u .

As with the VCM for the continuous HERS depression data, we assume that each element of the MTM parameters is an unspecified smooth function of u for individuals who dropped out of the HERS, while for administratively censored individuals, we assume that the MTM parameters are constant in u .

3. ESTIMATION

3.1 Joint likelihood

Suppose $\boldsymbol{\pi}$ indexes the dropout/administrative censoring time distribution $f(u, \delta \mid \mathbf{x}; \boldsymbol{\pi})$, and let Θ denote the set of parameters in the VCM for the outcome process. The likelihood from the i th individual can be partitioned as

$$L_i(\Theta, \boldsymbol{\pi} \mid \mathbf{y}_i, \mathbf{x}_i, u_i, \delta_i) \propto f(\mathbf{y}_i \mid \mathbf{x}_i, u_i, \delta_i; \Theta) f(u_i, \delta_i \mid \mathbf{x}_i; \boldsymbol{\pi}).$$

If the priors for $\boldsymbol{\pi}$ and Θ are independent, it follows that $\boldsymbol{\pi}$ is not a part of the posterior for Θ . The inference for $f(u, \delta \mid \mathbf{x}; \boldsymbol{\pi})$ can be based on the marginal likelihood $\prod_{i=1}^N f(u_i, \delta_i \mid \mathbf{x}_i; \boldsymbol{\pi})$, whereas the inference for Θ is based on $\prod_{i=1}^N f(\mathbf{y}_i \mid \mathbf{x}_i, u_i, \delta_i; \Theta)$.

Likelihood for the model with continuous data. In the HERS analysis reported in Section 4.1, the set of parameters Θ here includes those indexing the smooth functions for the regression coefficients and variance components when the dropout time is observed and the parameter vector (constant in u) when the dropout time is administratively censored. Using the same notation as in Section 2.1, the log-likelihood associated with the continuous outcome process can be written as

$$l = -N \log(2\pi)/2 - \log(|V_i|)/2 + \sum_{i=1}^N \left[-\{\mathbf{y}_i - \mathbf{x}_i \boldsymbol{\beta}_{\delta_i}(u_i)\}' V_i^{-1} \{\mathbf{y}_i - \mathbf{x}_i \boldsymbol{\beta}_{\delta_i}(u_i)\} / 2 \right],$$

where $V_i = \mathbf{z}_i \mathbf{G}_i \{\boldsymbol{\phi}_{\delta_i}(u_i)\} \mathbf{z}_i' + \mathbf{R}_i \{\boldsymbol{\phi}_{\delta_i}(u_i)\}$.

Likelihood for the model with binary data. In the HERS analysis reported in Section 4.1, we assume a first-order serial dependence structure. The likelihood contribution for the i th individual corresponding to the model in (2.2–2.4) can be written as

$$\begin{aligned} L_i &= f\{y_{i1} \mid \mathbf{x}_i(t_{i1}), u_i, \delta_i; \boldsymbol{\beta}_{\delta_i}(u_i)\} f\{y_{i2} \mid \mathbf{x}_i(t_{i2}), y_{i1}, u_i, \delta_i; \boldsymbol{\beta}_{\delta_i}(u_i), \boldsymbol{\alpha}_{1,\delta_i}(u_i)\} \\ &\quad \times \cdots \times f\{y_{i n_i} \mid \mathbf{x}_i(t_{i n_i}), y_{i, n_i-1}, u_i, \delta_i; \boldsymbol{\beta}_{\delta_i}(u_i), \boldsymbol{\alpha}_{1,\delta_i}(u_i)\} \\ &= \{\mu_{ij}^M(u_i)\}^{y_{i1}} \{1 - \mu_{ij}^M(u_i)\}^{(1-y_{i1})} \prod_{j=2}^{n_i} \{\mu_{ij}^C(u_i)\}^{y_{ij}} \{1 - \mu_{ij}^C(u_i)\}^{(1-y_{ij})}. \end{aligned}$$

3.2 Bayesian penalized splines

The smooth functions in (2.1) and (2.2–2.4) are modeled by Bayesian penalized splines with low-rank thin-plate spline bases (Ruppert *and others*, 2003; Crainiceanu *and others*, 2005).

The low-rank thin-plate spline representation of a scalar smooth function $\theta(\cdot)$ is

$$\theta(u; \boldsymbol{\eta}) = \zeta_0 + \zeta_1 \cdot u + \sum_{k=1}^K \psi_k \cdot |u - v_k|^3, \quad (3.1)$$

where $\boldsymbol{\eta} = (\zeta_0, \zeta_1, \psi_1, \dots, \psi_K)^\top$ is a vector of regression coefficients and $v_1 < \dots < v_K$ are fixed knots. We set v_k at the $k/(K+1)$ sample quantile of us (Ruppert, 2002; Ruppert *and others*, 2003; Crainiceanu *and others*, 2007). Let $\boldsymbol{\xi} = (\zeta_0, \zeta_1)^\top$, $\boldsymbol{\psi} = (\psi_1, \dots, \psi_K)^\top$, $\mathbf{U}_1 = (1, u)$, $\mathbf{U}_2 = (|u - v_1|^3, \dots, |u - v_K|^3)$, and $\boldsymbol{\Omega}$ be a $K \times K$ matrix whose (l, k) th entry is $|v_l - v_k|^3$. Using the reparameterization $\boldsymbol{\psi} = \boldsymbol{\Omega}^{1/2} \tilde{\boldsymbol{\psi}}$ and $\tilde{\mathbf{U}}_2 = \mathbf{U}_2 \boldsymbol{\Omega}^{-1/2}$, (3.1) can be rewritten as $\theta(u; \boldsymbol{\eta}) = \mathbf{U}_1 \boldsymbol{\xi} + \tilde{\mathbf{U}}_2 \tilde{\boldsymbol{\psi}}$.

In the HERS analysis reported in Section 4, we assign to $\boldsymbol{\xi}$ independent normal priors with mean zero and large variance and to $\tilde{\boldsymbol{\psi}}$ the prior $N(\mathbf{0}, \lambda \cdot \mathbf{I})$, where \mathbf{I} is a $K \times K$ identity matrix. Estimating the smoothing parameter λ is similar to estimating variance components in Bayesian hierarchical models (Gelman, 2006), and the curve estimation by penalized splines can be sensitive to the choice of prior for λ . Crainiceanu *and others* (2007) discussed this issue and found that inverse-Gamma priors can be used in practice when certain conditions are met such that the posterior inference of λ is insensitive to the hyperparameters in the prior for λ . In our applications, we use inverse-Gamma priors for λ and the estimated curves fit the observed data reasonably well. Additional analyses using Uniform priors for $\lambda^{1/2}$ give similar results for curve estimation. Therefore, we only present the results with inverse-Gamma priors for λ .

3.3 Bayesian bootstrap

In our VCM approach, we leave the dropout/administrative censoring time distribution $f(u_i, \delta_i \mid \mathbf{x}_i; \boldsymbol{\pi})$ completely unspecified and use Rubin's (1981) BB (Kim *and others*, 2005) to obtain the posterior for $P(U = u_i, \delta = \delta_i \mid \mathbf{x}_i)$.

We now briefly describe the BB procedure. Suppose $\mathcal{U} = (U_1, \dots, U_N)$ is a random sample from an unknown distribution. For simplicity, we assume that there are no ties in \mathcal{U} . The BB posterior for $\pi_i = P(U = U_i)$ can be obtained by

$$\text{BB posterior} \propto \text{empirical likelihood} \times \text{prior}.$$

Because there is only one observation at each U_i , the empirical likelihood is given by

$$L = \prod_{i=1}^N \pi_i.$$

Using a noninformative prior $\prod_{i=1}^N \pi_i^{-1}$ for (π_1, \dots, π_N) , we have Rubin's BB posterior

$$(\pi_1, \dots, \pi_N) \sim \text{Dirichlet}(1, \dots, 1).$$

In the HERS analysis reported in Section 4, at each iteration of the MCMC we then simulate $P(U = u_i, \delta = \delta_i \mid \mathbf{x}_i)$ from $\text{Dirichlet}(1, \dots, 1)$ for each combination of the discrete covariates.

3.4 Summarizing marginal covariate effects

To obtain inference on the marginal mean $E(\mathbf{Y}_i \mid \mathbf{x}_i)$, the empirical averages $\sum_{i=1}^{N^*} P(U = u_i, \delta = \delta_i \mid \mathbf{x}_i) E(\mathbf{Y}_i \mid \mathbf{x}_i, u_i, \delta_i)$ can be computed using the posterior samples of $P(U = u_i, \delta = \delta_i \mid \mathbf{x}_i)$ and $E(\mathbf{Y}_i \mid \mathbf{x}_i, u_i, \delta_i)$ from the VCM, where N^* is the sample size corresponding to a specific combination of the discrete covariate values.

For example, in (2.1), when the identity link is used for modeling the mean structure and $f(u, \delta \mid \mathbf{x}) = f(u, \delta)$, the marginal covariate effects can be approximated by $\sum_{i=1}^N P(U = u_i, \delta = \delta_i) \boldsymbol{\beta}_{\delta_i}(u_i)$. However, when other link functions are used for the mean structure and/or $f(u, \delta \mid \mathbf{x}) \neq f(u, \delta)$, the marginal covariate effects might not be readily available. Here, the effect of covariate difference $\mathbf{x} - \mathbf{x}'$ is

$$E(\mathbf{Y}_i \mid \mathbf{x}) - E(\mathbf{Y}_i \mid \mathbf{x}') = \int E(\mathbf{Y}_i \mid \mathbf{x}, u, \delta) f(u, \delta \mid \mathbf{x}; \boldsymbol{\pi}) du d\delta - \int E(\mathbf{Y}_i \mid \mathbf{x}', u, \delta) f(u, \delta \mid \mathbf{x}'; \boldsymbol{\pi}) du d\delta,$$

which cannot be simplified to $(\mathbf{x} - \mathbf{x}') E_{U, \delta} \{ \boldsymbol{\beta}_{\delta}(u) \}$ because of its dependence on \mathbf{x} . In a simple scenario with treatment groups and measurement times as the covariates, we can compute $\sum_{i=1}^{N^*} P(U = u_i, \delta = \delta_i \mid \mathbf{x}_i) E(\mathbf{Y}_i \mid \mathbf{x}_i, u_i, \delta_i)$ and plot summaries of the posteriors to demonstrate the marginal covariate effects. For other more complicated situations with many confounders or a number of quantitative covariates of interest, a simple summary of the marginal effects in PMMs might not be immediately obtainable (Fitzmaurice and Laird, 2000).

3.5 Markov Chain Monte Carlo

The prior specification for Bayesian penalized splines is discussed in Section 3.2. For the constant parameter vector in the administrative censoring group, independent vague normal priors with mean zero and large variance are assigned. We use MCMC to sample from the posterior distributions; summary statistics

such as posterior means and 95% credible bands are then used for inference. The MCMC is implemented in the WinBUGS package (version 1.4.1) and its development interface (WBDev; Spiegelhalter *and others*, 2003). The programs for the HERS analysis reported in Section 4 are provided in the supplementary material available at *Biostatistics* online.

4. APPLICATION: THE HERS STUDY

Our goal is to describe the depression changes over time by baseline characteristics, such as race (Black, White, Latina including others) and baseline CD4 count ($CD4 > 200$), for the 753 women who did not suffer HIV-related deaths in the HERS. We first present the analysis using the continuous CES-D data and then analyze the binary data using the cutoff $CES-D \geq 16$ to define clinical depression.

4.1 Continuous CES-D data

Models. We fit 3 models. The first is a LMM assuming that $(Y_{ij} | \mathbf{b}_i, \mathbf{X}_{ij}, \mathbf{Z}_{ij}) \sim N(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\mathbf{b}_i, \tau^2)$, where Y_{ij} is the CES-D score at time t_{ij} ; \mathbf{X}_{ij} is the covariate vector that includes intercept, race, baseline CD4 count, time, and the interaction between time and baseline CD4 count; and \mathbf{Z}_{ij} is the covariate vector associated with a random intercept and a random time slope $\mathbf{b}_i = (b_{i1}, b_{i2})^T$. In addition, we fit 2 VCMs. The first has unspecified smooth functions of u in the mean structure only (VCM1). Specifically, it is assumed that $(Y_{ij} | \mathbf{b}_i, \mathbf{X}_{ij}, \mathbf{Z}_{ij}, u, \delta_i) \sim N(\mathbf{X}_{ij}\boldsymbol{\beta}_{\delta_i}(u) + \mathbf{Z}_{ij}\mathbf{b}_i, \tau^2)$, where $\boldsymbol{\beta}_0(u) = \boldsymbol{\beta}_0$ is a vector that is constant in u for the administrative censoring group. The second VCM also includes the variance components as smooth functions of u (VCM2). Further details about variance-component parametrization, prior specification as well as posterior inference can be found in the supplementary material available at *Biostatistics* online.

Results. We first focus on the results from VCM2. Additional results from the LMM and VCM1 as well as an example of sensitivity analysis based on VCM2 can be found in the supplementary material available at *Biostatistics* online. Figure 2 gives the results for $\boldsymbol{\beta}_1(u)$ and $\boldsymbol{\beta}_0$. The intercept and race effects are fairly constant over u . The main effect of baseline CD4 count is decreasing as u increases. The main effect of time has a downward trend toward zero, which suggests that for the group with baseline $CD4 \leq 200$ earlier dropout was associated with steeper change in expected CES-D scores over time. The interaction between time and baseline CD4 count increases toward zero, which shows that the positive time slopes for the group with baseline $CD4 > 200$ are less steep than those with baseline $CD4 \leq 200$. Overall, we expect that VCM2 will adjust the expected CES-D profiles upward and the adjustment might differ between the baseline CD4 groups.

Figure 3 shows the estimated smooth functions for the variance components. Since later dropouts are generally associated with more observations within patients, we would expect that the estimated variability of the random intercepts, random slopes, and residual errors decreases as u increases. As seen from Figure 3, this is true for the estimated random-slope standard deviation (SD) and the error SD; but the estimated random-intercept SD increases as u increases. This upward trend for random-intercept SD suggests that those early dropouts in the HERS might be a more homogeneous group in terms of their baseline CES-D levels. Overall, all estimated variance components are not constant over u . In fact, by allowing the variance components to vary by u , the within-individual correlation structure in VCM2 is different from the one in VCM1 (see the supplementary material available at *Biostatistics* online for variance-component estimates from the LMM and VCM1). It is well known that with complete data and likelihood-based approaches, properly modeling the within-individual correlation structure can affect the variability estimates more than the point estimates of the mean regression coefficients (Diggle *and others*,

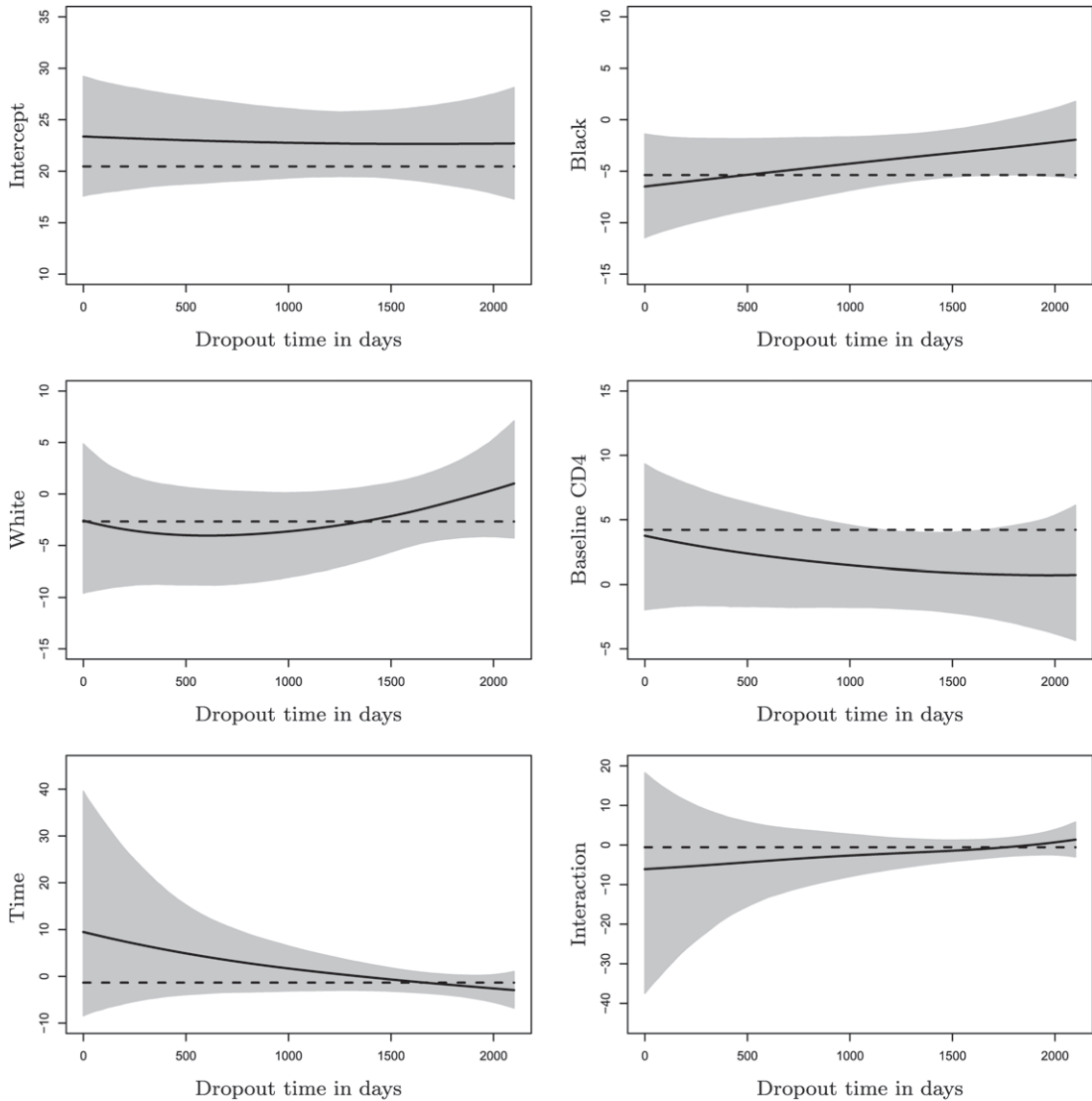


Fig. 2. Estimated smooth functions of the observed dropout times in the mean structure from VCM2 for the continuous CES-D data in the HERS. Gray shades are the pointwise 95% credible bands and dashed lines are the corresponding estimates of the regression coefficients β_0 in the administrative censoring group.

2002). However, with missing data, even point estimates can be biased if the correlation structure is not carefully modeled (Daniels and Hogan, 2008). In our case, because of the apparent dependence between variance components and dropout times in the observed data, we would expect that the 2 VCMs might provide different point estimates and variability estimates for the marginal covariate effects.

Table 1 gives the results for the intercept and the time slope in estimated CES-D profile by race and baseline CD4 count. The intercept estimates are close across all models, which is expected because in early study period the influence of dropout is minimal. However, both VCMs adjust the time slope estimates upward compared with the LMM, where for the group with baseline $CD4 \leq 200$, the adjustment is larger

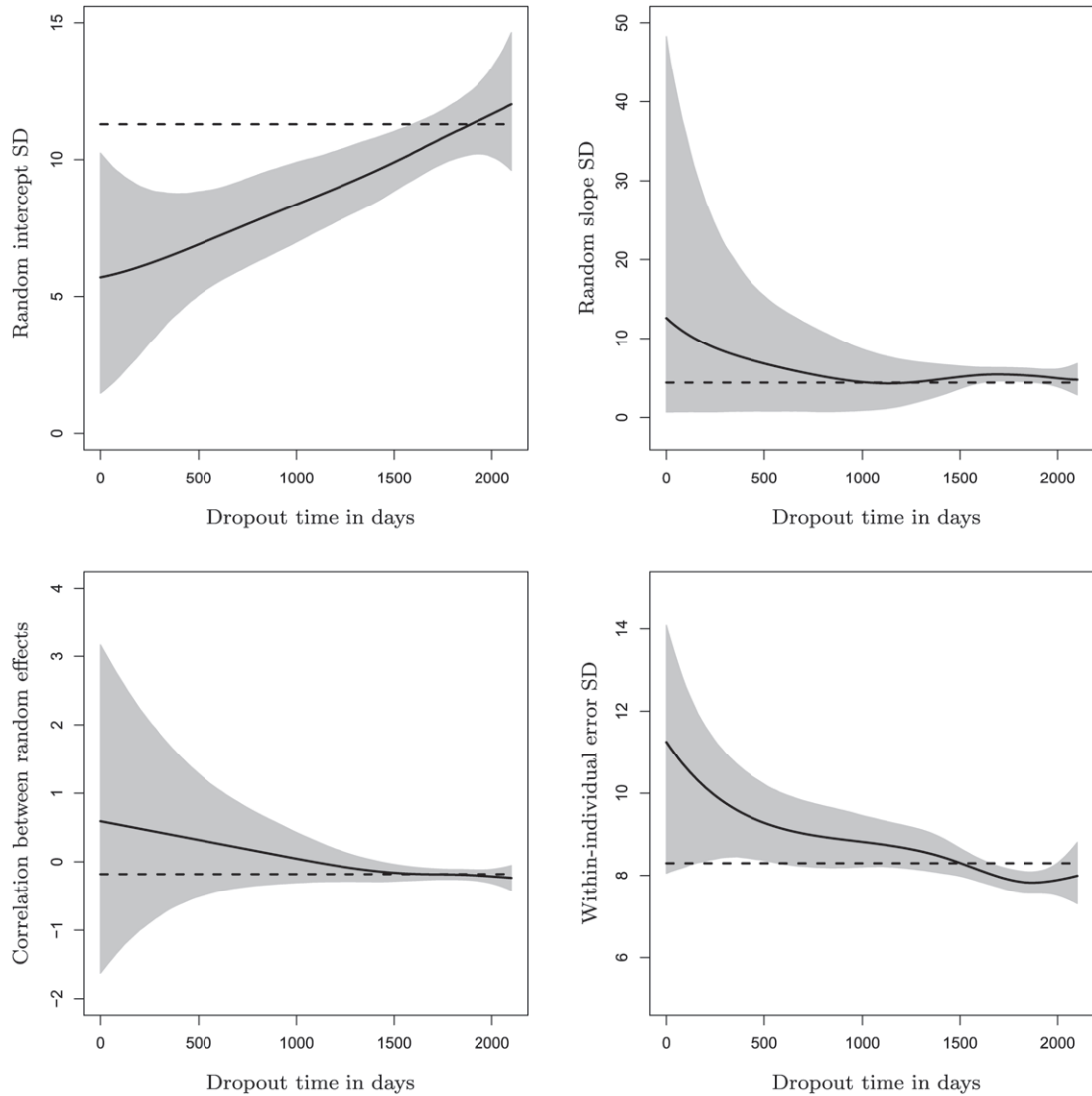


Fig. 3. Estimated smooth functions of the observed dropout times for the variance components from VCM2 for the continuous CES-D data in the HERS. Gray shades are the pointwise 95% credible bands and dashed lines are the corresponding estimates of the variance components in the administrative censoring group.

and the time slopes are changed completely to be positive. Therefore, without taking into account the dropout process such as in the LMM, we might incorrectly conclude that both baseline CD4 groups had downward trends for the CES-D scores over time. Further, both VCMs give similar time slope estimates for the group with baseline $CD4 > 200$, but the point estimates and variability estimates of the time slopes in the group with baseline $CD4 \leq 200$ differ between the VCMs. This might be explained by the different levels of missingness between the baseline CD4 groups. Allowing the variance components to vary by u in VCM2 therefore might have larger impact on the point and variability estimates for the group with baseline $CD4 \leq 200$.

Table 1. *Estimated intercept and time slope (posterior mean and 95% credible interval) in the expected linear CES-D profiles (by race and baseline CD4 groups) under 3 different models for the continuous CES-D data in the HERS*

		LMM		VCM1		VCM2	
		Intercept	Slope	Intercept	Slope	Intercept	Slope
CD4 ≤ 200	Latina	22.6 (19.8, 25.3)	-1.3 (-3.0, 0.2)	22.5 (19.5, 25.5)	2.7 (-2.6, 14.7)	22.6 (19.8, 25.4)	1.9 (-3.3, 10.3)
	Black	18.6 (16.2, 21.1)	-1.3 (-3.0, 0.2)	18.3 (15.5, 21.0)	1.5 (-2.3, 8.7)	18.4 (15.9, 21.0)	1.0 (-2.8, 6.3)
	White	20.5 (17.8, 23.4)	-1.3 (-3.0, 0.2)	20.1 (16.6, 23.6)	2.5 (-2.8, 16.4)	20.1 (16.7, 23.5)	1.6 (-3.5, 10.7)
CD4 > 200	Latina	24.3 (22.6, 26.1)	-2.1 (-2.6, -1.5)	24.5 (22.7, 26.3)	-1.0 (-3.0, 1.9)	24.5 (22.6, 26.3)	-1.0 (-3.0, 2.0)
	Black	20.4 (19.3, 21.4)	-2.1 (-2.6, -1.5)	20.3 (19.2, 21.4)	-1.4 (-2.6, 0.2)	20.3 (19.2, 21.4)	-1.4 (-2.7, 0.3)
	White	22.3 (20.7, 23.9)	-2.1 (-2.6, -1.5)	22.4 (20.6, 24.1)	-1.4 (-2.5, 0.0)	22.2 (20.6, 24.0)	-1.4 (-2.6, 0.1)

In summary, regardless of baseline CD4 count, we observed that Whites and Latinas (including others) had larger CES-D scores over time than Blacks. For all races, the expected CES-D scores for the patients with baseline CD4 ≤ 200 were increasing over time, while the CES-D scores for those with baseline CD4 > 200 were decreasing over time.

4.2 Binary CES-D data

Models. First we fit a MTM(1). The same set of covariates as in Section 4.1 is used. The marginal mean of depression follows $\text{logit}(\mu_{ij}^M) = \mathbf{X}_{ij}\boldsymbol{\beta}$, and the dependence structure is assumed to follow a first-order Markov model with constant serial dependence $\text{logit}(\mu_{ij}^C) = \Delta_{ij} + \alpha \cdot y_{ij-1}$. We then fit the varying-coefficient MTM(1) as in Section 2.2. The mean structure for the dropout group follows

$$\begin{aligned} \text{logit}\{\mu_{ij}^M(u_i)\} &= \beta_0 + \beta_1 \cdot \text{I(Black)} + \beta_2 \cdot \text{I(White)} + \beta_3(u_i) \cdot \text{I(baseline CD4 > 200)} \\ &+ \beta_4(u_i) \cdot t_{ij} + \beta_5(u_i) \cdot \text{I(baseline CD4 > 200)} \cdot t_{ij}, \end{aligned} \quad (4.1)$$

while the dependence structure follows $\text{logit}\{\mu_{ij}^C(u_i)\} = \Delta_{ij} + \alpha(u_i) \cdot y_{ij-1}$. Further details are provided in the supplementary material available at *Biostatistics* online.

Results. For individuals with observed dropout times, Figure 4 presents the estimated smooth functions. The main effect for baseline CD4 count shows a downward trend over u . The main effect for time decreases to approximately zero as u increases, which again suggests that earlier dropout was associated with larger time slope in depression probability for the group with baseline CD4 ≤ 200. The interaction between time and baseline CD4 count increases toward zero, which suggests that the group with baseline CD4 > 200 had time slopes that are less varying over u . Overall, we expect that the VCM could adjust the marginal probability profiles of depression upward for the group with baseline CD4 ≤ 200. The within-individual serial dependence is positive and increases slightly as u increases. Note that in Figure 4 the corresponding estimates from the administrative censoring group are close to those at the right boundary of the observed dropout times.

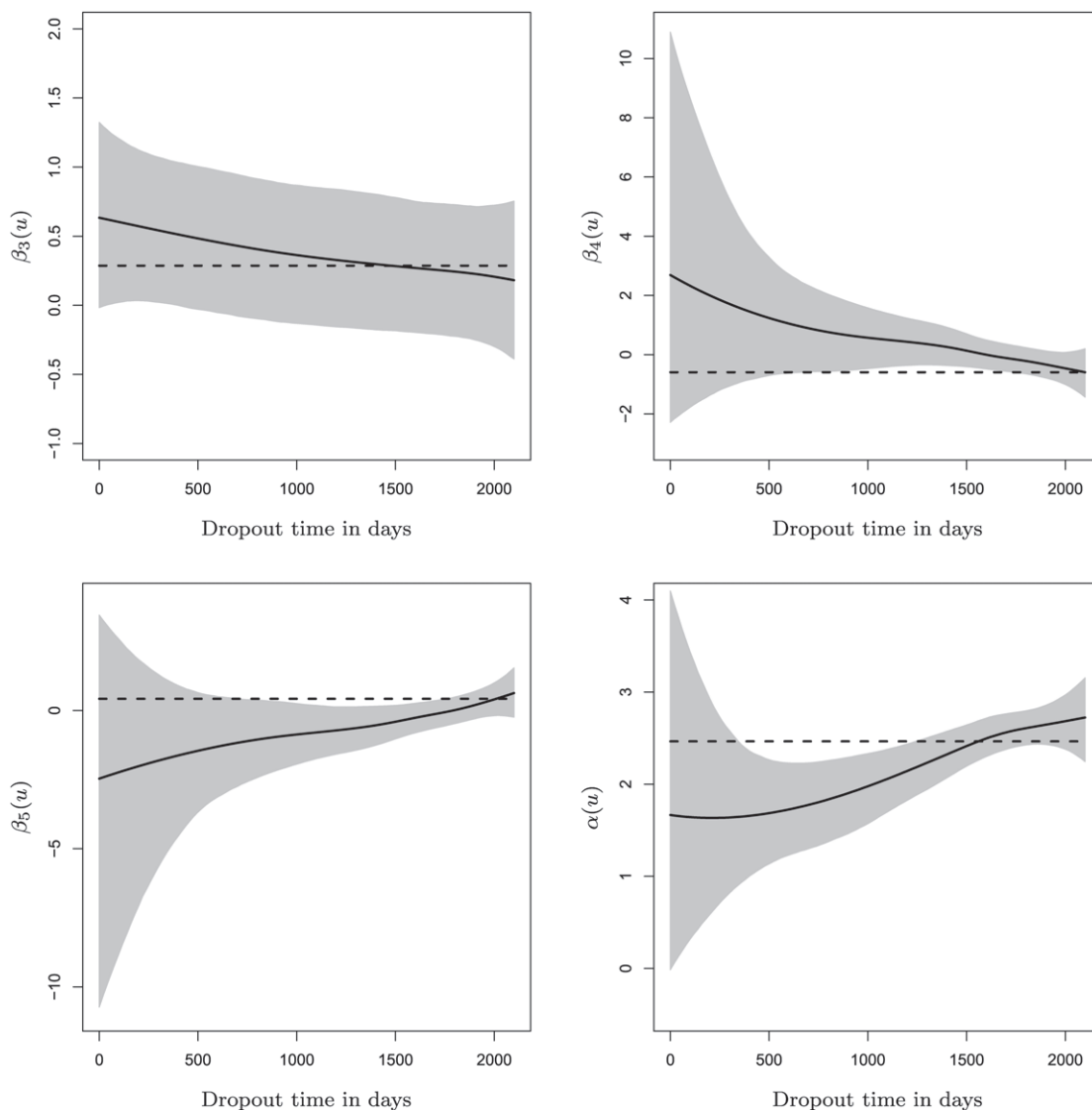


Fig. 4. Estimated smooth functions of the observed dropout times (on logit scale) in varying-coefficient MTM(1) for the binary CES-D data in the HERS; these include the coefficients for baseline CD4 > 200 ($\beta_3(u)$), the time ($\beta_4(u)$), the interaction between time and baseline CD4 count ($\beta_5(u)$), and the serial dependence ($\alpha(u)$). Gray shades are the pointwise 95% credible bands and dashed lines are the corresponding estimates in the administrative censoring group.

Table 2 presents the estimated marginal covariate effects from the fitted MTM(1) assuming MAR. The estimated interaction between time and baseline CD4 count is positive, which means that regardless of race, the group with baseline CD4 ≤ 200 had steeper decline in depression prevalence over time. Based on substantive knowledge and the results in Section 4.1, this is not sensible and may be an artefact selection bias due to informative dropout. Marginal probability profiles estimated from the VCM by race and baseline CD4 count are presented in Figure 4 of the supplementary material available at *Biostatistics*

Table 2. *Estimated covariate effects (posterior mean and 95% credible interval) from the fitted MTM(1) (assuming MAR) for the binary CES-D data in the HERS*

Parameter	Mean	LCL	UCL
Intercept	0.58	0.20	0.97
Black	-0.51	-0.73	-0.29
White	-0.32	-0.58	-0.06
Baseline CD4 > 200	0.25	-0.13	0.62
Time	-0.31	-0.66	0.02
Interaction between time and baseline CD4	0.10	-0.25	0.47

LCL, lower credible limit; UCL, upper credible limit.

online. Apparently, for the group with baseline CD4 ≤ 200 , the VCM adjusts the marginal probability of depression; and the downward trends shown in the MTM(1) under MAR are moved upward. The adjustment for the group with baseline CD4 > 200 is minimal. As a result, the estimated interaction between time and baseline CD4 count becomes negative in the VCM, which is shown by the difference between marginal probability profiles in Figures 5 and 6 of the supplementary material available at *Biostatistics* online. Note that after averaging over the dropout/administrative censoring time distribution, the effects of race, baseline CD4 count, and time are no longer independent. Results for the administrative censoring group are also given in the supplementary material available at *Biostatistics* online.

In summary, regardless of baseline CD4 count, we observed that Latinas (including others) had higher prevalence of depression over time than Blacks and Whites. Given the race groups, the patients with baseline CD4 ≤ 200 had similar depression prevalence over time as for the patients with baseline CD4 > 200; unlike the results based on continuous CES-D data, the depression prevalence remained relatively constant over time for all race and baseline CD4 groups. It should be noted that the analyses based on continuous and binary CES-D data focused on different scientific questions. With continuous CES-D data, we are interested in the covariate effects on the absolute levels of CES-D, while with binary CES-D data the targets are the covariate effects on the prevalence of clinical depression. We have seen that in both cases the race effects are similar, but the baseline CD4 effects differ.

5. DISCUSSION

We have proposed a Bayesian VCM approach for longitudinal data with continuous-time informative dropout. Our framework assumes that the parameters in the outcome process depend on the dropout time through unspecified functions, where administratively censored dropout times are handled separately and no modeling of the continuous dropout time distribution is needed in order to obtain the inference for marginal covariates effects. While the VCM is widely applicable, we used both continuous and binary data from an HIV longitudinal study to show that our approach has the potential to adjust for selection biases induced by early dropouts of poor responders.

Our VCM approach provides a convenient framework for sensitivity analysis because the unidentifiable part of the model can be distinguished from the identifiable part and for the latter the inferences remain the same regardless of the sensitivity parameters. In our analysis of the HERS depression data, we emphasized that sensitivity analysis should be based on those parameters that cannot be identified by the observed data. More in-depth research on this aspect is needed, building on general sensitivity analysis strategies developed for PMMs (Scharfstein *and others*, 1999; Daniels and Hogan, 2000; Molenberghs *and others*, 2003). For example, informative priors for sensitivity parameters can be introduced using expert opinions and/or prior elicitation based on previous studies (Lee, 2007).

Appropriate summary of marginal covariate effects is a challenge in the pattern-mixture modeling approach to informative dropout. In practice, we might prefer to specify the marginal covariate effects directly in the model. Thus, approaches to marginalizing PMMs are worth further research (Wilkins and Fitzmaurice, 2006, 2007; Roy and Daniels, 2007). In our ongoing research, we plan to extend the VCM for binary data by separately specifying the marginal model and the conditional model given the dropout/administrative censoring time, while constraints are imposed such that they are satisfied simultaneously.

In our VCM, we distinguished administrative censoring from other dropout. In the HERS application, we assumed that for the administrative censoring group the outcome model parameters do not vary with the administrative censoring times but are distinct from the parameters in the dropout group. However, our VCM specification is flexible, and in practice, similar unspecified smooth functions can also be used to capture the heterogeneity within the administrative censoring group with respect to the outcome process. This is particularly useful when study participants have staggered entry and the observed administrative censoring times vary considerably. When there is no dropout, variation in these administrative censoring times is usually ignorable. However, when informative dropout is present, for example, in the context of the HERS analysis, it is possible that the longer a participant stays on a study without dropping out, the less steep the patient's true depression trend is likely to be. In this situation, modeling the relationship between the outcome process and the administrative censoring times would be necessary.

Outcome-related death mixed with dropout is another problem that warrants further research. Because extrapolating the missing data beyond death is inappropriate, instead of modeling the marginal mean of the outcomes, a more meaningful quantify of interest would be the mean of the longitudinal outcomes conditional on being alive (Kurland and Heagerty, 2005). When the survival information is available, we could extrapolate the missing data in the VCM up to the observed survival times for summarizing marginal covariate effects. If survival times are censored, further work on joint modeling is needed.

ACKNOWLEDGMENTS

The authors thank Jeffrey Blume, Mike Daniels, Constantine Gatsonis, Patrick Heagerty, Tony Lancaster, the editor, and the referee for helpful comments and suggestions. *Conflict of Interest*: None declared.

FUNDING

The National Institutes of Health (R01-AI-50505, R01-HL-79457); the US Centers for Disease Control and Prevention (U64-CCU10675). Funding to pay the Open Access publication charges for this article was provided by Medical Research Council (UK) grant U.1052.00.009.

REFERENCES

- AZZALINI, A. (1994). Logistic regression for autocorrelated data with application to repeated measures. *Biometrika* **81**, 767–775.
- COOK, J., GREY, D., BURKE, J., COHEN, M., GURTMAN, A., RICHARDSON, J., WILSON, T., YOUNG, M. AND HESSOL, N. (2004). Depressive symptoms and AIDS-related mortality among a multisite cohort of HIV-positive women. *American Journal of Public Health* **94**, 1133–1140.
- CRAINICEANU, C., RUPPERT, D. AND CARROLL, R. J. (2007). Spatially adaptive penalized splines with heteroscedastic errors. *Journal of Computational and Graphical Statistics* **16**, 265–288.

- CRAINICEANU, C., RUPPERT, D. AND WAND, M. (2005). Bayesian analysis for penalized spline regression using WinBUGS. *Journal of Statistical Software* **14**, 1–24.
- DANIELS, M. AND HOGAN, J. (2000). Reparameterizing the pattern mixture model for sensitivity analyses under informative dropout. *Biometrics* **56**, 1241–1248.
- DANIELS, M. AND HOGAN, J. (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Monographs on Statistics and Applied Probability, Volume 101. New York: Chapman & Hall.
- DANIELS, M. AND ZHAO, Y. (2003). Modelling the random effects covariance matrix in longitudinal data. *Statistics in Medicine* **22**, 1631–1647.
- DIGGLE, P., HEAGERTY, P., LIANG, K.-Y. AND ZEGER, S. (2002). *Analysis of Longitudinal Data*. New York: Oxford University Press.
- DIGGLE, P. AND KENWARD, M. (1994). Informative dropout in longitudinal data analysis (with discussion). *Applied Statistics* **43**, 49–93.
- FITZMAURICE, G. M. AND LAIRD, N. M. (2000). Generalized linear mixture models for handling nonignorable dropouts in longitudinal studies. *Biostatistics* **1**, 141–156.
- FOLLMAN, D. AND WU, M. C. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics* **51**, 151–168.
- GELMAN, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**, 515–533.
- HASTIE, T. AND TIBSHIRANI, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society, Series B* **55**, 757–796.
- HEAGERTY, P. (2002). Marginalized transition models and likelihood inference for longitudinal categorical data. *Biometrics* **58**, 342–351.
- HENDERSON, R., DIGGLE, P. AND DOBSON, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1**, 465–480.
- HOGAN, J. W. AND LAIRD, N. M. (1997a). Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine* **16**, 239–257.
- HOGAN, J. W. AND LAIRD, N. M. (1997b). Model-based approaches to analyzing incomplete longitudinal and failure time data. *Statistics in Medicine* **16**, 259–272.
- HOGAN, J. W., LIN, X. AND HERMAN, B. (2004). Mixtures of varying coefficient models for longitudinal data with discrete or continuous non-ignorable dropout. *Biometrics* **60**, 854–864.
- ICKOVICS, J. R., HAMBURGER, M. E., VLAHOV, D., SCHOENBAUM, E. E., SCHUMAN, P., BOLAND, R. J. AND MOORE, J. FOR THE HIV EPIDEMIOLOGY RESEARCH STUDY GROUP (2001). Mortality, CD4 cell count decline, and depressive symptoms among HIV-seropositive women. *Journal of the American Medical Association* **285**, 1466–1474.
- KENWARD, M. AND MOLENBERGHS, G. (1999). Parametric models for incomplete continuous and categorical data. *Statistical Methods in Medical Research* **8**, 51–83.
- KIM, Y., LEE, J. AND KIM, J. (2005). Bayesian bootstrap analysis of doubly censored data using Gibbs sampler. *Statistica Sinica* **15**, 969–980.
- KURLAND, B. AND HEAGERTY, P. (2005). Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths. *Biostatistics* **6**, 241–258.
- LEE, J. Y. (2007). Sensitivity analysis and informative priors for longitudinal binary data with dropout, [PhD. Thesis]. Providence, RI: Brown University.
- LESERMAN, J. (2008). Role of depression, stress and trauma in HIV disease progression. *Psychosomatic Medicine* **70**, 539–545.

- LITTLE, R. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* **88**, 125–134.
- LITTLE, R. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika* **81**, 471–483.
- LITTLE, R. (1995). Modeling the dropout mechanism in repeated measures studies. *Journal of the American Statistical Association* **90**, 1112–1121.
- LITTLE, R. AND WANG, Y. (1996). Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics* **52**, 98–111.
- MOLENBERGHS, G., MICHIELS, B., KENWARD, M. G. AND DIGGLE, P. J. (1998). Monotone missing data and pattern-mixture models. *Statistica Neerlandica* **52**, 153–161.
- MOLENBERGHS, G., THIJIS, H., KENWARD, M. AND VERBEKE, G. (2003). Sensitivity analysis of continuous incomplete longitudinal outcomes. *Statistica Neerlandica* **57**, 112–135.
- RADLOFF, L. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement* **1**, 385–401.
- ROTNITZKY, A., SCHARFSTEIN, D., SU, T. AND ROBINS, J. (2001). Methods for conducting sensitivity analysis of trials with potentially nonignorable competing causes of censoring. *Biometrics* **57**, 103–113.
- ROY, J. AND DANIELS, M. J. (2007). A general class of pattern mixture models for nonignorable dropout with many possible dropout times. *Biometrics* **64**, 538–545.
- RUBIN, D. (1981). The Bayesian bootstrap. *Annals of Statistics* **9**, 130–134.
- RUPPERT, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics* **11**, 735–757.
- RUPPERT, D., WAND, M. AND CARROLL, R. (2003). *Semiparametric Regression*. Cambridge: Cambridge University Press.
- SCHARFSTEIN, D., ROBINS, J. AND ROTNITZKY, A. (1999). Adjusting for nonignorable nonresponse using semiparametric nonresponse models with time dependent covariates (with discussion). *Journal of the American Statistical Association* **94**, 1096–1146.
- SMITH, D., WARREN, D., VLAHOV, D., SCHUMAN, P., STEIN, M., GREENBERG, B. AND HOLMBERG, S. (1997). Design and baseline participant characteristics of the human immunodeficiency virus epidemiology research (HER) study: a prospective cohort study of human immunodeficiency virus infection in US women. *American Journal of Epidemiology* **146**, 459–469.
- SPIEGELHALTER, D., BEST, N., CARLIN, B. AND VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* **64**, 583–639.
- SPIEGELHALTER, D., THOMAS, A., BEST, N. AND LUNN, D. (2003). *WinBUGS Version 1.4 User Manual*. Cambridge: Medical Research Council Biostatistics Unit.
- TEN HAVE, T., KUNSELMAN, A., PULKSTENIS, E. AND LANDIS, J. (1998). Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics* **54**, 367–383.
- TSIATIS, A. AND DAVIDIAN, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* **14**, 809–834.
- VERBEKE, G., MOLENBERGHS, G., THIJIS, H., LESAFFRE, E. AND KENWARD, M. (2001). Sensitivity analysis for nonrandom dropout: a local influence approach. *Biometrics* **57**, 7–14.
- WILKINS, K. J. AND FITZMAURICE, G. M. (2006). A hybrid model for nonignorable dropout in longitudinal binary responses. *Biometrics* **62**, 168–176.
- WILKINS, K. J. AND FITZMAURICE, G. M. (2007). A marginalized pattern-mixture model for longitudinal binary data when nonresponse depends on unobserved responses. *Biostatistics* **8**, 297–305.

- WU, M. AND BAILEY, K. (1989). Estimation and comparison of changes in the presence of informative right censoring: conditional linear model (corr:v46 p889). *Biometrics* **45**, 939–955.
- WU, M. AND CARROLL, R. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44**, 175–188.
- WULFSOHN, M. AND TSIATIS, A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330–339.

*[Received July 17, 2008; first revision November 17, 2008; second revision March 18, 2009;
third revision August 6, 2009; accepted for publication September 14, 2009]*