University of Michigan School of Public Health
The University of Michigan Department of Biostatistics Working Paper Series

Year 2003 *Paper* 14

Mixtures of Varying Coefficient Models for Longitudinal Data with Discrete or Continuous Non-ignorable Dropout

Joseph W. Hogan* Xihong Lin†
Benjamin A. Herman‡

*Brown University, jhogan@stat.brown.edu
†University of Michigan, xlin@umich.edu
‡Brown University, Benjamin_A_Herman@Brown.edu

Mixtures of Varying Coefficient Models for Longitudinal Data with Discrete or Continuous Non-ignorable Dropout

Joseph W. Hogan, Xihong Lin, and Benjamin A. Herman

Abstract

The analysis of longitudinal repeated measures data is frequently complicated by missing data due to informative dropout. We describe a mixture model for joint distribution for longitudinal repeated measures, where the dropout distribution may be continuous and the dependence between response and dropout is semiparametric. Specifically, we assume that responses follow a varying coefficient random effects model conditional on dropout time, where the regression coefficients depend on dropout time through unspecified nonparametric functions that are estimated using step functions when dropout time is discrete (e.g., for panel data) and using smoothing splines when dropout time is continuous. Inference under the proposed semiparametric model is hence more robust than the parametric conditional linear model. The unconditional distribution of the repeated measures is a mixture over the dropout distribution. We show that estimation in the semiparametric varying coefficient mixture model can pro- ceed by fitting a parametric mixed-effects model and can be carried out on standard software platforms such as SAS. The model is used to analyze data from a recent AIDS clinical trial and its performance is evaluated using simulations.

Mixtures of Varying Coefficient Models for Longitudinal Data with Discrete or Continuous Non-ignorable Dropout

Joseph W. Hogan
Center for Statistical Sciences & Department of Community Health, Brown University.

Xihong Lin
Department of Biostatistics, University of Michigan

Benjamin Herman
Center for Statistical Sciences, Brown University.

Submitted to *Biometrics*
May 5, 2003

Abstract

The analysis of longitudinal repeated measures data is frequently complicated by missing data due to informative dropout. We describe a mixture model for joint distribution for longitudinal repeated measures, where the dropout distribution may be continuous and the dependence between response and dropout is semiparametric. Specifically, we assume that responses follow a varying coefficient random effects model conditional on dropout time, where the regression coefficients depend on dropout time through unspecified nonparametric functions that are estimated using step functions when dropout time is discrete (e.g., for panel data) and using smoothing splines when dropout time is continuous. Inference under the proposed semiparametric model is hence more robust than the parametric conditional linear model. The unconditional distribution of the repeated measures is a mixture over the dropout distribution. We show that estimation in the semiparametric varying coefficient mixture model can proceed by fitting a parametric mixed-effects model and can be carried out on standard software platforms such as SAS. The model is used to analyze data from a recent AIDS clinical trial and its performance is evaluated using simulations.

Key words: smoothing splines, linear mixed model, missing data, non-ignorable dropout, pattern-mixture model, pediatric AIDS, equivalence trial, selection bias, clinical trials.

Correspondence: Joseph W. Hogan, Center for Statistical Sciences, Box G-H, Brown University, Providence, RI 02912. Phone 401-863-9243. Fax 401-863-9182. Email jhogan@stat.brown.edu

1 Introduction

1.1 Informative dropout

Dropout and other types of missing data are common in long-term longitudinal studies; in many cases, dropout induces a missingness process that is *nonignorable* in the sense that missingness depends probabilistically on unobserved outcomes, even after conditioning on observable information. Most approaches to handling informative dropout in longitudinal data can be viewed as extensions of standard approaches such as multilevel modeling (Laird and Ware, 1982; Diggle, 1988; Breslow and Clayton, 1993) and marginal modeling (Liang and Zeger, 1986). Likelihood-based approaches include selection models (Wu and Carroll, 1988; Diggle and Kenward, 1994; Follman and Wu, 1995; Ten Have et al., 1998) and mixture models (Little, 1993, 1994; Hogan and Laird, 1997a; Wu and Bailey, 1989). Recent comprehensive surveys of parametric and likelihood-based approaches to handling dropout in longitudinal data can be found in Little (1995), Hogan and Laird (1997b), and Kenward and Molenberghs (1999). Moment-based methods also have been generalized to handle informative dropout under the selection modeling framework. See Robins, Rotnitzky and Zhao (1995), Rotnitzky, Robins and Scharfstein (1998), and Scharfstein, Robins and Rotnitzky (1999).

In this paper, we develop a general mixture modeling approach for continuous longitudinal repeated measures data where measurement times may be irregular across subjects and where dropout might be at continuous times and potentially non-ignorable. The conditional distribution of repeated measures given dropout follows a varying coefficient model (Zhang et al., 1998) where regression coefficients such as intercepts and slopes depend on dropout through unspecified nonparametric functions. The shapes of the functions are estimated using step functions when dropout time is discrete (e.g., for panel data), and using natural cubic smoothing splines (Green and Silverman, 1994) when dropout is continuous. We show that estimation in the proposed varying coefficient mixture model can proceed by fitting an augmented parametric mixed effect model. The complete data distribution is a mixture of the varying coefficient models over the dropout distribution, and the dropout distribution can be left completely unspecified.

This class of models can be viewed as a nonparametric extension of pattern mixture

models (Little, 1993, 1994) and conditional linear models (Wu and Bailey, 1989; Hogan and Laird, 1997a) described above. The former author mainly considered panel data, while the latter authors allowed for the dropout time to be continuous but assumed the regression coefficients to be parametric functions of the dropout time. Estimation can therefore be biased if the parametric functions are misspecified. The proposed varying coefficient mixture model relaxes the parametric assumption by providing a unified framework to allow for flexible dependence of the covariate effects on dropout patterns by assuming regression coefficients to be nonparametric functions of dropout times. Hence estimation of the covariate effects is more robust to misspecification of the dependence between longitudinal responses and dropout.

1.2 Motivating example

Protocol 128 of the ACTG was a randomized double-blind equivalency trial of high-dose (180mg per square meter body surface area, six times daily) versus low-dose (90mg) zidovudine (ZDV) for HIV-infected children (Brady et al., 1996). The study enrolled 424 children, randomized them to receive one of the two doses, and followed the children on a number of endpoints for up to five years. In this paper, we are concerned with comparing longitudinal trajectory of CD4 cell counts. Children were scheduled for measurement of CD4 count every 12 weeks, but actual measurement times varied considerably. In addition, only about half of the participants completed three years of follow-up (113/216 [52%] on low dose, 93/208 [45%] on high dose).

A simple but reasonable approach to analyzing these data is to estimate treatment-group-specific CD4 trajectories using a linear random effects model (Laird and Ware, 1982). This model provides valid inference under ignorable dropout (Laird, 1988; Diggle and Kenward, 1994; Little, 1995). Basic exploratory analysis suggests that for the observed data, mean square root of CD4 is well described by a linear time trend in both treatment arms (Figure 1). Using the random effects model, estimated change from baseline to week 200 is –12.7 (s.e. 0.8) in the low dose arm, –18.2 (s.e. 1.4) in the high dose arm, for a difference of –5.5 (s.e. 1.6), favoring low dose.

To explore the potential for bias due to outcome-related dropout, we plotted estimated

individual least-squares slopes versus followup time (Figure 2). A clear pattern is evident in both treatment arms, namely that lower slopes are associated with early dropout, which casts some doubt on the ignorable dropout assumption and hence on the validity of estimates from the linear random effects model, suggesting the need to utilize more elaborate models for addressing potential effects of informative dropout.

The remainder of our paper is organized as follows. The model is described in Section 2, and estimation procedures are detailed in Section 3; this includes estimation for discrete and continuous dropout times, and for settings with censored dropout times. In Section 4, we apply the proposed model to the clinical trial described in above. Section 5 presents a simulation study to evaluate the bias of the proposed method under departures from underlying assumptions. Summary and discussion follow in Section 6.

2 Mixtures of varying coefficient models for handling informative dropout

Suppose that the data consist of m subjects with the ith subject having n_i observations over time. For the ith subject, let $\boldsymbol{Y}_i$ be an $n_i \times 1$ observed outcome vector, $\boldsymbol{X}_i$ be an $n_i \times p$ covariate matrix associated with fixed effects, $\boldsymbol{Z}_i$ be an $n_i \times q$ covariate matrix associated with random effects, and U_i be the dropout time. The complete data distribution for the response vector is the mixture obtained by integrating the joint distribution $f(\boldsymbol{y}, u)$ over u. Mixture model approaches, therefore, require specification of $f(\boldsymbol{y} \mid u)$ and $f(u)$. When dropout times are discrete, it is usual to leave $f(u)$ unspecified and estimate it nonparametrically; for continuous u, it is possible but not always desirable to use a parametric model such as lognormal (DeGruttola and Tu, 1994; Schluchter, 1992). Our approach is to leave this marginal distribution unspecified.

To capture the dependence between $\boldsymbol{Y}$ and U, we assume that repeated measurements $\boldsymbol{Y}_i$ for those who drop out at u_i follow the varying coefficient random effects model

$$(\boldsymbol{Y}_i \mid U = u_i) = \boldsymbol{X}_i \boldsymbol{\beta}(u_i) + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{\epsilon}_i, \tag{1}$$

where $\boldsymbol{\beta}(u) = \{\beta_1(u), \ldots, \beta_p(u)\}^T$ is a $p \times 1$ vector of unknown regression coefficient functions of the dropout time u, $\boldsymbol{b}_i$ is a $q \times 1$ vector of random effects following $N\{\boldsymbol{0}, \boldsymbol{D}(\boldsymbol{\theta}, u_i)\}$, $\boldsymbol{\epsilon}_i$ is

an $n_i \times 1$ vector of residuals following $N\{\mathbf{0}, \boldsymbol{R}_i(\boldsymbol{\theta}, u_i)\}$, and $\boldsymbol{\theta}$ is a $c \times 1$ vector of variance components. Note that we allow the covariance matrices $\boldsymbol{D}$ and $\boldsymbol{R}$ to depend on the dropout time u_i. To better understand model (1), we write $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{ip})$, where $\boldsymbol{X}_{ij}$ is an $n_i \times 1$ vector of the values of the jth covariate measured over time for the ith subject. Equation (1) can be written as

$$(\boldsymbol{Y}_i \mid U = u_i) = \sum_{j=1}^{p} \boldsymbol{X}_{ij}\beta_j(u_i) + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i, \tag{2}$$

where $\beta_j(u)$ represents the jth covariate effect for those who drop out at time u.

In settings such as a panel design, where subjects are observed at prespecified finite time points (panels) and the underlying dropout times are discrete, we assume the $\beta_j(u)$ to be step functions (Little and Wang, 1996; Hogan and Laird, 1997a). If instead subjects are observed at different irregular time points and the underlying dropout times are continuous, we assume the $\beta_j(u)$ to be unspecified smooth functions. It follows that model (1) allows the covariate effects vary with dropout times nonparametrically and estimation of the covariate effects can be made more robust.

The varying coefficient specification (1) includes the pattern mixture model (Little, 1993, 1994), random effects pattern mixture model (Little, 1995; Hogan and Laird, 1997a), and conditional linear model (Wu and Bailey, 1989; Schluchter, 1992) as special cases. For example, if U has a discrete distribution with finite support, then the $\beta_j(u)$ are step functions and (1) is a pattern-mixture model. For continuous dropout times, consider the case where $\boldsymbol{X}_i = (\mathbf{1}, \boldsymbol{T}_i)$ and $\boldsymbol{T}_i = (t_{i1}, \ldots, t_{in_i})^T$, where t_{ik} is the kth followup time for the ith subject. If $\beta_1(u)$ and $\beta_2(u)$ are polynomial functions of u, model (1) reduces to a conditional linear model. If the $\beta_j(u)$ are constant in u, then the mixture distribution $f(\boldsymbol{y})$ has only one component and (1) reduces to a standard random effects model (Laird and Ware, 1982; Diggle, 1984).

Comparisons to selection models (e.g. Wu and Carroll, 1988; Diggle and Kenward, 1994) can be made by deriving the associated probability of dropout at time u as a function of $\boldsymbol{y}$. Let $\boldsymbol{u}^0 = (u_1^0, \ldots, u_{r-1}^0)^T$ denote the set of ordered, unique dropout times, and define an arbitrary time $u_r^0 > u_{r-1}^0$ for the completers. The selection function can be written in closed form as a logistic regression using baseline-category logits, where completers ($u = u_r^0$) define

the baseline category. Let $h_s(\boldsymbol{y}_i) = \text{pr}(U = u_s^0 \mid \boldsymbol{y}_i)/\text{pr}(U = u_r^0 \mid \boldsymbol{y}_i)$, which characterizes the odds of dropping out at time u_s^0 relative to completing the study, given repeated measures $\boldsymbol{y}_i$. Taking logs,

$$\log h_s(\boldsymbol{y}_i) = \log(\pi_r/\pi_s) + \log f(\boldsymbol{y}_i \mid U = u_s^0) - \log f(\boldsymbol{y}_i \mid U = u_r^0), \tag{3}$$

where $\pi_s = Pr(U_i = u_s^0)$, and

$$f(\boldsymbol{y}_i \mid U_i = u) = |\boldsymbol{V}_i(\boldsymbol{\theta}, u)|^{-n_i/2} \exp\left[-\left\{\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}(u)\right\}^T \boldsymbol{V}_i(\boldsymbol{\theta}, u)^{-1} \left\{\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}(u)\right\}/2\right],$$

with $\boldsymbol{V}_i(\boldsymbol{\theta}, u) = \boldsymbol{Z}_i\boldsymbol{D}_i(\boldsymbol{\theta}, u)\boldsymbol{Z}_i^T + \boldsymbol{R}_i(\boldsymbol{\theta}, u)$. In general, the selection model (3) is quadratic in $\boldsymbol{y}_i$. A more familiar version obtains when variance is constant across dropout times, i.e. when $\boldsymbol{V}_i(\boldsymbol{\theta}, u) = \boldsymbol{V}_i(\boldsymbol{\theta})$. Then (3) simplifies to

$$\log h_s(\boldsymbol{y}_i) = w\left\{\pi_s, \pi_r, \boldsymbol{\beta}(u_s^0), \boldsymbol{\beta}(u_r^0), \boldsymbol{V}_i(\boldsymbol{\theta})\right\} + \left\{\boldsymbol{\beta}(u_s^0) - \boldsymbol{\beta}(u_r^0)\right\}^T \boldsymbol{X}_i^T \boldsymbol{V}_i(\boldsymbol{\theta})^{-1}\boldsymbol{y}_i, \tag{4}$$

where $w(\cdot)$ comprises terms that do not depend on $\boldsymbol{y}_i$. Under this formulation, the log relative probability of dropout at u_s^0 is linear in $\boldsymbol{y}_i$, with coefficients that depend on the difference $\boldsymbol{\beta}(u_s^0) - \boldsymbol{\beta}(u_r^0)$, increasing in magnitude as $\boldsymbol{\beta}(u)$ depends more strongly on u. For binary U, (4) is consistent with the selection model restrictions used by Little (1994) and Little and Wang (1996) to identify parameters in a two-component pattern mixture model.

3 Estimation Procedures

In this section, we discuss estimation procedures when dropout times U_i are observed for all subjects. Under this circumstance, the likelihood of $(\boldsymbol{Y}_i, U_i)$ for the ith subject can be partitioned as

$$L_i(\boldsymbol{Y}_i, U_i; \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\pi}) = L_i(\boldsymbol{Y}_i \mid U_i; \boldsymbol{\beta}, \boldsymbol{\theta})\, L_i(U_i; \boldsymbol{\pi}).$$

It follows that $(\boldsymbol{\beta}, \boldsymbol{\theta})$ can be estimated by maximizing the conditional likelihood $\prod_{i=1}^m L_i(\boldsymbol{Y}_i \mid U_i; \boldsymbol{\beta}, \boldsymbol{\theta})$, and $\boldsymbol{\pi}$ can be estimated by maximizing the marginal likelihood $\prod_{i=1}^m L_i(U_i; \boldsymbol{\pi})$. We consider estimation procedures first for the situation where subjects are observed at a common set of time points, and then where the set of observation times may be misaligned across subjects.

3.1 Estimation procedures for fixed and common observation times

For a fixed design, such as a panel design, subjects are observed at n (often small) prespecified time points $(t_1, \ldots, t_n)$ and the number of possible dropout times is small. Thus, the nonparametric functions $\beta_j(u)$ are assumed to be step functions.

Let r be the number of observed distinct values of u, with $r \leq n$. As indicated in Section 2, let $\boldsymbol{u}^0 = (u_1^0, \ldots, u_r^0)^T$ be an $r \times 1$ vector of ordered distinct dropout times, where we assume completers take $u_r^0 = t_{n+1}$ for some value t_{n+1}. Hence $\beta_j(u)$ is fully determined by $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jr})^T$, where β_{jk} represents the jth covariate effect for those dropping out at u_k^0. Denote the incidence matrix by $\boldsymbol{N}_i = (N_{i1}, \ldots, N_{ir})^T$ where $N_{ik} = 1$ if $u_i = u_k^0$ and 0 otherwise. Some calculations show that model (2) can then be written as a linear mixed model

$$\boldsymbol{Y}_i = \sum_{j=1}^{p} \boldsymbol{X}_{ij} \boldsymbol{N}_i^T \boldsymbol{\beta}_j + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{\epsilon}_i.$$

Equivalently, we have

$$\boldsymbol{Y} = \sum_{j=1}^{p} \widetilde{\boldsymbol{X}}_j \boldsymbol{\beta}_j + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\epsilon}, \tag{5}$$

where $\widetilde{\boldsymbol{X}}_j = \boldsymbol{X}_j \otimes \boldsymbol{N}$, $\boldsymbol{X}_j = (\boldsymbol{X}_{1j}^T, \ldots, \boldsymbol{X}_{mj}^T)^T$, $\boldsymbol{N} = (\boldsymbol{N}_1^T, \ldots, \boldsymbol{N}_m^T)^T$, and $\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{b}$, and $\boldsymbol{\epsilon}$ are defined similarly to $\boldsymbol{N}$. Here $\boldsymbol{A} \otimes \boldsymbol{B}$ denotes a direct product of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$.

It follows that estimation of $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_p^T)^T$ proceeds by solving the linear mixed model normal equation

$$(\widetilde{\boldsymbol{X}}^T \boldsymbol{V}^{-1} \widetilde{\boldsymbol{X}}) \boldsymbol{\beta} = \widetilde{\boldsymbol{X}}^T \boldsymbol{V}^{-1} \boldsymbol{Y},$$

where $\widetilde{\boldsymbol{X}} = (\widetilde{\boldsymbol{X}}_1, \ldots, \widetilde{\boldsymbol{X}}_p)$, $\boldsymbol{V} = \mathrm{diag}(\boldsymbol{V}_i)$ and $\boldsymbol{V}_i = \mathrm{cov}(\boldsymbol{Y}_i) = \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^T + \boldsymbol{R}_i$. The covariance matrix of the estimator of $\boldsymbol{\beta}$ is $(\widetilde{\boldsymbol{X}}^T \boldsymbol{V}^{-1} \widetilde{\boldsymbol{X}})^{-1}$. Estimation of $\boldsymbol{\theta}$ can be obtained using the REML estimating equation

$$-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{P}\frac{\partial \boldsymbol{V}}{\partial \theta_j}\right) + \frac{1}{2}(\boldsymbol{Y} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta})^T \boldsymbol{V}^{-1} \frac{\partial \boldsymbol{V}}{\partial \theta_j} \boldsymbol{V}^{-1} (\boldsymbol{Y} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta}) = 0,$$

where $\boldsymbol{P} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\widetilde{\boldsymbol{X}}(\widetilde{\boldsymbol{X}}^T \boldsymbol{V}^{-1} \widetilde{\boldsymbol{X}})^{-1} \widetilde{\boldsymbol{X}}^T \boldsymbol{V}^{-1}$. The (j,k)th component of the information matrix of the estimator of $\boldsymbol{\theta}$ is $(1/2)\mathrm{tr}\{\boldsymbol{P}(\partial \boldsymbol{V}/\partial \theta_j)\boldsymbol{P}(\partial \boldsymbol{V}/\partial \theta_k)\}$. If $\boldsymbol{D}$ and $\boldsymbol{R}$ depend on u, the identifiability of some components of $\boldsymbol{\theta}$ based on the observed data might require some constraints on $\boldsymbol{\theta}$ or a sensitivity analysis (Little, 1993; Daniels and Hogan, 2000). For

example, consider the case with two time points ($n = 2$); if dropout is informative, there is no information in the observed data to fit a random intercept and slope model for those who drop out at time 2, and sensitivity analyses are needed.

The necessary and sufficient condition for the $\boldsymbol{\beta}_j$ to be identifiable from model (5) can be stated as follows:

> Let $\mathcal{S}_k = \{i_{k1}, \ldots, i_{km_k}\}$ denote the indexes of those m_k subjects who drop out at u_k^0, and $\mathcal{X}_k = (\boldsymbol{X}_{i_{k1}}^T, \ldots, \boldsymbol{X}_{i_{km_k}}^T)$. If $\text{rank}(\mathcal{X}_k) = p$ for all $k = 1, \ldots, r$, then the $\boldsymbol{\beta}_j$ $(j = 1, \ldots, p)$ are identifiable.

This condition states that when dropout times are discrete and finite, regression coefficients must be separately estimable for each dropout pattern. If some components of $\boldsymbol{\beta}_j$ are not identifiable for some dropout patterns, then the observed data do not have information about these components and either parameter constraints or sensitivity analyses are needed. For example, consider again the case with two time points ($n = 2$), if dropout is informative, there is no information in the observed data to estimate the mean of the outcome at time 2 for those who drop out at time 2, and sensitivity analysis is hence needed. See Little and Wang (1996) and Daniels and Hogan (2000) for detailed discussion.

Note that the jth regression coefficient vector $\boldsymbol{\beta}_j$ measures the jth covariate effect conditional on dropout times. Primary interest is in the marginal jth covariate effect $\widetilde{\beta}_j = E_U\left\{\boldsymbol{\beta}_j(u)\right\} = \boldsymbol{\pi}^T\boldsymbol{\beta}_j$, where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_r)^T$ and $\pi_j = P(U = u_k^0)$ is the probability that a subject drops out at u_k^0. It follows that one can estimate $\widetilde{\beta}_j$ by $\widehat{\boldsymbol{\pi}}^T\widehat{\boldsymbol{\beta}}_j$, where $\widehat{\boldsymbol{\pi}}_j = m_j/m$, m_j is the number of subjects who drop out at u_k^0, and $\widehat{\boldsymbol{\beta}}$ is the maximum likelihood estimator from fitting model (5).

3.2 Estimation for random or misaligned measurement times

For situations when the measurement times are random or misaligned across subjects, repeated measures of each subject are observed at different time points and the underlying dropout times are often continuous. We therefore assume that the regression coefficient functions $\beta_j(u)$ are twice-differentiable smooth functions and estimate them using cubic smoothing splines. A key feature of cubic smoothing spline estimation is that we can esti-

mate all model components — including the nonparametric regression coefficient functions $\beta_j(u)$, variance components $\boldsymbol{\theta}$ and smoothing parameters — within an augmented parametric linear mixed model framework.

Following the notation in Section 3.1, let r be the number of ordered distinct values of the U_i, $\boldsymbol{u}^0 = (u_1^0, \ldots, u_r^0)^T$ be the ordered distinct values of the U_i, and $\boldsymbol{\beta}_j = \{\beta_j(u_1^0), \ldots, \beta_j(u_r^0)\}^T$ be the values of $\beta_j(u)$ evaluated at $\boldsymbol{u}^0$. Given the variance components $\boldsymbol{\theta}$, the conditional loglikelihood of $\boldsymbol{\beta}_j$ given the u_i is

$$\begin{aligned}\ell(\boldsymbol{Y}|\boldsymbol{u};\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_p) &= \sum_{i=1}^{m} -\frac{1}{2}\ln|\boldsymbol{V}_i| - \frac{1}{2}\left\{\boldsymbol{Y}_i - \sum_{j=1}^{p}\boldsymbol{X}_{ij}\beta_j(u_i)\right\}^T \boldsymbol{V}_i^{-1}\left\{\boldsymbol{Y}_i - \sum_{j=1}^{p}\boldsymbol{X}_{ij}\beta_j(u_i)\right\} \\ &= -\frac{1}{2}\ln|\boldsymbol{V}| - \frac{1}{2}\left(\boldsymbol{Y} - \sum_{j=1}^{p}\widetilde{\boldsymbol{X}}_j\boldsymbol{\beta}_j\right)^T \boldsymbol{V}^{-1}\left(\boldsymbol{Y} - \sum_{j=1}^{p}\widetilde{\boldsymbol{X}}_j\boldsymbol{\beta}_j\right)\end{aligned}$$

where $\widetilde{\boldsymbol{X}}_j$, $\boldsymbol{V}_i$, $\boldsymbol{V}$, and $\boldsymbol{Y}$ were defined in Section 3.1.

Following O'Sullivan, Yandell and Raynor (1986), one can show that the natural cubic smoothing spline estimators of the $\beta_j(u)$ maximize the following penalized conditional loglikelihood

$$\ell(\boldsymbol{Y}|\boldsymbol{u};\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_p) - \frac{1}{2}\sum_{j=1}^{p}\lambda_j\int_{A_1}^{A_2}[\beta_j''(u)]^2du = \ell(\boldsymbol{Y}|\boldsymbol{u};\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_p) - \frac{1}{2}\sum_{j=1}^{p}\lambda_j\boldsymbol{\beta}_j^T\boldsymbol{K}\boldsymbol{\beta}_j, \quad (6)$$

where the λ_j are smoothing parameters controlling the balance between goodness-of-fit and the smoothness of the estimated $\boldsymbol{\beta}_j(u)$, A_1 and A_2 specify the range of u, and $\boldsymbol{K}$ is the nonnegative definite natural cubic smoothing spline smoothing matrix constructed using $\boldsymbol{u}_0$ and defined in equation (2.3) of Green and Silverman (1994).

For fixed smoothing parameters $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)^T$ and the variance components $\boldsymbol{\theta}$, differentiation of (6) with respect to $(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p)$ gives their estimating equations as

$$\begin{bmatrix} \widetilde{\boldsymbol{X}}_1^T\boldsymbol{V}^{-1}\widetilde{\boldsymbol{X}}_1 + \lambda_1\boldsymbol{K} & \widetilde{\boldsymbol{X}}_1^T\boldsymbol{V}^{-1}\widetilde{\boldsymbol{X}}_2 & \cdots & \widetilde{\boldsymbol{X}}_1^T\boldsymbol{V}^{-1}\widetilde{\boldsymbol{X}}_p \\ \vdots & \vdots & \vdots & \vdots \\ \widetilde{\boldsymbol{X}}_p^T\boldsymbol{V}^{-1}\widetilde{\boldsymbol{X}}_1 & \widetilde{\boldsymbol{X}}_p^T\boldsymbol{V}^{-1}\widetilde{\boldsymbol{X}}_2 & \cdots & \widetilde{\boldsymbol{X}}_p^T\boldsymbol{V}^{-1}\widetilde{\boldsymbol{X}}_p + \lambda_p\boldsymbol{K} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_p \end{bmatrix} = \begin{bmatrix} \widetilde{\boldsymbol{X}}_1^T\boldsymbol{V}^{-1}\boldsymbol{Y} \\ \vdots \\ \widetilde{\boldsymbol{X}}_p^T\boldsymbol{V}^{-1}\boldsymbol{Y} \end{bmatrix}. \quad (7)$$

One can solve equation (7) using a backfitting algorithm as follows

$$\widehat{\boldsymbol{\beta}}_j = (\widetilde{\boldsymbol{X}}_j^T\boldsymbol{V}^{-1}\widetilde{\boldsymbol{X}}_j + \lambda_j\boldsymbol{K})^{-1}\widetilde{\boldsymbol{X}}_j^T\boldsymbol{V}^{-1}(\boldsymbol{Y} - \sum_{k\neq j}\widetilde{\boldsymbol{X}}_k\boldsymbol{\beta}_k)$$

for $j = 1, \ldots, p$.

Following Zhang et al. (1998), we show that the above cubic smoothing spline estimators of the $\boldsymbol{\beta}_j$ can be easily obtained by fitting an augmented linear mixed model. Specifically, $\boldsymbol{\beta}_j$ can be written via a one-to-one transformation as

$$\boldsymbol{\beta}_j = \boldsymbol{U}\boldsymbol{\gamma}_j + \boldsymbol{B}\boldsymbol{a}_j,$$

where $\boldsymbol{U} = (\mathbf{1}, \boldsymbol{u}^0)$, $\boldsymbol{B} = \boldsymbol{L}(\boldsymbol{L}^T\boldsymbol{L})^{-1}$, $\boldsymbol{L}$ is an $r \times (r-2)$ full rank matrix satisfying $\boldsymbol{K} = \boldsymbol{L}\boldsymbol{L}^T$ and $\boldsymbol{L}^T\boldsymbol{U} = 0$, $\boldsymbol{\gamma}_j$ is a 2×1 unknown vector and $\boldsymbol{a}_j$ is a $(r-2) \times 1$ unknown vector. It can be shown that $\boldsymbol{\beta}_j^T\boldsymbol{K}\boldsymbol{\beta} = \boldsymbol{a}_j^T\boldsymbol{a}_j$. The penalized loglikelihood (6) becomes

$$\ell(\boldsymbol{Y}|\boldsymbol{U}; \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p) - \frac{1}{2}\sum_{j=1}^{p} \lambda_j \boldsymbol{a}_j^T\boldsymbol{a}_j.$$

It follows that the nonparametric natural cubic spline estimators of the $\boldsymbol{\beta}_j$ can be obtained by fitting the parametric linear mixed model

$$\boldsymbol{Y} = \sum_{j=1}^{p}(\widetilde{\boldsymbol{X}}_j\boldsymbol{U})\boldsymbol{\gamma}_j + \sum_{j=1}^{p}(\widetilde{\boldsymbol{X}}_j\boldsymbol{B})\boldsymbol{a}_j + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\epsilon}, \tag{8}$$

where $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^T, \ldots, \boldsymbol{\gamma}_p^T)^T$ is a vector of regression coefficients, $\boldsymbol{a} = (\boldsymbol{a}_1^T, \ldots, \boldsymbol{a}_p^T)^T$ and $\boldsymbol{b}$ are independent random effects with $\boldsymbol{a} \sim N(\mathbf{0}, \boldsymbol{\Lambda}(\boldsymbol{\tau}))$, $\boldsymbol{b} \sim N(\mathbf{0}, \mathrm{diag}\{\boldsymbol{D}(\boldsymbol{\theta})\})$, $\boldsymbol{\Lambda}(\boldsymbol{\tau}) = \mathrm{diag}(\tau_j\boldsymbol{I})$, $\tau_j = 1/\lambda_j$ and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_p)^T$, and $\boldsymbol{\epsilon}_i \sim N\{\mathbf{0}, \boldsymbol{R}_i(\boldsymbol{\theta}, u_i)\}$.

Estimation of $\boldsymbol{\gamma}$ and $\boldsymbol{a}$ can proceed using the BLUP estimator by solving the normal equation

$$\begin{bmatrix} \boldsymbol{H}^T\boldsymbol{V}^{-1}\boldsymbol{H} & \boldsymbol{H}^T\boldsymbol{V}^{-1}\boldsymbol{G} \\ \boldsymbol{G}^T\boldsymbol{V}^{-1}\boldsymbol{H} & \boldsymbol{G}^T\boldsymbol{V}^{-1}\boldsymbol{G} + \boldsymbol{\Lambda}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{a} \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}^T\boldsymbol{V}^{-1}\boldsymbol{Y} \\ \boldsymbol{G}^T\boldsymbol{V}^{-1}\boldsymbol{Y} \end{bmatrix}, \tag{9}$$

where $\boldsymbol{H} = (\widetilde{\boldsymbol{X}}_1\boldsymbol{U}, \ldots, \widetilde{\boldsymbol{X}}_p\boldsymbol{U})$ and $\boldsymbol{G} = (\widetilde{\boldsymbol{X}}_1\boldsymbol{B}, \ldots, \widetilde{\boldsymbol{X}}_p\boldsymbol{B})$. Denoting by $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{a}}$ the solution of (9), the natural cubic smoothing spline estimator of $\boldsymbol{\beta}_j$ is $\widehat{\boldsymbol{\beta}}_j = \boldsymbol{U}\hat{\boldsymbol{\gamma}}_j + \boldsymbol{B}\hat{\boldsymbol{a}}_j$. One can show that the $\widehat{\boldsymbol{\beta}}_j$ from (9) are identical to those obtained from solving (7). The natural cubic spline estimators $\widehat{\boldsymbol{\beta}}_j$ are unique when $\boldsymbol{H}$ is of full rank.

We have so far assumed that the smoothing parameters $\boldsymbol{\lambda}$ and the variance components $\boldsymbol{\theta}$ are known when estimating the $\boldsymbol{\beta}_j$. They are usually unknown in practice and need to be estimated from the data. Examination of the modified linear mixed model (8) suggests that

$\boldsymbol{\tau}$ behaves like variance components; therefore, following Zhang, et al. (1998), we estimate the smoothing parameters $\boldsymbol{\tau}$ and the variance components $\boldsymbol{\theta}$ simultaneously using restricted maximum likelihood (REML) by treating $\boldsymbol{\tau}$ as extra variance components in addition to $\boldsymbol{\theta}$ in (8). Some calculations show that the REML estimating equations for $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ are

$$\begin{aligned} -\frac{1}{2}\mathrm{tr}\left(\widetilde{\boldsymbol{P}}\frac{\partial \boldsymbol{V}}{\partial \theta_k}\right) + \frac{1}{2}(\boldsymbol{Y} - \sum_{j=1}^{p} \boldsymbol{X}_j\boldsymbol{\beta}_j)^T \boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial \theta_k}\boldsymbol{V}^{-1}(\boldsymbol{Y} - \sum_{j=1}^{p} \boldsymbol{X}_j\boldsymbol{\beta}_j) &= 0 \\ -\frac{1}{2}\mathrm{tr}\left(\widetilde{\boldsymbol{P}}\widetilde{\boldsymbol{X}}_j\boldsymbol{B}\boldsymbol{B}^T\widetilde{\boldsymbol{X}}_j^T\right) + \frac{1}{2}(\boldsymbol{Y} - \sum_{j=1}^{p} \boldsymbol{X}_j\boldsymbol{\beta}_j)^T \boldsymbol{V}^{-1}\widetilde{\boldsymbol{X}}_j\boldsymbol{B}\boldsymbol{B}^T\widetilde{\boldsymbol{X}}_j^T\boldsymbol{V}^{-1}(\boldsymbol{Y} - \sum_{j=1}^{p} \boldsymbol{X}_j\boldsymbol{\beta}_j) &= 0, \end{aligned}$$

where $\widetilde{\boldsymbol{P}} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}(\boldsymbol{H},\boldsymbol{G})\boldsymbol{C}^{-1}(\boldsymbol{H},\boldsymbol{G})^T\boldsymbol{V}^{-1}$, and $\boldsymbol{C}$ is the coefficient matrix on the left hand side of (9). Parameters from the varying coefficient mixture model (1) can therefore be obtained by fitting the parametric linear mixed model (8) by SAS Proc Mixed. The smoothing matrix $\boldsymbol{B}$ needs to be computed in advance.

3.3 Inference for marginal regression coefficients

The marginal jth covariate effect $\widetilde{\beta}_j$ is $\widetilde{\beta}_j = E_U\{\beta_j(u)\} = \int \beta_j(u)dF(u)$, where $F(u)$ is the c.d.f. of u. Using the estimated cubic smoothing spline $\widehat{\beta}_j(u)$ and the empirical c.d.f. $\widehat{F}(u)$, one can estimate $\widetilde{\beta}_j$ by $\int \widehat{\beta}_j(u)d\widehat{F}(u) = \widehat{\boldsymbol{\pi}}^T\widehat{\boldsymbol{\beta}}_j$, where $\widehat{\boldsymbol{\pi}} = \sum_{i=1}^{m} \boldsymbol{N}_i/m$. The delta method has been used for standard error estimation for mixture models where the support of U is very small relative to the number of subjects (Hogan and Laird, 1997a; Fitzmaurice and Laird, 2000). In our application with continuous dropout times, we found that the delta method performed poorly; the bootstrap was used instead, treating the subject as the basic resampling unit. Details are provided in the application.

4 Application to AIDS clinical trial

In this section we analyze the ACTG data using several different mixture model formulations representing different assumptions about the missing data mechanism and provide detailed interpretation of the model parameters.

4.1 Variable transformations and candidate models

As indicated in Section 1.2, observed CD4 data were transformed to the square root scale to reduce positive skewness. We fit three models (for computational tractability, the models were fit separately by dose). The first is a standard random effects model (REM) with subject-specific intercepts and linear time trends. The REM, briefly summarized in Section 1.2, assumes that the complete data in each treatment arm follow the linear mixed model

$$Y_{il} \quad = \quad \beta_1 + \beta_2 t^*_{il} + b_{1i} + b_{2i} t^*_{il} + \epsilon_{il}, \tag{10}$$

where Y_{il} is square root of CD4 count at time t_{il}, $l = 1, \ldots, n_i$, $\boldsymbol{b}_i = (b_{1i}, b_{2i}) \sim N(\boldsymbol{0}, \boldsymbol{D})$ and $\epsilon_{il} \sim N(0, \sigma^2)$ (i.e. $\boldsymbol{R}_i(\boldsymbol{\theta}) = \sigma^2 \boldsymbol{I}_{n_i}$), with independence between ϵ_{il} and $\boldsymbol{b}_i$. The time axis is rescaled using $t^*_{il} = (t_{il} - \overline{U})/\text{range}(t_{il})$ so that the new time scale has range 1 and is centered at the sample mean of dropout times (the range is computed for the pooled sample from both dose arms). This is a standard random effects model, but can be viewed as a special case of (1) where $\beta_j(u)$ is constant in u for $j = 1, 2$.

In addition to the REM, we fit a conditional linear model (CLM) in which individual intercepts and slopes are linear functions of dropout time, and a varying coefficient model (VCM), where intercepts and slopes are unspecified smooth functions of dropout time. The CLM is precisely model (8) under the assumption that $\boldsymbol{\tau} = \boldsymbol{0}$ (equivalently, $\boldsymbol{a}_j = \boldsymbol{0}$ for all j), and therefore it is just a specialized version of the VCM. Separately by treatment, the CLM elaborates (10) such that $\beta_1(u) = \gamma_1 + \gamma_2 u^*$ and $\beta_2(u) = \gamma_3 + \gamma_4 u^*$, where $u^* = (u - \overline{U})/\text{range}(t_{il})$. In our parameterization with the rescaled time axis, γ_1 is the mean of $\sqrt{\text{CD4}}$ at $u = \overline{U}$, γ_2 is the 'main effect' of dropout, which is the slope of $\beta_1(u)$, γ_3 is mean change in $\sqrt{\text{CD4}}$ from baseline for those with $U_i = \overline{U}$, and γ_4 is the effect of interaction between dropout and change from baseline, representing the slope of $\beta_2(u)$ on u. These parameters are estimated in straightforward fashion by fitting with SAS PROC MIXED a standard linear mixed model $Y_{ij} = \boldsymbol{H}_{ij}^T \gamma + \boldsymbol{Z}_{ij}^T \boldsymbol{b}_i + \epsilon_{ij}$, where $\boldsymbol{H}_{ij} = (1, U^*_{ij}, t^*_{ij}, t^*_{ij} U^*_{ij})^T$, and $\boldsymbol{Z}_{ij} = (1, t^*_{ij})^T$.

Rescaling the time axis has some practical advantages in terms of model fitting and interpretation: (i) it makes the estimates more stable by increasing the variance of individual

slopes away from zero; (ii) for the conditional linear model, the intercept and slope main effect are already averaged over dropout time (Fitzmaurice, Laird and Shneyer, 2001); (iii) for the REM and CLM, the slope parameter corresponds to average total change in square root CD4 from baseline to the longest follow up time for the study; (iv) for the CLM, the parameters γ_2 and γ_4 contrast mean intercept and slope, respectively, between those who drop out immediately (directly after $U = 0$) and those who complete the study protocol ($U = \max_{il} t_{il}$).

The VCM uses the same design matrix for $\boldsymbol{\gamma}$ as the CLM. Estimation of the $\boldsymbol{a}_j$ is made by specifying the $r - 2$ columns of $\widetilde{\boldsymbol{X}}_1\boldsymbol{B}$ and $\widetilde{\boldsymbol{X}}_2\boldsymbol{B}$ as independent random effects with respective variances τ_1^2 and τ_2^2.

4.2 Summary of fitted models

4.2.1 Parameter estimates

To get a crude understanding of whether MAR may be a valid assumption, we begin by summarizing regression coefficients for parameterizations of the conditional distribution $f(\boldsymbol{y} \mid u)$ given by the REM and CLM; these appear in Table 1. For both models, the parameter γ_3 quantifies average change from baseline to the maximum follow up time. Recall that in the CLM, $\beta_0(u) = \gamma_1 + \gamma_2(u^* - \overline{u^*})$ and $\beta_1(u) = \gamma_3 + \gamma_4(u^* - \overline{u^*})$, so that if we assume MAR is violated according to linear dependence between intercept and dropout and/or slope and dropout, the parameters γ_2 and γ_4 will quantify the degree to which MAR fails to hold. The parameter estimates from Model 1 suggest that intercepts vary considerably by dropout time ($\widehat{\gamma}_2 = 14.6$ and 13.0 [s.e. 3.3 and 3.5] for low dose and high dose, respectively); the same pattern is indicated for slopes, where $\widehat{\gamma}_4 = 26.1$ and 33.7 (s.e. 4.8 and 6.5) for low and high dose. These parameters are interpreted as the average difference in intercept (γ_2) and slope (γ_4) between those who dropped out immediately after enrolling and those who were followed for the maximum time (220 weeks). Hence the CLM indicates that dropouts have lower intercepts and slopes than completers, leading to selection bias for end-of-study comparisons due to missing data on less healthy participants. Because dropout time is centered at its mean in the CLM, $\widehat{\gamma}_3$ is the expected change in square root CD4 from week 0 to 220 if a subject were followed for the whole study period (note however that the standard

error is from the conditional distribution of $\boldsymbol{Y}$ given U). A comparison of the estimates of $\boldsymbol{\gamma}$ under REM and CLM suggests that the effect of accounting for dropout via CLM is to correct the average change in $\sqrt{\text{CD4}}$ downward.

In the VCM we allow both $\beta_1(u)$ and $\beta_2(u)$ to be completely unspecified for both treatment arms, and estimate them using cubic smoothing splines. Plots of the estimated functionals from the VCM, together with empirical Bayes estimates of individual intercepts and slopes, appear in Figure 3. In both treatment arms, baseline mean and CD4 slope are highly associated with dropout time: those who remain in the study for longer periods have higher values of both. Except for the baseline mean on the low dose arm, the associations appear to be highly nonlinear.

4.2.2 Comparison of treatment effect inferences across models

Because the complete data likelihood factors over $(\boldsymbol{\beta}, \boldsymbol{\theta})$ and $\boldsymbol{\pi}$, and because our three models differ only in the specification of the conditional factor $f(\boldsymbol{y} \mid \boldsymbol{u}; \boldsymbol{\beta}, \boldsymbol{\theta})$, model selection can in principle be based on criteria for the conditional $(\boldsymbol{y} \mid \boldsymbol{u})$ model. However, formal model selection procedures comparing parametric and nonparametric mixed effects models are not currently well developed and are beyond the scope of this paper; furthermore, there are potential difficulties associated with comparing likelihoods for models that are not properly nested in a traditional way, e.g., due to boundary-value problems.

Table 2 lists estimates and associated standard errors for the marginal regression coefficients $\widetilde{\beta}_j = \int \beta_j(u)\ dF(u)$, for $j = 1, 2$, estimated according to the procedure described in Section 3.3 (note that the marginal coefficients — and not the conditional coefficients reported in Table 1 — are of direct scientific interest). Standard errors were calculated using bootstrap resampling based on 100 replicated datasets sampled with replacement. Not surprisingly, standard errors associated with the VCM are increased relative to the CLM, reflecting uncertainty about the functional form of $\beta_j(u)$. For the low dose arm, the increase is relatively modest. Comparing slopes in the low-dose arm, for example, s.e.$(\widehat{\beta}_2) = 1.0$ for the CLM and 1.3 for the VCM. No appreciable difference in standard errors is seen in the estimated intercepts.

Table 2 also indicates the degree to which adjusting for selection bias affects the final

inferences about treatment. Under the MAR assumption (REM), estimated mean difference in total change in $\sqrt{\text{CD4}}$ is -5.5, with Z-statistic $= -3.4$; adjustment under the CLM gives the same estimated effect, with Z-statistic $= -2.8$. Both lead to the conclusion that low dose is superior to high dose because the decline in CD4 is less steep. Under the VCM, the correction for possible selection biases on low dose changes the slope estimate from -12.7 (REM) to -17.1; the correction is less severe on high dose (–18.2 for REM, compared to –20.1 for VCM); the effect is to narrow the gap in treatment effect to –3.0, with $Z = -1.1$, representing a change of 1.56 standard errors relative to the REM, and 1.25 standard errors relative to the CLM.

The VCM also provides an important substantive insight, namely that participants who drop out of low dose (the experimental dose in this trial) tend to have steeper decline in their CD4 counts, compared to those on high dose. The trial was designed to see whether the lower dose, known to be associated with fewer side effects in adults, would have efficacy equal to the high dose. The form of the $\beta_2(u)$ functions suggests that among the early dropouts, rate of change in CD4 for those on low dose is substantially less than for those on high dose. In an MAR analysis, early dropouts contribute less information to the estimate of population slope because they have fewer observed data points, leading to the potential selection bias seen in the REM.

5 Simulation study

Our model gives the analyst considerable flexibility in specifying dependence between outcome and dropout in the context of a mixture model, and avoids biases that are possible if the functional form of this dependence is assumed to be known. The primary innovation of the VCM over CLM is that $\boldsymbol{\beta}(u)$ can be left unspecified, but this generalization relies on the key assumption that $\boldsymbol{\beta}(u)$ is a (vector of) smooth, twice differentiable functions of u. We designed a brief simulation study to investigate the performance of our model under violations of this assumption.

Each simulation uses datasets with $n = 50$ subjects having up to 15 unique dropout times, with $\beta(u)$ taking three different forms. We compare estimates of mean change from baseline

from a standard random effects model, the conditional linear model with components of $\boldsymbol{\beta}(u)$ assumed linear in u, and from the varying coefficient model with $\boldsymbol{\beta}(u)$ left unspecified. Specifically, we assume

$$y_{il} = \beta_1(u_i) + \beta_2(u_i)t_{il} + b_{1i} + b_{2i}t_{il} + e_{il},$$

where $(b_{1i}, b_{2i})^T \sim N(\mathbf{0}, \boldsymbol{D})$, $e_{il} \sim N(0, \sigma^2)$. There are 15 time points $\{t_{il}\}$, equally spaced between 0 and 1. This simulation uses $d_{11} = 4, d_{22} = 0.1$, $d_{12} = -0.1$ (correlation ≈ -0.15), and $\sigma^2 = 1$, which implies that between-subject variation exceeds within-subject variation by a factor of about 4.

Dropout is generated from a beta mixture of binomial distributions as follows: $p \sim$ Beta$(1.5, 1.5)$ (mean 0.5); $U^* \sim$ Bin$(15, p)$, and dropout time $U = U^*/15 \in (0, 1)$. Finally, we assume $\beta_1(u) = 0$ and vary the functional form of $\beta_2(u)$; candidate functions are:

(i) $-\exp(\alpha u)$,

(ii) $\exp(\alpha u)I(u < t^*) + \exp(\alpha t^*)I(u \geq t^*)$
(exponential with plateau effect for dropouts beyond t^*),

(iii) $\alpha_1 I(u < t^*) + \alpha_2 I(u \geq t^*)$
(two-piece step function).

Case (i) actually meets the assumptions for the VCM, and is included for validating our simulation and estimation routines; case (ii) violates the smoothness assumption and case (iii) violates both smoothness and continuity assumptions.

For (i), at $\alpha = -4$, completers ($U = 1$) have mean change from baseline $\beta_2(1) = \exp(-4) \approx 0.02$ and early dropouts ($U = 0$) have mean change -1, a difference of about 3 SD (because $d_{22} = 0.1$). Under (ii), we keep $\alpha = -4$ and invoke the plateau effect at $t^* = 2/3$, leading to a structure wherein those who complete 2/3 of the study or more have average change from baseline equal to $\exp(-4 \times 2/3) \approx 0.07$. For (iii), we keep $t^* = 2/3$ and set $\alpha_1 = 0$, $\alpha_2 = 1$.

Results are reported in Table 3. As expected, the VCM gives virtually unbiased estimation of the true slope for case (i), where $\beta_2(u)$ is both continuous and smooth, while both

the REM and CLM show substantial upward bias. This comparison is not as trivial as it would appear, however, because exploratory plots of OLS slopes versus dropout time (e.g. Figure 2) do not always reveal an obvious functional form for $\beta(u)$, particularly in the early part of the time axis. One advantage to the VCM is its effectiveness in finding a signal from noisy data.

The VCM shows only very little bias for estimating the true slope for the continuous but not everywhere-differentiable function from case (ii), but exhibits more bias than the CLM for the discontinuous function in case (iii). In all cases, however, the VCM outperforms the random effects model.

6 Discussion

We have proposed a mixture-modeling approach to analyzing longitudinal data with outcome-dependent dropout. Our model assumes that covariate effects depend on dropout time through unspecified functions $\{\beta_j(u)\}$, where u is dropout time. When dropout times are discrete, the $\beta_j(u)$ are step functions, and when dropout is continuous, $\beta_j(u)$ are assumed to be unspecified smooth functions of u. This formulation generalizes pattern-mixture models (Little, 1993, 1994) and random-effects mixture models (Hogan and Laird, 1997a, Wu and Bailey, 1988, 1989, Follman and Albert, 2000) for continuous response data. Using an example from an AIDS clinical trial, we show that the model has the potential to adjust for selection biases induced by poor responders dropping out early.

The primary innovation in our approach is that the functional dependence between covariate effects can be left unspecified. Our simulation study shows that when the parametric form of the dependence of the regression coefficients on the dropout time is misspecified, conditional linear models yield biased estimates, while varying-coefficient mixture models still yield unbiased estimates and are more robust. In many applications, this is a decided advantage over the CLM model because the form of $\boldsymbol{\beta}(u)$ will rarely be known or intuitive. Moreover, it is our experience that using polynomials leads to overfitting and/or extrapolations well outside the range of data, particularly when the polynomial has degree greater than 2. When u is continuous, our simulation also shows that inferences are unlikely to be

sensitive to lack of smoothness in $\boldsymbol{\beta}(u)$, but could be affected by discontinuities. In both cases, however, bias is substantially less than under an MAR analysis.

Another advantage to mixture modeling in general is that extrapolations of the missing data are transparent, and lend themselves well both to substantive critique and to empirical sensitivity analysis (e.g. Rubin, 1977; Little and Wang, 1996; Daniels and Hogan, 2000). In our application, for example, we assume dropouts at time u have the same slope for $t > u$ as for $t \leq u$; on the surface this is a strong assumption but it is relatively easily modified and is a sensible starting point for sensitivity analysis. For example, one could perform a sensitivity analysis by varying the slope after the dropout time. It should be noted that the observed data do not contain any information about noninformative dropout mechanism. One needs either to make an assumption about what happens after the time of dropout or perform a sensitivity analysis. Because mixture models are specified in terms of the conditional distribution of $\boldsymbol{Y}$ given U, the extrapolation $E(\boldsymbol{Y}_{\text{mis}} \mid \boldsymbol{Y}_{\text{obs}}, U = u)$ is obvious (and in the mixture-of-normals model is linear in $\boldsymbol{Y}_{\text{obs}}$); by contrast, the same extrapolation for selection models is not obvious in most cases (but see Rotnitzky et al., 2001 for a discussion of this issue for semiparametric selection models). Further work will consider censoring in the dropout distribution, sensitivity analyses, and generalization to discrete distributions.

Acknowledgment

Work on this project was funded by grants R01-AI-50505 (JH, XL), and additionally by grant P30-AI-42853 (JH), both from the US Public Health Service. The authors are grateful to Rusty Tchernis for assistance with computing related to the simulation and implementation of the bootstrap, and to Jason Roy for helpful comments on an earlier draft.

References

Albert, P. S. and Follmann, D. (2000). Modeling repeated count data subject to informative dropout. *Biometrics* **56** 667–677.

Brady, M. T., McGrath, N., Brouwers, P., Gelber, R., Fowler, M. G., Yogev, R., Hutton, N., Bryson, Y. J., Mitchell, C. D., Fikrig, S., Borkowsky, W., Jimenez, E., McSherry, G., Rubinstein, A., Wilfert, C. M., McIntosh, E., Elkins, M. M. and Weintrub, P.S., for the Pediatric AIDS clinical trial. (1996). Randomized study of the tolerance and efficacy of high-versus low-dose zidovudine in human immunodeficiency virus-infected children with mild to moderate symptoms(ACTG 128). *Journal of Infectious disease* **173**, 1097–1106.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25.

Daniels, M. J. and Hogan, J. W.(2000). Reparameterizing the pattern mixture model for sensitivity analyses under informative dropout. *Biometrics* **56**, 1241–1248.

DeGruttola, V. and Tu, X. M. (1994). Modeling the Progression of CD4-Lymphocyte Count and Its Relationship to Survival Time. *Biometrics* **50**, 1003–1014.

Diggle, P. J. (1988). An Approach to the Analysis of Repeated Measurements. *Biometrics* **44**, 959–971.

Diggle, P. and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Applied Statistics* **43**, 49–73.

Fitzmaurice, G. M. and Laird, N. M. (2000). Generalized linear mixture models for handing nonignorable dropouts in longitudinal studies. *Biostatistics* **1**, 141–156.

Fitzmaurice, G. M., Laird, N. M., and Shneyer, L. (2001). An alternative parameterization of the general linear mixture model for longitudinal data with non-ignorable drop-outs. *Statistics in Medicine* **20** 1009–1021.

Follman, D. and Wu, M. C. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics* **51**, 151–168.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach.* Chapman & Hall (London).

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 analysing incomplete longitudinal and failure time data. *Statistics in Medicine* **16**, 259–272.

Kenward, M. G. , and Molenberghs, G. (1999). Parametric models for incomplete continuous and categorical longitudinal data. *Statistical Methods in Medical Research* **8** 51–83.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.

Laird, N. M. (1988). Missing data in Longitudinal Studies. *Statistics in Medicine* **7**, 305–315.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* **88**, 125–134.

Little, R. J. A. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika* **81**, 471–483.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* **90**, 1112–1121.

Little, R. J. A., and Wang, Y. (1996). Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics* **52**, 98–111.

O'Sullivan, F., Yandell, B. S. , and Raynor, W. J. Jr (1986). Automatic smoothing of regression functions in generalized linear models. *Journal of the American Statistical Association* **81**, 96–103.

Robins, J., Rotnitsky, A. and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **90**, 106–121.

Rotnitzky, A , Scharfstein, D., Su, T.L., and Robins, J. (2001). Methods for conducting sensitivity analysis of trials with potentially nonignorable competing causes of censoring. *Biometrics* **57**, 103–113.

Rotnitzky, A., Robins, J. M. and Scharfstein, D. O. (1998). Semiparametric Regression for Repeated outcomes with Non-Ignorable Non Response. *Journal of the American Statistical Association* **93**, 1321–1339.

Rubin, D. B. (1977). Formalizing subjective notions about the effect of nonrespondents in sample surveys. *Journal of the American Statistical Association* **72**, 538-543

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.

Schluchlter, M. D. (1992). Methods for the Analysis of Informatively Censored Longitudinal Data. *Statistics in Medicine* **11**, 1861–1870.

Ten Have, T. R., Kunselman, A. R., Pulkstenis, E. P. and Landis, J. R. (1998). Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics* **54**, 367–383

Wu, M. C. and Bailey, K. (1988). Analysing changes in the presence of informative right censoring caused by death and withdrawl. *Statistics in Medicine* **7**, 337–346.

Wu, M. C. 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. C. and Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44**, 175–188.

Zhang, D., Lin, X., Raz, J. and Sowers M. (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association* **93**, 710–719.

Table 1: Parameter estimates from conditional part of joint model, assuming linear random effects structure (REM) and conditional linear model (CLM).

		AZT Dose	
Model	Parameter*	Low (90mg)	High (180mg)
REM	γ_1	28.6 (0.8)	30.1 (0.9)
	γ_3	–12.7 (0.8)	–18.2 (1.4)
CLM	γ_1	28.9 (0.8)	30.3 (0.8)
	γ_2	14.6 (3.3)	13.0 (3.5)
	γ_3	–15.9 (0.9)	–21.4 (1.4)
	γ_4	26.1 (4.8)	33.7 (6.5)

* See Section 4 for definitions.

Table 2: Estimated intercept and slope characterizing marginal mean of CD4 trajectory under three different specifications for conditional part of joint model, with standard errors estimated via bootstrap.

Model	Parameter	Low Dose	High Dose	Difference (s.e.)	Z
REM	β_1	28.6 (0.8)	30.1 (0.9)		
	β_2	–12.7 (0.8)	–18.2 (1.4)	–5.5 (1.6)	–3.4
CLM	β_1	28.9 (0.8)	30.3 (0.9)		
	β_2	–15.9 (1.0)	–21.4 (1.8)	–5.5 (2.0)	–2.8
VCM	β_1	29.0 (0.7)	29.9 (0.8)		
	β_2	–17.1 (1.3)	–20.1 (2.5)	–3.0 (2.8)	–1.1

Table 3: Results from simulation to characterize bias. REM = linear random effects model; CLM = conditional linear model with $\beta_2(u)$ linear; VCM = varying coefficient model with $\beta_2(u)$ unspecified. Each estimate represents a sample average of estimated slopes over 100 replicated datasets, each having 100 subjects with up to 15 repeated measures. Standard errors for simulation-based estimated mean appear in parentheses.

		Estimated Slope		
Underlying model*	True Slope (β_2)	REM	CLM	VCM
(i) Continuous, smooth	–0.159**	–0.073	–0.119	–0.160
		(0.018)	(0.035)	(0.028)
(ii) Continuous, not smooth	–0.170**	–0.062	–0.100	–0.166
		(0.015)	(0.027)	(0.033)
(iii) Discontinuous	–0.587**	–0.211	–0.622	–0.715
		(0.018)	(0.031)	(0.038)

* See Section 5 for model descriptions
** Computed to nearest 0.001 via Monte-Carlo simulation

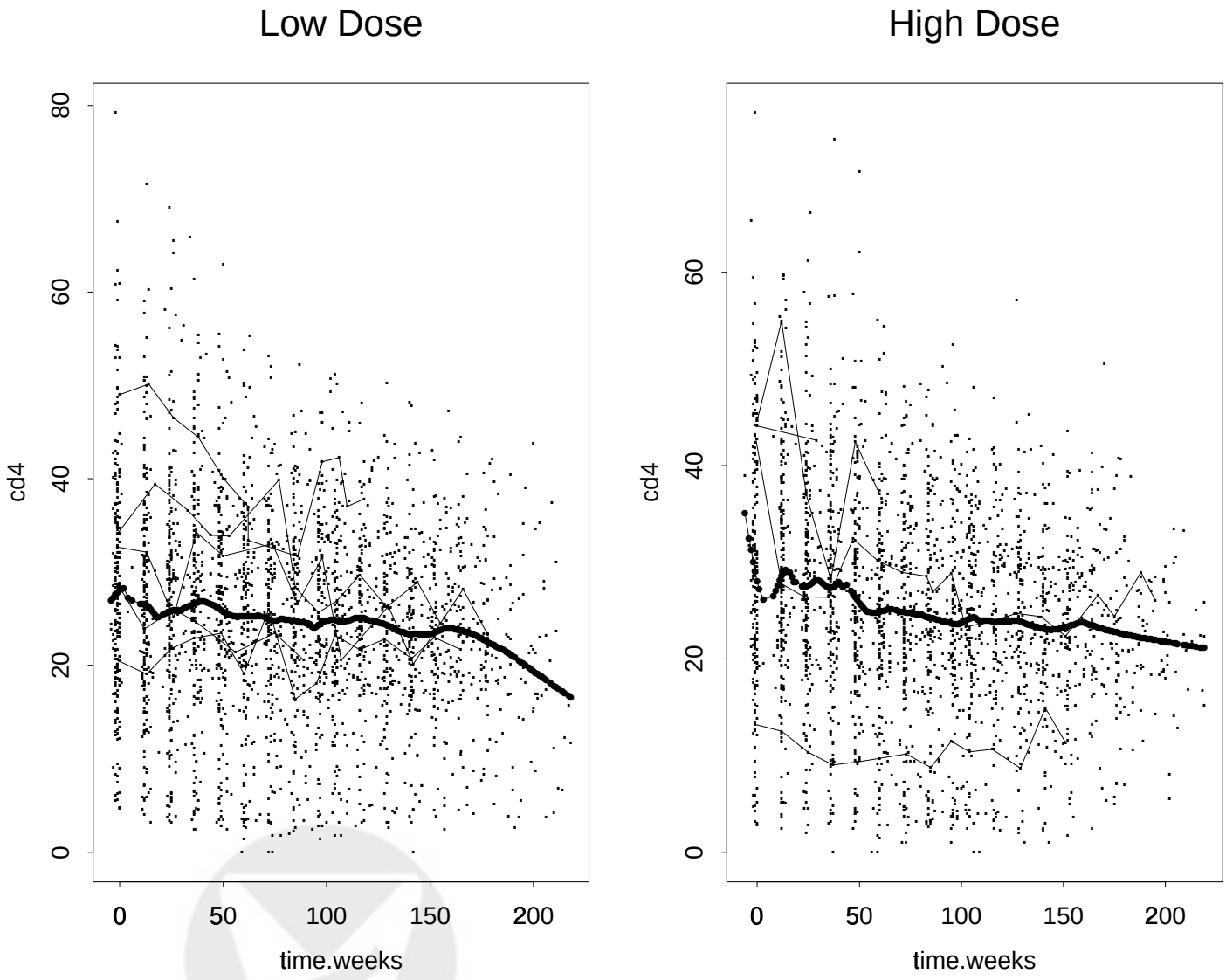

Figure 1: Observed (square root) CD4 counts versus time and stratified by dose, with lowess regression line fit to pooled sample and five individual profiles highlighted.

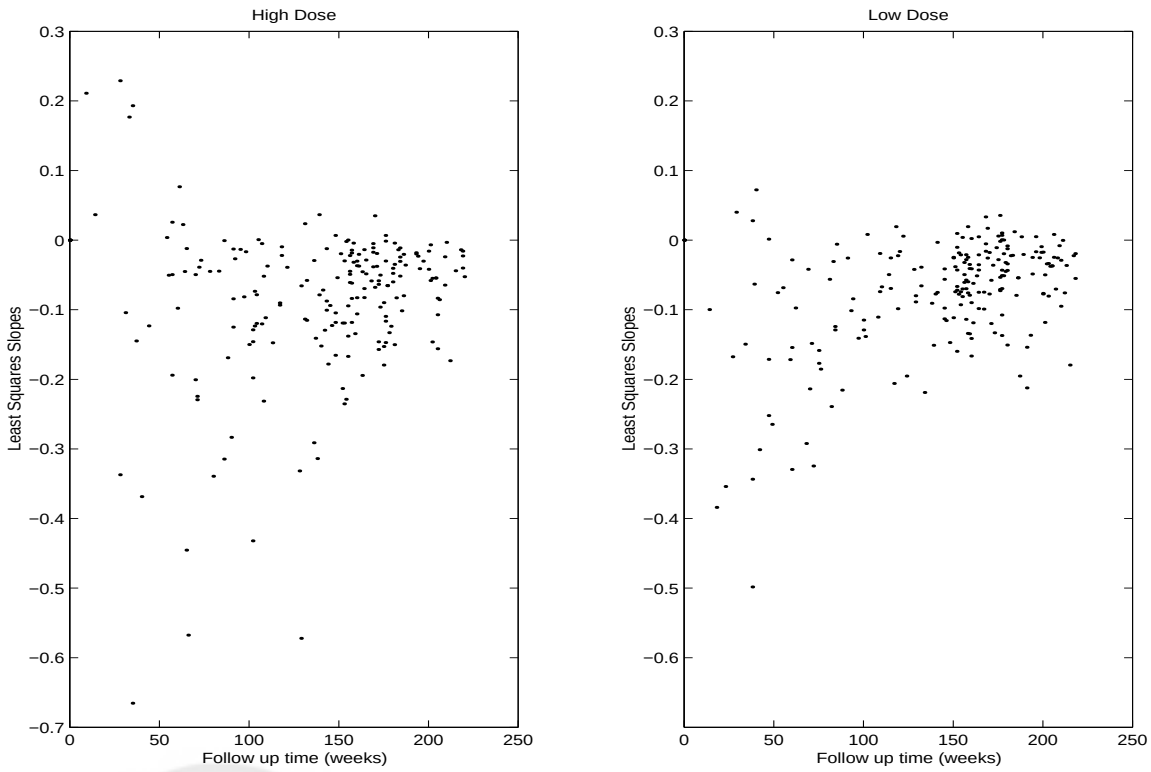

Figure 2: Individual-specific OLS slopes for square-root CD4 versus as a function of follow-up time, stratified by dose.

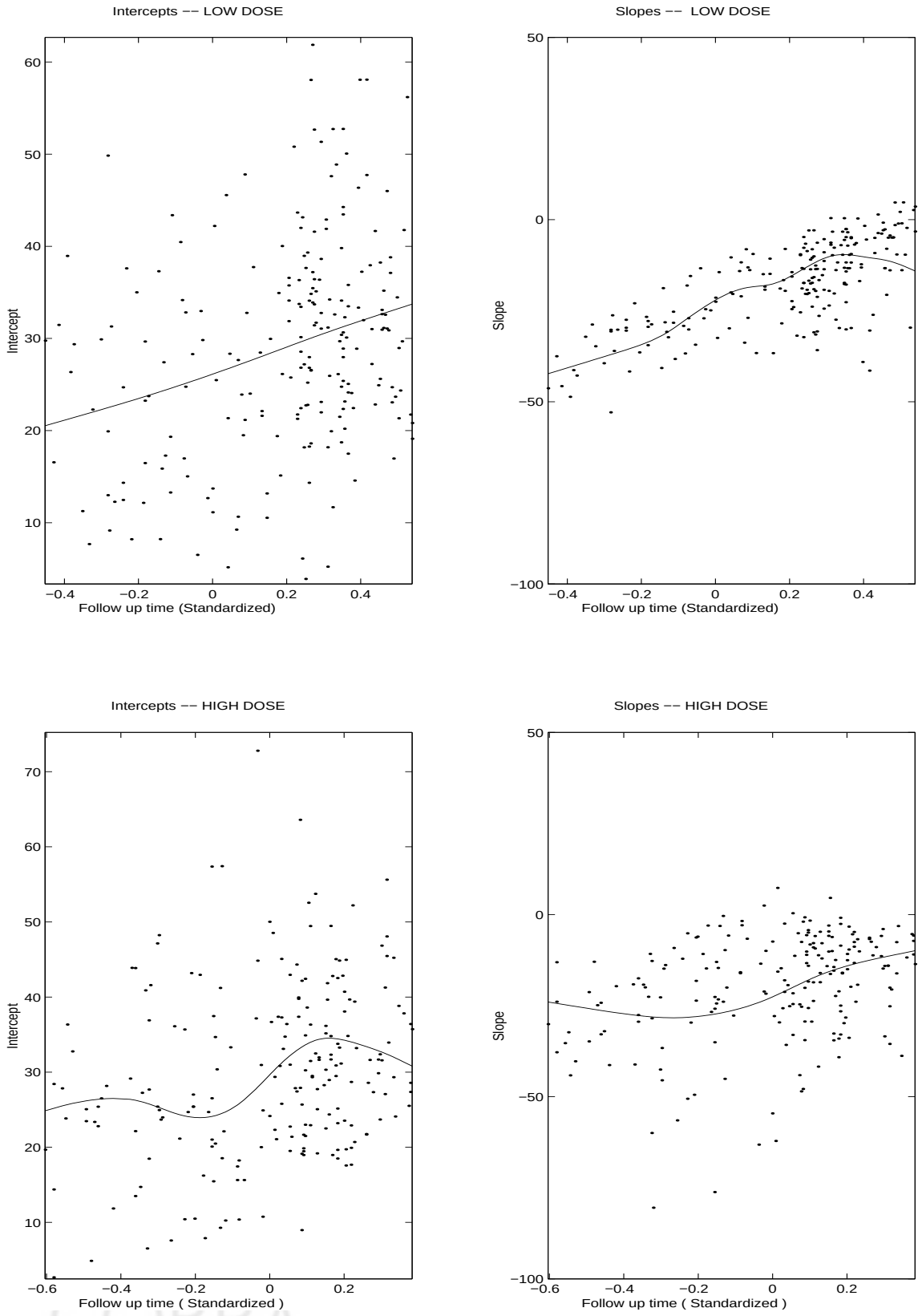

Figure 3: Estimated functions $\beta_1(u)$ and $\beta_2(u)$ for low- and high-dose ZDV arms, together with empirical Bayes estimates of individual intercepts and slopes (on square root CD4 scale). Slopes correspond to change from baseline to week 200. Standardized follow up times correspond to deviation from the average from the combined sample, and one unit represents 200 weeks.