

# Innovative Applications of Shared Random Parameter Models for Analyzing Longitudinal Data Subject to Dropout

Paul S. Albert, Rajeshwari Sundaram, and Alexander C. McLain

**Abstract** Shared random parameter (SRP) models provide a framework for analyzing longitudinal data with missingness. We discuss the basic framework and review the most relevant literature for the case of a single outcome followed longitudinally. We discuss estimation approaches, including an approximate approach which is relatively simple to implement. We then discuss three applications of this framework in novel settings. First, we show how SRP models can be used to make inference about pooled or batched longitudinal data subject to non-ignorable dropout. Second, we show how one of the estimation approaches can be extended for estimating high dimensional longitudinal data subject to dropout. Third, we show how to use jointly model complex menstrual cycle length data and time to pregnancy in order to study the evolution of menstrual cycle length accounting for non-ignorable dropout due to becoming pregnant and to develop a predictor of time-to-pregnancy from repeated menstrual cycle length measurements. These three examples demonstrate the richness of this class of models in applications.

## 1 Introduction

Modeling longitudinal data subject to missingness has been an active area of research in the last few decades. The missing-data mechanism is said to be missing completely at random if the probability of missing is independent of both the observed and unobserved data. Further, the mechanism is not missing

---

P.S. Albert (✉) • R. Sundaram

*Eunice Kennedy Shriver*, National Institute of Child Health and Human Development,  
Bethesda, MD, USA

e-mail: [albertp@mail.nih.gov](mailto:albertp@mail.nih.gov); [Sundaramr2@mail.nih.gov](mailto:Sundaramr2@mail.nih.gov)

A.C. McLain

University of South Carolina, Columbia, SC, USA

e-mail: [mclaina@mailbox.sc.edu](mailto:mclaina@mailbox.sc.edu)

at random (NMAR) if the probability of missingness depends on the unobserved data (Rubin 1976; Little and Rubin 1987). It is well known that naive methods that do not account for NMAR can lead to biased estimation. The use of shared (random) parameter models has been one approach that accounts for non-random missing data. In this formulation, a model for the longitudinal response measurements is linked with a model for the missing-data mechanism through a set of random effects that are shared between the two processes. Wu and Carroll (1988) proposed a model whereby the response process, which was modeled with a linear mixed model with a random intercept and slope was linked with the censoring process by including an individual's random slope as a covariate in a probit model for the censoring process. When the probit regression coefficient for the random slope is not zero, there is a dependence between the response and missing-data processes. Failure to account for this dependence can lead to biased estimation of important model parameters. Shared-parameter models (Follmann and Wu 1995) induce a type of non-randomly missing-data mechanism that has been called "informative missingness" (Wu and Carroll 1988). For a review and comparison with other methods, see Little (1995), Hogan and Laird (1997), and Vonesh et al. (2006). More recently Molenberghs et al. (2012) have discussed a fundamental non-identifiability of shared random effects models. Specifically, these models make non-verifiable assumptions about data not seen and there are multiple models in a wide class that can equally explain the observed data. Thus, shared random parameter (SRP) models make implicit assumptions that need to be justified from on a scientific basis and cannot be completely verified empirically.

This article discusses some applications of SRPs to some interesting novel applications. In Sect. 2, we set up the general model formulation and show how this mechanism induces a special type of nonignorable missingness. We also discuss both a full maximum-likelihood approach and conditional approach for parameter estimation that is easier to implement. We discuss some examples where a single longitudinal measurement is subject to non-ignorable dropout. Section 3 shows an example of batched laboratory data and how a SRP model can be used to account for the apparent non-ignorable missingness. In Sect. 4 we provide an example of the joint modeling of multiple or high dimensional longitudinal biomarker and time-to-event data. Section 5 shows an example of jointly modeling complex menstrual cycle data and time-to-pregnancy using a SRP approach. Lastly, we present a discussion in Sect. 6.

## 2 Model Formulation and Estimation

Let  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iJ})'$  be a vector of longitudinal outcomes for the  $i$ th subject ( $i = 1, 2, \dots, I$ ) observed on  $J$  occasions  $t_1, t_2, \dots, t_J$ , and let  $\mathbf{R}_i = (R_{i1}, R_{i2}, \dots, R_{iJ})'$  be a vector of random variables reflecting the missing data status (e.g.,  $R_{ij} = 0$  denoting a missed visit). Further, let  $\mathbf{b}_i = (b_{i1}, b_{i2}, \dots, b_{iL})'$  be an  $L$ -element vector of random effects for the  $i$ th subject which can be shared between the response and

missing data mechanism. We assume that  $b_i$  is multivariate normal with mean vector 0 and covariance matrix  $\Sigma_b$ . Covariates  $X_{ij}$  are also measured which can influence both  $Y_{ij}$  and  $R_{ij}$ .

The joint distribution of  $Y_i, R_i, b_i$  can be written as

$$f(y_i, r_i, b_i) = g(y_i|b_i, r_i)m(r_i|b_i)h(b_i).$$

We make the assumption that conditional on the random effects, the responses do not depend on the missing data status, thus  $g(y_i|b_i, r_i) = g(y_i|b_i)$ . Furthermore, the elements of  $Y_i$  are conditionally independent given  $b_i$ . By conditional independence, the density for the response vector  $Y_i$  conditional on  $b_i$ ,  $g(y_i|b_i)$  can be decomposed into the product of the densities for the observed and unobserved values of  $Y_i$ . Namely,  $g(y_i|b_i) = g(y_i^o|b_i)g(y_i^m|b_i)$ , where  $y_i^o$  and  $y_i^m$  are vectors of observed and missing data responses, respectively, for the  $i$ th subject. The density of the observed random variables can be expressed as

$$\begin{aligned} f(y_i^o, r_i) &= \int_{y_i^m} \int_b f(y_i^o, y_i^m, r_i) db dy_i^m \\ &= \int_{y_i^m} \int_b g(y_i^o|b_i)g(y_i^m|b_i)m(r_i|b_i)h(b_i) db dy_i^m \\ &= \int_b g(y_i^o|b_i)m(r_i|b_i)h(b_i) \left\{ \int_{y_i^m} g(y_i^m|b) dy_i^m \right\} db_i \\ &= \int_b g(y_i^o|b_i)m(r_i|b_i)h(b_i) db_i. \end{aligned} \tag{1}$$

Although the conditional independence of  $Y_i|b_i$  is easy to verify when there is no missing data, it is difficult to verify for SRP models. Serial correlation could be incorporated (conditional on the random effects) using autoregressive or lagged responses (see, [Zeger and Qaqish 1988](#); [Albert 2000](#); [Sutradhar and Mallick 2010](#)). These lag-response modeling components can be formulated with the addition of a shared random effect that links the response and missing data mechanism together. Alternatively, [Albert et al. \(2002\)](#) link together the response and missing data mechanism with a shared latent process where the subject-specific random effect  $b$  is replaced by a random process  $b_i = (b_{i1}, b_{i2}, \dots, b_{iJ})'$ . They consider a random process that follows a continuous-time exponential correlation structure since observations are not equally spaced. Although the shared latent processes model is an attractive approach, it requires computationally intensive techniques such as Monte-Carlo EM for parameter estimation. In the remainder of this article, we focus on the SRP rather than the shared latent process model.

[Tsiatis and Davidian \(2004\)](#) provide a concise discussion of how the joint density is obtained for the case where missingness is monotone (i.e., patients only drop out of the study) and measured in continuous time.

The choice of a density function  $g$  depends on the type of longitudinal response data being analyzed. For Gaussian longitudinal data,  $g$  can be specified as a Gaussian distribution, and the model formulation can be specified as a linear mixed model

(Laird and Ware 1982). A simple linear mixed model which can be used as an illustration is

$$Y_{ij}|X_i, b_i = \beta_0 + \beta_1 X_i + b_i + \varepsilon_{ij}, \quad (2)$$

where  $X_i$  is a subject-specific covariate such as treatment group,  $b_i$  is a random effect which is often assumed normally distributed, and  $\varepsilon_{ij}$  is an error term which is assumed normally distributed. Alternatively, for discrete or dichotomous longitudinal responses,  $g$  can be formulated as a generalized linear mixed model (Follmann and Wu 1995; Ten Have et al. 1998; Albert and Follmann 2000).

The choice of the density for the missing data indicators,  $m$ , depends on the type of missing data being incorporated. When missing data is a discrete time to dropout, a monotone missing data mechanism, then a geometric distribution is often used for  $m$  (Mori et al. 1994). For example, the probability of dropping out is

$$\Phi^{-1}\{P(R_{ij} = 0|R_{ij} = 1)\} = \alpha_0 + \alpha X_i + \theta b_i. \quad (3)$$

Various authors have proposed shared random effects models for the case in which dropout is a continuous event time (Schluchter 1992; Schluchter et al. 2001; DeGruttola and Tu 1994; Tsiatis et al. 1995; Wulfson and Tsiatis 1997; Tsiatis and Davidian 2001; Vonesh et al. 2006). When missing data includes only intermittently missed observations without dropout, then the product of Bernoulli densities across each of the potential observations may be a suitable density function for  $g$ . Alternatively, when multiple types of missing data such as both intermittent missingness and dropout need to be incorporated, a multinomial density function for  $g$  can be incorporated (Albert et al. 2002).

The shared random effects model accounts for a MNAR data mechanism, which can be seen with the following argument. Suppressing the index  $i$  for notational simplicity, suppose that the random effect  $b$  is a scalar with  $R_j$  indicating whether  $Y_j$  is observed. MAR implies that the conditional density of  $R_j$  given the complete data  $Y$  does not depend on  $Y^m$ , while a MNAR implies that this conditional density depends on  $Y^m$ . The conditional density of  $R_j$  given  $Y = (Y^o, Y^m)$  is

$$f(r_j|y^m, y^o) = \frac{\int g(r_j|b)g(y^m, y^o|b)h(b)db}{\int g(y^m, y^o|b)h(b)db} \\ \int g(r_j|b)h(b|y^o, y^m)db.$$

A MNAR data mechanism follows since the conditional density depends on  $y^m$  and since  $h(b|y^o, y^m)$  depends on  $y^m$ . It is interesting to note that for models (2) and (3) when the residual variance is very small, error is very small,  $h(r_j|y^m, y^o) \approx h(r_j|y^o)$  since  $h(b|y^m, y^o) \approx h(b|y^o)$ . In this situation, the missing data mechanism will be close to MAR, so simply fitting a likelihood-based model for  $y^o$  will result in valid inference.

Albert and Follmann (2007) discuss various SRP modeling formulation for analyzing binary longitudinal data with applications to an opiates clinical trial.

There are various approaches for parameter estimation. First, maximization of the likelihood  $L = \prod_{i=1}^I f(y_i^o, r_i)$ , where  $f$  is given by (1) can be used to obtain the maximum-likelihood estimates (MLEs). Maximizing the likelihood may be computationally intensive since it involves integrating over the random effects distribution. For a high dimensional random effects distribution, this involves the numerically difficult evaluation of a high dimensional integral. Approaches such as Monte-Carlo EM or Laplace approximations of the likelihood (Gao 2004) may be good alternatives to direct evaluation of the integral. Fortunately, many applications involve only one or two shared random effects where the integral can be evaluated more straightforwardly with Gaussian quadrature, adaptive Gaussian quadrature, or other numerical integration techniques. Various statistical software packages can be used to fit these models including procedures in SAS and specialized code in R.

An alternative approach for parameter estimation, which conditions on  $R_i$ , has been proposed (Wu and Carroll 1988; Follmann and Wu 1995; Albert and Follmann 2000). In developing this approach, first note that the joint distribution of  $(Y_i^o, R_i, b_i)$  can be re-written as

$$\begin{aligned} f(y_i^o, r_i, b_i) &= f(y_i^o, b_i | r_i) m(r_i) \\ &= f(y_i^o | b_i, r_i) h(b_i | r_i) m(r_i) \\ &= g(y_i^o | b_i) h(b_i | r_i) m(r_i). \end{aligned}$$

Thus, the conditional likelihood for  $y_i^o | r_i$  is given by  $L = \prod_{i=1}^I \int g(y_i^o | b_i) h(b_i | r_i) db_i$ . Note that this approximate conditional model can be directly viewed as a pattern mixture model as

$$\begin{aligned} f(y^o, r) &= \int g(y^o | b) h(b | r) db m(r) \\ &= p(y | r) m(r), \end{aligned}$$

Little (1993).

For illustration, we can estimate the treatment effect  $\beta_1$  in the non-random dropout model (2) and (3) by noting that  $b_i | d_i$  can be approximated by a normal distribution with mean  $\omega_0 + \omega_1 d_i$ . The conditional model can then be characterized by a linear mixed model of the form  $Y_{ij} = \beta_0^* + \omega_1 d_i + \beta_1^* + b_i + \varepsilon_{ij}$ . An important point is that the parameters of this model are conditional on the dropout time  $d_i$  and are easily interpretable. What is of interest are inferences on the marginal distribution of  $Y_{ij}$ . To estimate  $\beta_1$  in model (2), we need to marginalize over the dropout time distribution. Specifically,  $E(Y|x) = E(E(Y|d,x)) = \beta_0^* + \beta_1^* x + \omega_1 E(d|x)$ , where  $E(d|x)$  is the conditional distribution of dropout given the covariate  $x$ . We can estimate this conditional distribution in a two group comparison with  $\hat{E}(d|x) = \bar{d}_x$ . Thus,  $\beta_1 = E(Y|x = 1) - E(Y|x = 0) = \beta_1^* + \omega_1 (\bar{d}_1 - \bar{d}_0)$ . Variance estimation for the MLE approach can be obtained through standard asymptotic techniques (i.e., inverting the observed Fisher information matrix). For the conditional modeling

approach, the simplest approach is to perform a bootstrap procedure (Efron and Tibshirani 1993) where all measurements on a chosen individual are sampled with replacement.

### 3 An Analysis of Longitudinal Batched Gaussian Data Subject to Non-random Dropout

Due to cost or feasibility, longitudinal data may be measured in pools or batches whereby samples are combined or averaged across individuals at a given time point. In these studies, interest may be on comparing the longitudinal measurements across groups. A complicating factor may be that subjects are subject to dropout from the study. An example of this type of data structure is a large mouse study examining the effect of an experimental antioxidant on the weight profiles over time in mice. It was suspected that animals receiving the treatment would reach a lower adult weight than control animals and that the decline in weight among treated animals would be less than that for control animals. Ninety-five genetically identical animals were enrolled into the treatment and control groups, respectively (190 total animals). Within a group, five animals were placed in each of the 19 cages at birth. Due to the difficulty in repeatedly weighting each animal separately, the average weight per cage was recorded at approximately bi-weekly intervals over the life span of the animals (2–3 years). At each follow-up time, average batch weight was measured as the total batch weight divided by the number of animals alive in that batch.

Albert and Shih (2011) proposed a SRP model for each of the two groups separately. Initially, we present the model when individual longitudinal data are observed and then develop the model for batched longitudinal data. Denote  $Y_{ij}$  as the  $j$ th longitudinal observation at time  $t_j$  for the  $i$ th subject. As described in Sect. 2, for a dropout process where an individual dies between the  $(d-1)$ th and the  $d$ th time point,  $R_{i1} = R_{i2} = \dots = R_{id-1} = 1$  and  $R_{id} = 0$ . The dropout time for the  $i$ th subject is denoted as  $d_i$ .

We assume a linear mixed model in each group of the form

$$Y_{ij} = \beta_0 + \beta_1 t_j + b_{0i} + b_{1i} t_j + \varepsilon_{ij}, \quad (4)$$

where  $b_i = (b_{0i}, b_{1i}) \sim N(0, \Sigma_b)$  and  $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$  is independent of  $b_i$ . Further, we denote

$$\Sigma_b = \begin{pmatrix} \sigma_{b_0}^2 & \sigma_{b_0 b_1} \\ \sigma_{b_0 b_1} & \sigma_{b_1}^2 \end{pmatrix}.$$

Model (4) can be made more general by including a change point or additional polynomial functions of time to the fixed and random effects. Similar to (2) and (3), the dropout mechanism can be modeled with a geometric distribution in

which, conditional on the random intercept  $b_{0i}$  and random slope  $b_{1i}$ ,  $R_{ij}|(R_{ij-1} = 1, b_{0i}, b_{1i})$  is Bernoulli with probability

$$P(R_{ij} = 0 | R_{ij-1} = 1, b_{0i}, b_{1i}) = \Phi(\alpha(t_j) + \theta_1 b_{0i} + \theta_2 b_{1i}), \tag{5}$$

where  $\alpha(t_j)$  is a function of follow-up time  $t_j$ . As discussed in Sect. 2, incorporating SRPs between the response and dropout process induces a non-ignorable dropout mechanism.

For the longitudinal animal study, interest is on estimating changes in the longitudinal process over time while accounting for potential informative dropout. For batched samples, we do not observe the actual  $Y_{ij}$ , but rather the average measurement in each batch. At the beginning of the study, subjects are placed into batches, and these batches are maintained throughout. Since subjects are dying over time and the batch structure is maintained, there may be very few subjects in a batch as the study draws to an end. Define  $B_{lj}$  as the set of subjects who are alive in the  $l$ th batch at the  $j$ th time point. Define  $n_{lj}$  as the number of subjects contained in  $B_{lj}$ . Further, define  $X_{lj} = \frac{1}{m_j} \sum_{i \in B_{lj}} y_{ij}$ , where in each group,  $l = 1, 2, \dots, L$ , and where

$L$  is the number of batches in that group. Animals are grouped into batches of five animals that are repeatedly weighed in the same cage. In this study, there are 19 batches in each group ( $L = 19$  in each of the two groups).

When individual longitudinal measurements  $Y_{ij}$ 's are observed, maximum-likelihood estimation is relatively simple as described earlier in Sect. 2. Estimation is much more difficult when longitudinal measurements are collected in batches. In principle, we can obtain MLEs of the parameters in model (4)–(5), denoted by  $\eta$ , by directly maximizing the joint likelihood, where the individual contribution of the likelihood for the  $l$ th batch is

$$L(X_l, d_l; \eta) = \int_{b_l} f(X_l | b_l) f(d_l | b_l) f(b_l) db_l, \tag{6}$$

where  $X_l = (X_{l1}, X_{l2}, \dots, X_{lJ_l})'$ ,  $J_l$  is the last observed time-point immediately before the last subject in the  $l$ th batch dies ( $n_{lj} = 0$  for  $j > J_l$ ) and  $b_l$  and  $d_l$  are a vectors of all the random effects and dropout times, respectively, for individuals in the  $l$ th batch. In the application considered here,  $b_l$  is a vector contain ten random effects and  $d_l$  is a vector containing the dropout times for a batch size of five mice per cage. In (6),  $f(X_l | b_l) = \prod_{j=1}^{J_l} f(X_{lj} | b_l)$ , where  $f(X_{lj} | b_l)$  is a univariate normal density with mean given by  $\beta_0 + \beta_1 t_j + \frac{1}{m_j} \sum_{i \in B_{lj}} (b_{0i} + b_{1i} t_j)$  (see (4)) and variance  $\sigma_\epsilon^2 / n_{lj}$ .

Further,  $f(b_l)$  is a multivariate normal with block diagonal matrix (under an ordering where random effects on the same subjects are grouped together) and  $f(d_l | b_l)$  is the product of geometric probabilities.

One approach to maximize the likelihood is to use the E–M algorithm. In the E-step we compute the expected value of the complete-data log likelihood (the

log-likelihood we would have if we observed  $b_l$ ) given the observed data  $X_l$  and  $d_l$ , and in the M-step, we maximize the resulting expectation. Specifically,

$$\max_{\eta} \sum_{l=1}^L E[\log L(X_l, b_l, d_l; \eta) | X_l, d_l], \tag{7}$$

where  $\eta$  is a vector of all parameters of the shared parameter model and  $\log L(X_l, b_l, d_l; \eta)$  is the log of the complete data likelihood for the  $l$ th batch. The standard E–M algorithm is implemented by iterating between an E- and an M-step, whereby the expectation in (7) is evaluated in the E-step and the parameters are updated through the maximization of (7) in the M-step. Unfortunately, the E-step is difficult to implement in closed form. As an alternative [Albert and Shih \(2011\)](#) proposed a Monte-Carlo (MC) EM algorithm where the E-step is evaluated using the Metropolis–Hastings algorithm; the details are included in this paper.

Although the shared parameter modeling approach is feasible, it can be computational intensive due to the Monte-Carlo Sampling. An alternative approach that is simpler to implement for the practitioner is the conditional model discussed in Sect. 2. The conditional approach can easily be adapted for approximate parameter estimation for the shared random effects model with batched longitudinal data.

The approximate conditional model approach discussed for *unbatched* longitudinal data can be applied to the batched data (i.e., observing  $X_{lj}$ 's rather than  $Y_{ij}$ 's). Since  $Y_i|d_i$  in (4) is multivariate normal,  $X_l|d_l$  is also multivariate normal. Denote

$$\Sigma_c = \begin{pmatrix} \sigma_{c0}^2 & \sigma_{c0c1} \\ \sigma_{c0c1} & \sigma_{c1}^2 \end{pmatrix}.$$

The conditional distribution of  $X_l|d_l$  is multivariate normal with means and covariance matrix given by

$$E(X_{lj}|d_l) = \omega_0 + \omega_1 t_j + \omega_2 \sum_{i \in B_{lj}} d_i/n_{lj} + \omega_3 \sum_{i \in B_{lj}} d_i t_j/n_{lj}, \tag{8}$$

$$\text{Cov}(X_{lj}, X_{lj'}|d_l) = \frac{\min(n_{lj}, n_{lj'})}{n_{lj}n_{lj'}} (\sigma_{c0}^2 + t_j t_{j'} \sigma_{c1}^2 + (t_j + t_{j'}) \sigma_{c0c1}), \tag{9}$$

for  $j \neq j'$ , and

$$\text{Var}(X_{lj}|d_l) = \frac{1}{n_{lj}} (\sigma_{c0}^2 + t_j^2 \sigma_{c1}^2 + 2t_j \sigma_{c0c1} + \sigma_{\epsilon}^2). \tag{10}$$

The multivariate normal likelihood with mean and variance given by (8)–(10) can be maximized using a quasi-Newton Raphson algorithm. This has been implemented in R using the `optimum` function. Once the MLEs of the conditional



model are computed, inference about the average intercept and slope can be performed by marginalizing over the dropout times. Similar to what was described for un-batched longitudinal data in Sect. 2, the average intercept and slope can be estimated by  $\hat{\omega}_0 + \hat{\omega}_2 \bar{d}_1$  and  $\hat{\omega}_1 + \hat{\omega}_3 \bar{d}_2$ , respectively. Similar to variance estimation for the SRP model, standard errors for the estimated mean intercept and slope can be estimated using the bootstrap by re-sampling cage-specific data.

We examined the statistical properties of the maximum-likelihood and the approximate conditional approach using simulations. First, we simulated data according to the SRP model (4) and (5) and fit the correctly specified SRP model, the approximate conditional model (APM), and an ignorable model (IM) that simply fit (4) without regard to the dropout process. Data are simulated under model (4) and (5) with  $\sigma_{b0} = \sigma_{b1} = \sigma_\epsilon = 1$ ,  $\sigma_{b0b1} = 0$ ,  $\alpha(t_j) = -1$ ,  $\theta_1 = \theta_2 = 0.25$ , and an intercept and slope of 0 and 1, respectively. The average estimated slopes under the SRP model, ACM, and the IM were 0.99 (SD = 0.12), 1.07 (0.11), and 0.83 (0.08), respectively. Not surprisingly, the SRP model is unbiased under the correct specification and the IM is highly biased. The ACM is approximately unbiased which is consistent with our previous theoretical discussions. Second, we simulated data according to the ACM. In this case, the ACM is unbiased, but both the SRP model and IM are severely biased. These simulations suggest that the ACM model may be more robust (under different model formulations) than the SRP model.

A detailed analysis of these data is presented in [Albert and Shih \(2011\)](#). We will summarize the analysis here. Scientific interest was on estimating and comparing the weight in animals at full growth (15 months) and subsequently the decline in weight in older age (slope) between the treatment group (an agent called Tempol) and a control group. Table 1 shows estimates and standard errors for the IM (simply fit the longitudinal model and discard the relationship between the two processes), SRP, and conditional approximation approaches. All methods show that Tempol treated animals have a statistically significant lower early adult weight (intercept) and a slower decline in weight into later adulthood (slope) as compared with genetically identical control animals.

## 4 Jointly Modeling Multivariate Longitudinal Measurements and Discrete Time-to-Event Data

An exciting area in biomedical research is investigating the relationship between biomarker measurements and time-to-event. For example, developing a predictor of the risk of pre-term birth from biomarker data is an important goal in obstetrical medicine. SRP models that link the two processes provides a nice way to do this. Unfortunately, this is problematic in relatively high dimensions.

Denote  $Y_{1i} = (Y_{1i1}, Y_{1i2}, \dots, Y_{1iJ_i})'$ ,  $Y_{2i} = (Y_{2i1}, Y_{2i2}, \dots, Y_{2iJ_i})'$ ,  $\dots$ ,  $Y_{Pi} = (Y_{Pi1}, Y_{Pi2}, \dots, Y_{PiJ_i})'$  as the  $P$  biomarkers measured repeatedly at  $j = 1, 2, \dots, J_i$  time points. Further, define  $Y_{pi}^* = (Y_{pi1}^*, Y_{pi2}^*, \dots, Y_{piJ_i}^*)'$  as the longitudinal measurements without

**Table 1** Batched mouse data example

Method	Parameters	Est. tempol	Est. control
Ignorable	Intercept	28.1 (0.24)	38.4 (0.76)
	Slope	-1.81 (0.19)	-2.47 (0.28)
Shared random parameter	Intercept	27.9 (0.24)	36.9 (0.42)
	Slope	-1.87 (0.17)	-2.20 (0.20)
Conditional	Intercept	27.9 (0.23)	38.3 (0.72)
	Slope	-1.48 (0.19)	-2.36 (0.31)

The shared random parameter model and conditional model estimates of intercept and slope in the Tempol and control groups. Estimates are presented with standard errors in ( ). Standard errors were estimated using 250 samples of a non-parametric bootstrap

measurement error for the  $p$ th biomarker and  $Y_i^* = (Y_{1i}^*, Y_{2i}^*, \dots, Y_{pi}^*)$ . We consider a joint model for multivariate longitudinal and discrete time-to-event data in which the discrete event time distribution is modeled as a linear function of previous true values of the biomarkers without measurement error on the probit scale. Specifically,

$$P(R_{ij} = 0 | R_{i,j-1} = 1; X_i^*) = \Phi\left(\alpha_{0j} + \sum_{p=1}^P \alpha_p Y_{pi(j-1)}^*\right), \tag{11}$$

where  $i = 1, 2, \dots, I$ ,  $j = 1, 2, \dots, J_i$ ,  $R_{i0}$  is taken as 1,  $\alpha_{0j}$  governs the baseline discrete event time distribution and  $\alpha_p$  measures the effect of the  $p$ th biomarker ( $p = 1, 2, \dots, P$ ) at time  $t_{j-1}$  on survival at time  $t_j$ .

The longitudinal data is modeled assuming that the fixed and random effect trajectories are linear. Specifically, the multivariate longitudinal biomarkers can be modeled as

$$Y_{pij} = Y_{pij}^* + \varepsilon_{pij}, \tag{12}$$

where

$$Y_{pij}^* = \beta_{p0} + \beta_{p1}t_j + \gamma_{pi0} + \gamma_{pi1}t_j, \tag{13}$$

where  $\beta_{p0}$  and  $\beta_{p1}$  are the fixed effect intercept and slope for the  $p$ th biomarker, and  $\gamma_{pi0}$  and  $\gamma_{pi1}$  are the random effect intercept and slope for the  $p$ th biomarker on the  $i$ th individual. Denote

$$\beta = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}, \dots, \beta_{P0}, \beta_{P1})'$$

and

$$b_i = (b_{i0}, b_{i1}, b_{2i0}, b_{2i1}, \dots, b_{Pi0}, b_{Pi1})'$$

We assume that  $b_i$  is normally distributed with mean 0 and variance  $\Sigma_b$ , where  $\Sigma_b$  is a  $2P \times 2P$  dimensional variance matrix and  $\epsilon_{pij}$  are independent error terms which are assumed to be normally distributed with mean 0 and variance  $\sigma_{pe}^2$  ( $p = 1, 2, \dots, P$ ).

Albert and Shih (2010b) proposed a two-stage regression calibration approach for estimation, which can be described as follows. In the first stage, multivariate linear mixed models can be used to model the longitudinal data. In the second stage, the time-to-event model is estimated by replacing the random effects with corresponding empirical Bayes estimates. There are three problems with directly applying this approach. First, estimation in the first stage is complicated by the fact that simply fitting multivariate linear mixed models results in bias due to informative dropout; this is demonstrated by Albert and Shih (2010a) for the case of  $P = 1$ . Second, parameter estimation for multivariate linear mixed models can be computationally difficult when the number of longitudinal measurements ( $P$ ) is even moderately large. Third, calibration error in the empirical Bayes estimation needs to be accounted for in the time-to-event model. The proposed approach will deal with all three of these problems.

The bias from informative dropout is a result of differential follow-up whereby the longitudinal process is related to the length of follow-up. That is, in (13), patients with large values of  $Y_{pij}^*$  are more likely to have an early event when  $\alpha_p > 0$  for  $p = 1, 2, \dots, P$ . There would be no bias if all  $J$  follow-up measurements were observed on all patients. As proposed by Albert and Shih (2010a) for univariate longitudinal data, we can avoid this bias by generating complete data from the conditional distribution of  $Y_i = (Y_{1i}, Y_{2i}, \dots, Y_{Pi})$  given  $d_i$ , denoted as  $Y_i|d_i$ . Since  $Y_i|d_i$  under model (11–12) does not have a tractable form, we propose a simple approximation for this conditional distribution. The distribution of  $Y_i|d_i$  can be expressed as

$$P(Y_i|d_i) = \int h(Y_i|b_i, d_i)g(b_i|d_i)db_i. \tag{14}$$

Since  $d_i$  and the values of  $Y_i$  are conditional independent given  $b_i$ ,  $h(Y_i|b_i, d_i) = h(Y_i|b_i)$ , where  $h(Y_i|b_i) = \prod_{p=1}^P h(Y_{pi}|b_{pi0}, b_{pi1})$ . The distribution of  $Y_i|d_i$  can be expressed as a multivariate linear mixed model if we approximate  $g(b_i|d_i)$  by a normal distribution. Under the assumption that  $g(b_i|d_i)$  is normally distributed with mean  $\mu_{d_i} = (\mu_{01d_i}, \mu_{11d_i}, \mu_{02d_i}, \mu_{12d_i}, \dots, \mu_{0Pd_i}, \mu_{1Pd_i})'$  and variance  $\Sigma_{bd_i}^*$ , and by re-arranging mean structure parameters in the integrand of (14) so that the random effects have mean zero,  $Y_i|d_i$  corresponds to the following multivariate linear mixed model

$$Y_{pij}|(d_i, b_{ip0d_i}^*, b_{ip1d_i}^*) = \beta_{p0d_i}^* + \beta_{p1d_i}^* t_j + b_{ip0d_i}^* + b_{ip1d_i}^* t_j + \varepsilon_{pij}^*, \quad (15)$$

where  $i = 1, 2, \dots, I$ ,  $j = 1, 2, \dots, J_i$ , and  $p = 1, 2, \dots, P$ . The parameters  $\beta_{p0d_i}^*$  and  $\beta_{p1d_i}^*$  are intercept and slope parameters for the  $p$ th longitudinal measurement and for patients who have an event time at time  $d_i$  or who are censored at time  $t_j$ . In addition, the associated random effects  $b_{id_i}^* = (b_{i10d_i}^*, b_{i11d_i}^*, b_{i20d_i}^*, b_{i21d_i}^*, \dots, b_{ip0d_i}^*, b_{ip1d_i}^*)'$  are multivariate normal with mean 0 and variance  $\Sigma_{bd_i}^*$ , and the residuals  $\varepsilon_{pij}^*$  are assumed to have an independent normal distribution with mean zero and variance  $\sigma_{\varepsilon_p}^{*2}$ . Thus, this conditional model involves estimating separate fixed effect intercept and slope parameters for each potential event-time and for subjects who are censored at time  $t_j$ . Likewise, separate random effects distributions are estimated for each of these discrete time points. For example, the intercept and slope fixed-effect parameters for the  $p$ th biomarker for those patients who have an event at time  $d_i = t_3$  is  $\beta_{p0t_3}^*$  and  $\beta_{p1t_3}^*$ , respectively. Further, the intercept and slope random effects for all  $P$  biomarkers on those patients who have an event at time  $d_i = t_3$ ,  $b_{it_3}^*$ , is multivariate normal with mean 0 and variance  $\Sigma_{bt_3}^*$ . A similar approximation has been proposed by [Albert and Shih \(2010a\)](#) for univariate longitudinal data ( $P = 1$ ).

Recall that by generating complete data from (15) we are able to avoid the bias due to informative dropout. However, when  $P$  is large, direct estimation of model (15) is difficult since the number of elements in  $b_{id_i}^*$  grows exponentially with  $P$ . For example, the dimension of the variance matrix  $\Sigma_{bd_i}^*$  is  $2P$  by  $2P$  for  $P$  longitudinal biomarkers. [Fieuws and Verbeke \(2005\)](#) proposed estimating the parameters of multivariate linear mixed models by formulating bivariate linear mixed models on all possible pairwise combinations of longitudinal measurements. In the simplest approach, they proposed fitting bivariate linear mixed models on all  $\binom{P}{2}$  combinations of longitudinal biomarkers and averaging ‘‘overlapping’’ or duplicate parameter estimates. Thus, we estimate the parameters in the fully specified model (15) by fitting  $\binom{P}{2}$  bivariate longitudinal models that only include pairs of longitudinal markers. Fitting these bivariate models is computationally feasible since only four correlated random effects are contained in each model. (i.e.,  $\Sigma_{bd_i}^*$  is a four-by-four dimensional matrix for each discrete event-time  $d_i$ .) Duplicate estimates of fixed effects and random-effect variances from all pairwise bivariate models are averaged to obtain final parameter estimates of the fully specified model (15). For example, when  $P = 4$  there are  $(P - 1) = 3$  estimates of  $\beta_{p0d_i}^*$ ,  $\beta_{p1d_i}^*$ ,  $\sigma_{\varepsilon_p}^{*2}$  for the  $p$ th longitudinal biomarker that need to be averaged.

Model (15) is then used to construct complete longitudinal pseudo data sets which in turn are used to estimate the mean of the posterior distribution of an individual’s random effects given the data. Specifically, multiple complete longitudinal data sets can be constructed by simulating  $Y_{pij}$  values from the approximation to the distribution of  $Y_i|d_i$  given by (15) where the parameters are replaced by their estimated values. Since the simulated data sets have complete follow-up on each individual, the bias in estimating the posterior mean of  $b_i$  caused by informative dropout will be much reduced.

The posterior mean of distribution  $b_i$  given the data can be estimated by fitting (11)–(13) to the generated complete longitudinal pseudo data. However, similar to fitting the conditional model (15), fitting model (11)–(13) is difficult due to the high dimension of  $\Sigma_b$ . Thus, we again use the pairwise estimation approach of Fieuws and Verbeke (2005), whereby we estimate the parameters of (2)–(3) by fitting all pairwise bivariate models and averaging duplicate parameter estimates to obtain final parameter estimates. For each generated complete longitudinal pseudo data set, the estimate of the posterior mean, denoted as  $\hat{b}_i = (\hat{b}_{1i0}, \hat{b}_{1i1}, \dots, \hat{b}_{Pi0}, \hat{b}_{Pi1})'$  can be calculated as

$$\hat{b}_i = \Sigma_b Z_i' V_i^{-1} (X_i - Z_i \hat{\beta}), \tag{16}$$

where  $Z_i$  is a  $PJ \times 2P$  design matrix corresponding to the fixed and random effects in (11)–(13), where  $Z_i = \underbrace{\text{diag}(A', A', \dots, A')}_{P \text{ Times}}$ ,

$$A = \begin{pmatrix} 1 & 1 & \dots & 1 \\ t_1 & t_2 & \dots & t_J \end{pmatrix},$$

and  $V_i$  is the variance of  $X_i$ . Estimates of  $Y_{pij}^*$ , denoted as  $\hat{Y}_{pij}^*$ , are obtained by substituting  $(\hat{\beta}_{p0}, \hat{\beta}_{p1}, \hat{b}_{pi0}, \hat{b}_{pi1})$  for  $(\beta_{p0}, \beta_{p1}, b_{pi0}, b_{pi1})$  in (13).

To account for the measurement error in using  $\hat{b}_i$  as compared with using  $b_i$  in (11), we note that

$$P(R_{ij} = 0 | R_{i(j-1)} = 1; \hat{Y}_i^*) = \Phi \left( \frac{\alpha_{0j} + \sum_{p=1}^P \alpha_p \hat{Y}_{pij}^*}{\sqrt{1 + \text{Var} \left\{ \sum_{p=1}^P \alpha_p (\hat{Y}_{pij}^* - Y_{pij}^*) \right\}}} \right), \tag{17}$$

where  $\text{Var} \left\{ \sum_{p=1}^P \omega_p (\hat{Y}_{pi(j-1)}^* - Y_{pi(j-1)}^*) \right\} = C_{ij}' \text{Var}(\hat{b}_i - b_i) C_{ij}$ ,  $C_{ij} = (\omega_1, \omega_1 t_{j-1}, \omega_2, \omega_2 t_{j-1}, \dots, \omega_p, \omega_p t_{j-1})$ ,  $\text{Var}(\hat{b}_i - b_i) = \Sigma_b - \Sigma_b Z_i' \{V_i^{-1} - V_i^{-1} Z_i Q Z_i' V_i^{-1}\} Z_i \Sigma_b$ , and where  $Q = \sum_{i=1}^I (Z_i' V_i^{-1} Z_i)^{-1}$  (Laird and Ware 1982). Expression (17) follows from the fact that  $E[\Phi(a + V)] = \Phi[(a + \mu) / \sqrt{1 + \tau^2}]$ , where  $V \sim N(\mu, \tau^2)$ .

In the second stage,  $\alpha_{0j}$  ( $j = 1, 2, \dots, J$ ) and  $\alpha_p$  ( $p = 1, 2, \dots, P$ ) can be estimated by maximizing the likelihood

$$L = \prod_{i=1}^I \left[ \prod_{j=1}^{J_i} \{1 - P(R_{ij} = 0 | R_{i(j-1)} = 1; \hat{Y}_i^*)\} \right] P(R_{i(J_i+1)} = 0 | R_{iJ_i} = 1; \hat{Y}_i^*)^{J_i < J}, \tag{18}$$

where  $P(R_{ij} = 0 | R_{i(j-1)} = 1, \hat{Y}_i^*)$  is given by (17). Thus, we propose the following algorithm for estimating  $\alpha_{0j}$  and  $\alpha_p$  ( $p = 1, 2, \dots, P$ ).

1. Estimate the parameters of model (15) by fitting  $\binom{P}{2}$  bivariate models to each of the pairwise combinations of longitudinal measurements and averaging duplicate parameter estimates. The bivariate models can be fit in R using code presented in [Doran and Lockwood \(2006\)](#).
2. Simulate complete longitudinal pseudo measurements (i.e.,  $Y_{pij}$  for  $p = 1, 2, \dots, P$ ,  $i = 1, 2, \dots, I$ ,  $j = 1, 2, \dots, J$ ) from model (15) with model parameters estimated from step 1.
3. Estimate the parameters in model (12)–(13) without regard to the event time distribution from complete longitudinal pseudo measurements (simulated in step 2) by fitting all possible  $\binom{P}{2}$  bivariate longitudinal models and averaging duplicate model parameter estimates.
4. Calculate  $\hat{b}_i$  using (7) and  $\hat{Y}_{pij}^*$  using (13) with  $b_i$  replaced by  $\hat{b}_i$  and  $\beta$  being replaced by  $\hat{\beta}$  estimated in step 3.
5. Estimate  $\alpha_{0j}$  ( $j = 1, 2, \dots, J$ ) and  $\alpha_p$  ( $p = 1, 2, \dots, P$ ) using (17) and (18).
6. Repeat steps 2 to 5  $M$  times and average  $\hat{\alpha}_{0j}$  and  $\hat{\alpha}_p$  to get final estimates.

We choose  $M = 10$  in the simulations and data analysis since this was shown to be sufficiently large for univariate longitudinal modeling discussed in [Albert and Shih \(2010a\)](#). Asymptotic standard errors of  $\hat{\alpha}_{0j}$  and  $\hat{\alpha}_p$  cannot be used for inference since they fail to account for the missing data uncertainty in our procedure. The bootstrap ([Efron and Tibshirani 1993](#)) can be used for valid standard error estimation.

This approach is most useful in situations where the number of longitudinal measurements is very large (e.g., panels of cytokine measurements followed longitudinally). For computational simplicity, we focus on a simulated example where three biomarkers are measured longitudinally at five time points on 300 individuals. Table 2 shows the results of these simulations. The results show that simply using the observed longitudinal data will result in severely biased estimation. The proposed approach results in unbiased estimation for the parameters of the joint model. The table also includes estimates for the situation in which we observe the biomarkers without measurement error (only possible to do in simulations). This *strawman* case simply shows us that we could do better in terms of efficiency if the biomarkers could be assessed with less measurement error.

**Table 2** Simulation shows estimates for a method that uses the true values of the markers without measurement error (NoME), the proposed method (Proposed), and an approach that uses the observed biomarkers (Observed)

Parameters	True values	NoME	Proposed	Observed
$\alpha_0$	-1.75	-1.77 (0.115)	-1.76 (0.180)	-1.37 (0.089)
$\alpha_1$	0.40	0.408 (0.060)	0.405(0.089)	0.221(0.042)
$\alpha_2$	0	0.00 (0.058)	0.00(0.077)	0.001(0.042)
$\alpha_3$	0.40	0.405 (0.062)	0.400(0.092)	0.229(0.043)

We simulate according to (11)–(13) with  $\alpha_{0j} = \alpha_0$  for all  $j$  and with stated values of  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$ . Remaining values are given in Table 1 of [Albert and Shih \(2010b\)](#)

## 5 Jointly Modeling of Menstrual Cycle Length and Time-to-Pregnancy

An important scientific problem in reproductive epidemiology is both to characterize the menstrual cycle patterns in women who are attempting pregnancy and to develop predictive models for time-to-pregnancy. A SRP model that links together the complex menstrual cycle pattern with time-to-pregnancy is important for valid statistical analysis in both problems. Specifically, when interest focuses on characterizing longitudinal changes in the menstrual cycle, it is important to account for the dependence between the two processes since failure to do so results in informative dropout in the longitudinal process. Further, incorporating dependence between the two processes is important for developing a flexible class of prediction models of time-to-pregnancy.

The menstrual cycle pattern is complex since it is well known to be long tailed with a proportion of cycles being unusually long while a majority appear within normal ranges. Various authors (e.g., Guo et al. 2006 and references within) have proposed two component mixture models with one component reflecting a distribution with a long right tail and the other reflecting a normal distribution. McLain et al. (2012) propose a class a mixture model with a normal distribution for “normal” cycles and a long tailed distribution (reflecting the possibility of both extremely long and short irregular cycles). Further, McLain et al. introduce random effects which are shared between the two models as well as with a discrete time survival model characterizing time-to-pregnancy.

For illustration, we present a simplified version of McLain et al.’s modeling approach without external covariates. The menstrual cycle is modeled as a mixture of two components. First, for normal cycles

$$\begin{aligned} Y_{N,ij} &= \mu_i + \varepsilon_{N,ij} \\ &= \beta_0 + b_{N,i} + \varepsilon_{N,ij}, \quad \varepsilon_{N,ij} \sim N(0, \sigma_i^2), \end{aligned} \tag{19}$$

where  $\sigma_i^2 = \sigma_0^2 \exp(b + b_{S,i})$ . For abnormal cycles, the following representation is assumed

$$Y_{A,ij} = \beta_0 + b_{A,i} + \varepsilon_{A,ij}, \quad \varepsilon_{A,ij} \sim EVD(0, \eta), \tag{20}$$

where  $EVD(0, \eta)$  denotes an extreme-value type I distribution with location 0 and scale  $\eta$ . Further, for implementation it was assumed that  $b_{A,ij} = \varepsilon b_{N,ij}$ , that the between subject heterogeneity is a scalar shift between the normal and abnormal cycles. The distribution of the menstrual cycle length  $Y_{ij}$  is completed through the specification of the mixture,  $Y_{ij} = g_{ij}Y_{N,ij} + (1 - g_{ij})Y_{A,ij}$ , where  $g_{ij}$  is an indicator that the  $i$ th women’s  $j$ th cycle is normal, and  $p = P(g_{ij} = 1)$ . Finally, time-to-pregnancy is specified with a discrete-time survival model as in (5) and (11) with a different link function,

$$P(R_{ij} = 0 | R_{ij-1} = 1, b_{N,i}, b_{S,i}) = 1 - \exp(-V_{ij} \exp(\omega + \phi_{\mu} b_{A,i} + \phi_{\sigma} b_{S,i})), \quad (21)$$

where  $V_{ij}$  is an indicator of whether the  $i$ th subject had sexual intercourse during the fertile window (period in which they can conceive) during the  $j$ th cycle. The parameters  $\phi_{\mu}$  and  $\phi_{\sigma}$  characterize the dependence between time-to-pregnancy and the mean structure and variance structure (among normal cycles), respectively. Negative values of these parameters characterize positive associations between the two processes in terms of mean and variance of the longitudinal process.

Parameter estimation was conducted by maximizing the joint likelihood as described in their manuscript. The major complication was evaluating the bi-variate integral in the joint likelihood (integrating over the bi-variate random effects). McLain et al. used Gaussian quadrature (Abramowitz and Stegun 1972) for this numerical integration.

McLain et al. (2012) fit this model to interesting time-to-pregnancy cohort data. Of interest is that they found estimates of  $\phi_{\mu}$  which were positive and estimates of  $\phi_{\sigma}$  which were negative ( $\phi_{\sigma}$  estimates were significantly different from zero), reflecting that women with shorter cycles and more normal cycle variability had a longer time to pregnancy.

## 6 Discussion

This paper presents a summary of methodology and approaches for using SRP models to analyze longitudinal data subject to missingness. The basic approach which started with Wu and Carroll (1988) has been expanded in many directions. In this paper we reviewed some of these expansions focusing on making inference in batched or pooled longitudinal data subject to missingness, joint modeling of high dimensional longitudinal data and time-to-event data, and in the joint modeling of time-to-pregnancy and complex menstrual cycle patterns.

We discuss both a direct maximum-likelihood approaches and approximating approaches for fitting SRP models. The direct modeling approaches, although feasible for univariate longitudinal responses, can be very computationally intensive for either batched or high-dimensional longitudinal data. In fact, direct maximum-likelihood approach would be infeasible for joint modeling of high-dimensional and time-to-event data with all known approaches. The approximate conditional model, although not ideal in certain cases, provides a feasible solution to this difficult problem.

Most SRP models assume that the random effects are normally distributed. Various authors have investigated the robustness of inferences for different settings. Davidian et al. have shown that for a joint model of longitudinal and survival inferences on joint model parameters are relatively insensitive to random effects misspecification, particularly when the number of longitudinal measurements is relatively large.



Selection models where the non-ignorable missing data mechanism is incorporated by modeling the probability of missing as a direct function of the missed observation had it been observed are alternatives to SRP models. Pattern mixture modeling, another commonly used technique for analyzing longitudinal data with missingness, is one where we condition on the missing data pattern and make inference marginalized over the missing data pattern. In spirit, the pattern mixture model is similar to the conditional model which is proposed as an approximation to the SRP model.

We recommend that model adequacy be examined in traditional ways such as the examination of residuals and fitted values as well as through goodness of fit tests. However, as pointed out by [Molenberghs et al. \(2012\)](#), SRP models (as well as other models for missing data) require assumptions about underlying mechanism that are impossible to fully verify empirically. Knowledge about the subject at hand needs to be incorporated in model development for proper inference.

**Acknowledgments** We thank the referee and editor for their thoughtful and constructive comments which lead to an improved manuscript. We also thank the audience of the International Symposium in Statistics (ISS) on Longitudinal Data Analysis Subject to Outliers, Measurement Errors, and/or Missing Values. This research was supported by the Intramural Research Program of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development.

## References

- Abramowitz, M., Stegun, I.A.: Handbook of Mathematical Functions. Dover, New York (1972)
- Albert, P.S.: A transition model for longitudinal binary data subject to nonignorable missing data. *Biometrics* **56**, 602–608 (2000)
- Albert, P.S., Follmann, D.A.: Modeling repeated count data subject to informative dropout. *Biometrics* **56**, 667–677 (2000)
- Albert, P.S., Follmann, D.A.: Random effects and latent processes approaches for analyzing binary longitudinal data with missingness: a comparison of approaches using opiate clinical trial data. *Stat. Methods Med. Res.* **16**, 417–439 (2007)
- Albert, P.S., Follmann, D.A., Wang, S.A., Suh, E.B.: A latent autoregressive model for longitudinal binary data subject to informative missingness. *Biometrics* **58**, 631–642 (2002)
- Albert, P.S., Shih, J.H.: On estimating the relationship between longitudinal measurements and time-to-event data using a simple two-stage procedure. *Biometrics* **66**, 983–991 (2010a)
- Albert, P.S., Shih, J.H.: An approach for jointly modeling multivariate longitudinal measurements and discrete time-to-event data. *Ann. Appl. Stat.* **4**, 1517–1532 (2010b)
- Albert, P.S., Shih, J.H.: Modelling batched Gaussian longitudinal weight data in mice subject to informative dropout. Published on-line before print. *Stat. Methods Med. Res.* (2011)
- DeGruttola, V.T., Tu, X.M.: Modeling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics* **50**, 1003–1014 (1994)
- Doran, H.C., Lockwood, J.R.: Fitting value-added models in R. *J. Educ. Behav. Stat.* **31**, 205–230 (2006)
- Efron, B., Tibshirani, R.J.: An Introduction to the Bootstrap. Chapman and Hall, New York (1993)
- Fieuws, S., Verbeke, G.: Pairwise fitting of mixed models for the joint modelling of multivariate longitudinal profiles. *Biometrics* **62**, 424–431 (2005)

- Follmann, D., Wu, M.: An approximate generalized linear model with random effects for informative missing data. *Biometrics* **51**, 151–168 (1995)
- Gao, S.: A shared random effect parameter approach for longitudinal dementia data with non-ignorable missing data. *Stat. Med.* **23**, 211–219 (2004)
- Guo, Y., Manatunga, A.K., Chen, S., Marcus, M.: Modeling menstrual cycle length using a mixture distribution. *Biostatistics* **7**, 151–168 (2006)
- Hogan, J.W., Laird, N.M.: Model-based approaches to analyzing incomplete longitudinal and failure time data. *Stat. Med.* **16**, 259–272 (1997)
- Laird, N.M., Ware, J.H.: Random -effects models for longitudinal data. *Biometrics* **38**, 963–978 (1982)
- Little, R.J.A.: Pattern-Mixture models for multivariate incomplete data. *J. Am. Stat. Assoc.* **88**, 125–134 (1993)
- Little, R.J.A.: Modeling the drop-out mechanism in repeated-measures studies. *J. Am. Stat. Assoc.* **90**, 1112–1121 (1995)
- Little, R.J.A., Rubin, D.B.: *Statistical Analysis with Missing Data*. Wiley, New York (1987)
- McLain, A., Lum, K.J., Sundaram, R.: A joint mixed effects dispersion model for menstrual cycle length and time-to-pregnancy. *Biometrics* **68**, 648–656 (2012)
- Molenberghs, G., Njeru, E., Kenward, M.G., Verbeke, G.: Enriched-data problems and essential non-identifiability. *Int. J. Stat. Med. Res.* **1**, 16–44 (2012)
- Mori, M., Woolson, R.F., Woodworth, G.G.: Slope estimation in the presence of informative censoring: modeling the number of observations as a geometric random variable. *Biometrics* **50**, 39–50 (1994)
- Rubin, D.B.: Inference and missing data. *Biometrika* **63**, 581–592 (1976)
- Schluchter, M.D.: Methods for the analysis of informative censored longitudinal data. *Stat. Med.* **11**, 1861–1870 (1992)
- Schluchter, M.D., Greene, T., Beck, G.J.: Analysis of change in the presence of informative censoring: application to a longitudinal clinical trial of progressive renal disease. *Stat. Med.* **20**, 989–1007 (2001)
- Sutradhar, B.C., Mallick, T.S.: Modified weights based on generalized inferences in incomplete longitudinal binary models. *Can. J. Stat.* **38**, 217–231 (2010)
- Ten Have, T.R., Pulkstenis, E., Kunselman, A., Landis, J.R.: Mixed effects logistic regression models for longitudinal binary response data with informative dropout. *Biometrics* **54**, 367–383 (1998)
- Tsiatis, A.A., Davidian, M.: Joint modeling of longitudinal and time-to-event data: an overview. *Stat. Sin.* **14**, 809–834 (2004)
- Tsiatis, A.A., Davidian, M.: A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* **88**, 447–459 (2001)
- Tsiatis, A.A., DeGruttola, V., Wulfsohn, M.S.: Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *J. Am. Stat. Assoc.* **90**, 27–37 (1995)
- Vonesh, E.F., Greene, T., Schluchter, M.D.: Shared parameter models for the joint analysis of longitudinal data with event times. *Stat. Med.* **25**, 143–163 (2006)
- Wu, M.C., Carroll, R.J.: Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44**, 175–188 (1988)
- Wulfson, M.S., Tsiatis, A.A.: A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330–339 (1997)
- Zeger, S.L., Qaqish, B.: Markov regression models for time series: a quasilikelihood approach. *Biometrics* **44**, 1019–1031 (1988)