Discussion of "Missing Data Methods in Longitudinal Studies: A Review" by Ibrahim and Molenberghs

Michael J. Daniels and Chenguang Wang

Jan. 18, 2009

First, we would like to thank Joe and Geert for a carefully written review paper on longitudinal data. We would like to expand on several points discussed in this paper. Specifically, we would like to expand on 1) the interpretation of covariate effects and use of identifying restrictions with covariates in mixture models (Section 4.2.2) and 2) issues with sensitivity analyses in parametric models for the full **fix wording** and in selection models in general (Section 4.2.3).

1 Mixture Models

In the following, we focus on the setting of covariates that are collected at baseline with no missingness.

1.1 Interpretation of covariate effects

In longitudinal studies, as discussed here, the main focus of inference is usually on the marginal distribution, $p(\boldsymbol{y})$. In mixture models, the full-data model $p(\boldsymbol{y}|\boldsymbol{x})$ is a mixture of component distributions with regard to different missing patterns r, i.e.

$$p(\boldsymbol{y}|\boldsymbol{x}) = \sum_{r \in \mathscr{R}} p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{r}) p(\boldsymbol{r}|\boldsymbol{x}).$$

Similarly, $E[\boldsymbol{Y}|\boldsymbol{x}]$ is

$$E(\boldsymbol{Y}|\boldsymbol{x}) = \sum_{\boldsymbol{r} \in \mathscr{R}} E(\boldsymbol{Y}|\boldsymbol{x}, \boldsymbol{r}) p(\boldsymbol{r}|\boldsymbol{x}).$$

So, assessing the covariate effects on the marginal mean has to be done by averaging over patterns and needs to consider (1) whether the mean is linear in covariates; (2) whether marginal distribution of missingness depends on covariates, and (3) whether covariates effects are time-invariant. In this discussion, we will focus on the first issue.

For mixture models with an identity link, averaged covariates effects for the full-data distribution have a simple form as a weighted average over pattern-specific covariate effects and have a straightforward interpretation (Fitzmaurice et al., 2001). As an example, consider the full-data response $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$ are to be observed at time points $\{t_1, \ldots, t_n\}$ and denote the baseline covariates by $\boldsymbol{X}$. Assume drop out is monotone and independent of $\boldsymbol{X}$ and let S be the dropout time with $\phi_s = P(S = t_s)$ for $s = 1, \ldots, n$ and $\sum_{s=1}^{n} \phi_s = 1$.

When the link function, denoted by g, is non-linear, and the within-pattern s ($S = t_s$) mean model is

$$g\{\mathrm{E}(Y_{ij}|\boldsymbol{X} = x_{ij})\} = x'_{ij}\boldsymbol{\beta},$$

and we have in general

$$g(\mu_j(\boldsymbol{x})) - g(\mu_j(\boldsymbol{x}') \neq (\boldsymbol{x} - \boldsymbol{x}')\boldsymbol{\beta}^{(s)}\phi^{(s)}.$$

So it can be difficult to capture the covariate effects compactly (Fitzmaurice et al., 2001; Wilkins and Fitzmaurice, 2006). Roy and Daniels (2008) proposed to specify marginalized models and impose constraints on the conditional mean. This is in the spirit of earlier work by Azzalini (1994) and Heagerty (1999). A simple version of the model in Roy and Daniels is illustrated below.

First, the marginal mean is specified as

$$g\{\mathrm{E}(Y_{ij}|\boldsymbol{X} = x_{ij})\} = x'_{ij}\boldsymbol{\beta}.$$

Second, a conditional model is specified to account for within-subject correlation and dependencies between the response and missingness pattern. We assume Y_{ij}, conditional on random effects b_i and missingness pattern S_i, are from exponential family and have distribution

$$f(Y_{ij}|b_i, S_i, \boldsymbol{X}) = \exp\{[Y_{ij}\eta_{ij} - \psi(\eta_{ij})]/\phi + h(Y_{ij}, \phi)\},$$

where

$$\eta_{ij} = g(\mathrm{E}(Y_{ij}|b_i, S_i = s, \boldsymbol{X} = x_{ij})) = \Delta_{ij} + b_i + x_{ij}\alpha^{(s)}.$$

The conditional model has to be compatible with the marginal model. In particular, the intercepts Δ_{ij} are determined by the relationship

$$\mathrm{E}(Y_{ij}|\boldsymbol{X}) = \sum_s \phi_s \int \mathrm{E}(Y_{ij}|b_i, S_i, \boldsymbol{X}) p(b_i) db_i$$

and are functions of other parameters including $\boldsymbol{\beta}$ in the model. Note that this is marginalized over both missingness patterns and subject-specific random effects. Serial correlation within pattern can be addressed by augmenting the conditional model with a Markov components (Heagerty, 2002).

1.2 Identifying restrictions with covariates

Identifying restrictions can be problematic in pattern mixture models with baseline covariates with time-invariant coefficients. We will focus on the available case missing value (ACMV) restriction (Little, 1993; Molenberghs et al., 1998) here which corresponds to MAR. Missing at random (MAR) is often taken as a starting point for analysis of incomplete data (Troxel et al., 2004; Zhang and Heitjan, 2006).

To illustrate, consider $\boldsymbol{Y} = (Y_1, Y_2)$ being a bivariate normal response with missing data only in Y_2. The missing data indicator R equals 1 or 0 corresponding to Y_2 being observed or missing. Assume

$$R \sim \mathrm{Bern}(\phi) \qquad \text{and} \qquad \boldsymbol{Y}|R = r \sim \mathrm{N}(\boldsymbol{\mu}^{(r)}, \boldsymbol{\Sigma}^{(r)})$$

where

$$\boldsymbol{\mu}^{(r)} = \begin{bmatrix} \mu_1^{(r)} \\ \mu_2^{(r)} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}^{(r)} = \begin{bmatrix} \sigma_{11}^{(r)} & \sigma_{12}^{(r)} \\ \sigma_{12}^{(r)} & \sigma_{22}^{(r)} \end{bmatrix}$$

for $r = 0, 1$. For the bivariate case, the ACMV restriction is

$$Y_2|Y_1, R = 0 \simeq Y_2|Y_1, R = 1,$$

where $\simeq$ denotes the equality in distribution. This requires that for all Y_1,

$$\begin{aligned} \mu_2^{(0)} + \frac{\sigma_{21}^{(0)}}{\sigma_{11}^{(0)}}(Y_1 - \mu_1^{(0)}) &= \mu_2^{(1)} + \frac{\sigma_{21}^{(1)}}{\sigma_{11}^{(1)}}(Y_1 - \mu_1^{(1)}) \\ \sigma_{22}^{(0)} - \frac{(\sigma_{21}^{(0)})^2}{\sigma_{11}^{(0)}} &= \sigma_{22}^{(1)} - \frac{(\sigma_{21}^{(1)})^2}{\sigma_{11}^{(1)}}, \end{aligned}$$

which in turn implies that

$$\begin{aligned} \mu_2^{(0)} &= \mu_2^{(1)} + \frac{\sigma_{21}^{(1)}}{\sigma_{11}^{(1)}}(\mu_1^{(0)} - \mu_1^{(1)}) \\ \sigma_{21}^{(0)} &= \frac{\sigma_{21}^{(1)}\sigma_{11}^{(0)}}{\sigma_{11}^{(1)}} \\ \sigma_{22}^{(0)} &= \sigma_{22}^{(1)} + \frac{(\sigma_{21}^{(1)})^2}{\sigma_{11}^{(1)}}(\frac{\sigma_{11}^{(0)}}{\sigma_{11}^{(1)}} - 1). \end{aligned}$$

This restriction identified the full data response distribution.

When there are baseline covariates with time-invariant coefficients, we have that

$$\boldsymbol{\mu}^{(r)} = \begin{bmatrix} \mu_1^{(r)} + X\beta^{(r)} \\ \mu_2^{(r)} + X\beta^{(r)} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}^{(r)} = \begin{bmatrix} \sigma_{11}^{(r)} & \sigma_{12}^{(r)} \\ \sigma_{12}^{(r)} & \sigma_{22}^{(r)} \end{bmatrix}$$

for $r = 0, 1$, where $\boldsymbol{x}$ does not contain an intercept.

The MAR assumption requires that for all X and Y_1,

$$\mu_2^{(0)} + X\beta^{(0)} + \frac{\sigma_{21}^{(0)}}{\sigma_{11}^{(0)}}(Y_1 - \mu_1^{(0)} - X\beta^{(0)}) = \mu_2^{(1)} + X\beta^{(1)} + \frac{\sigma_{21}^{(1)}}{\sigma_{11}^{(1)}}(Y_1 - \mu_1^{(1)} - X\beta^{(1)}).$$

By simple algebra, we can see that this restricts $\beta^{(0)}$ to be equal to $\beta^{(1)}$. Note that both $\beta^{(0)}$ and $\beta^{(1)}$ are *identified* from the observed data. Therefore, the ACMV restricton/MAR assumption causes over-identification and has impact on the model fit to the observed data. This is against the principle of applying identifying restrictions (Little, 1994). Ways to remedy this (and associated problems) are explored in Wang and Daniels (working paper).

2 Issues with Sensitivity Analysis

Sensitivity analysis is critical in longitudinal analysis of incomplete data with informative drop-out as stated in this paper. In the setting of missing data, the full-data model can be factored into an extrapolation model and an observed data model,

$$p(\boldsymbol{y}, \boldsymbol{r}|\boldsymbol{\omega}) = p(\boldsymbol{y}_{\text{mis}}|\boldsymbol{y}_{\text{obs}}, \boldsymbol{r}, \boldsymbol{\omega}_E)p(\boldsymbol{y}_{\text{obs}}, \boldsymbol{r}|\boldsymbol{\omega}_I),$$

where $\boldsymbol{\omega}_E$ are parameters indexing the extrapolation model and $\boldsymbol{\omega}_I$ are parameters indexing the observed data model and are identifiable from observed data (Daniels and Hogan, 2008). Full-data model inference requires unverifiable assumptions about the extrapolation model $p(\boldsymbol{y}_{\text{mis}}|\boldsymbol{y}_{\text{obs}}, \boldsymbol{r}, \boldsymbol{\omega}_E)$. A sensitivity analysis explores the sensitivity of inferences of interest about the full data response model to unverifiable assumptions about the extrapolation model. This is typically done by varying sensitivity parameter, which we define next. Suppose there exists a reparameterization $\boldsymbol{\xi}(\boldsymbol{\omega}) = (\boldsymbol{\xi}_S, \boldsymbol{\xi}_M)$ such that (1) $\boldsymbol{\xi}_s$ is a non-constant function of $\boldsymbol{\omega}_E$, (2) the observed likelihood $L(\boldsymbol{\xi}_S, \boldsymbol{\xi}_M|\boldsymbol{y}_{\text{obs}}, \boldsymbol{r})$ is a constant as a function of $\boldsymbol{\xi}_S$ and (3) given $\boldsymbol{\xi}_S$ fixed, $L(\boldsymbol{\xi}_S, \boldsymbol{\xi}_M|\boldsymbol{y}_{\text{obs}}, \boldsymbol{r})$ is a non-constant function of $\boldsymbol{\xi}_M$. A parameter $\boldsymbol{\xi}_S$ that satisfies these three conditionals is a *sensitivity parameter* and can be used for sensitivity analysis and/or for incorporation of prior information (Daniels and Hogan, 2008).

2.1 In parametric models

Unfortunately, fully parametric selection models and shared parameter models do *not* allow sensitivity analysis as sensitivity parameters cannot be found (Daniels and Hogan, Chapter 8, 2008). Examining sensitivity to distributional assumptions, e.g., random effects, will provide *different* fits to the observed data, $(\boldsymbol{y}_{\text{obs}}, \boldsymbol{r})$. In such cases, a sensitivity analysis cannot be done since varying the distributional assumptions does not provide equivalent fits to the observed data (Daniels and Hogan, 2008). It then becomes a problem of model selection. Next, we provide an example of the inability to find sensitivity parameters in a simple parametric selection model for binary data.

As an example, consider the situation when $Y = (Y_1, Y_2)$ is a bivariate binary response with missing data only in Y_2. Let $R = 1$ if Y_2 is observed and $R = 0$ otherwise.

Let $\omega^{(r)}_{y_1,y_2}$ be $P(Y_1 = y_1, Y_2 = y_2, R = r)$ and $\omega^{(0)}_{y_1+}$ be $P(Y_1 = y_1, R = 0)$. A multinomial parameterization of the full-data model of Y and R is shown in Table 1.

Table 1: A multinomial parameterization full-data model for Y

R	Y_1	Y_2	$p(y_1, y_2, r\|\boldsymbol{\omega})$
0	0	0	$\omega^{(0)}_{00}$
0	0	1	$\omega^{(0)}_{01}$
0	1	0	$\omega^{(0)}_{10}$
0	1	1	$\omega^{(0)}_{11}$
1	0	0	$\omega^{(1)}_{00}$
1	0	1	$\omega^{(1)}_{01}$
1	1	0	$\omega^{(1)}_{10}$
1	1	1	$\omega^{(1)}_{11}$

In this example, the set of parameters

$$\boldsymbol{\omega_I} = \{\omega^{(1)}_{00}, \omega^{(1)}_{01}, \omega^{(1)}_{10}, \omega^{(1)}_{11}, \omega^{(0)}_{0+}, \omega^{(0)}_{1+}\}$$

are identified by observed data without any modeling assumption. When a selection model is fully parametric, all its parameters can be identified by the observed data. To see this, we specify a parametric model for the bivariate binary example:

$$\begin{aligned} \text{logit}\, P(Y_1 = 1) &= \beta_0 \\ \text{logit}\, P(Y_2 = 1|Y_1) &= \beta_0 + \beta_1 Y_1 \\ \text{logit}\, P(R_2 = 1|Y_1, Y_2) &= \phi_0 + \tau Y_2. \end{aligned}$$

Note that we assume

$$P(Y_2 = 1|Y_1 = 0) = P(Y_1 = 1) \tag{1}$$

and

$$P(R_2 = 1|Y_1, Y_2) = P(R_2 = 1|Y_2). \tag{2}$$

We will show that the full-data model is identified under the parametric assumptions by showing all parameters, $(\beta_0, \beta_1, \phi_0, \tau)$ can be written as a function of $\boldsymbol{\omega_I}$, the identified ω's.

First, note that

$$\beta_0 = \operatorname{logit} P(Y_1 = 1) = \operatorname{logit}(\omega_{10}^{(1)} + \omega_{11}^{(1)} + \omega_{1+}^{(0)}).$$

Also, by (1),

$$\beta_0 = \operatorname{logit} P(Y_2 = 1|Y_1 = 0) = \operatorname{logit} \frac{\omega_{01}^{(1)} + \omega_{01}^{(0)}}{\omega_{00}^{(1)} + \omega_{01}^{(1)} + \omega_{0+}^{(0)}}.$$

This gives

$$\omega_{01}^{(0)} = (\omega_{10}^{(1)} + \omega_{11}^{(1)} + \omega_{1+}^{(0)})(\omega_{00}^{(1)} + \omega_{01}^{(1)} + \omega_{0+}^{(0)}) - \omega_{01}^{(1)} \quad \text{and} \quad \omega_{00}^{(0)} = \omega_{0+}^{(0)} - \omega_{01}^{(0)}. \tag{3}$$

As a consequence, since τ has the interpretation that

$$\tau = \log \left\{ \frac{P(R_2 = 1, Y_2 = 1, Y_1 = 0)}{P(R_2 = 0, Y_2 = 1, Y_1 = 0)} \Big/ \frac{P(R_2 = 1, Y_2 = 0, Y_1 = 0)}{P(R_2 = 0, Y_2 = 0, Y_1 = 0)} \right\},$$

thus it is identified by

$$\tau = \log \frac{\omega_{01}^{(1)} \omega_{00}^{(0)}}{\omega_{00}^{(1)} \omega_{01}^{(0)}},$$

where $\omega_{00}^{(0)}$ and $\omega_{01}^{(0)}$ are identified by (3).

Further, since τ can also be expressed as

$$\tau = \log \left\{ \frac{P(R_2 = 1, Y_2 = 1, Y_1 = 1)}{P(R_2 = 0, Y_2 = 1, Y_1 = 1)} \Big/ \frac{P(R_2 = 1, Y_2 = 0, Y_1 = 1)}{P(R_2 = 0, Y_2 = 0, Y_1 = 1)} \right\} = \log \frac{\omega_{11}^{(1)} \omega_{10}^{(0)}}{\omega_{10}^{(1)} \omega_{11}^{(0)}},$$

hence we have that $\omega_{11}^{(0)}$ and $\omega_{10}^{(0)}$ are identified as

$$\omega_{11}^{(0)} = \omega_{1+}^{(0)} \frac{1}{1 + \frac{\omega_{00}^{(0)} \omega_{01}^{(1)} \omega_{10}^{(1)}}{\omega_{01}^{(0)} \omega_{00}^{(1)} \omega_{11}^{(1)}}} \qquad \text{and} \qquad \omega_{10}^{(0)} = \omega_{1+}^{(0)} - \omega_{11}^{(0)}.$$

Therefore, in this parametric selection model, the parameters $\omega_{00}^{(0)}$, $\omega_{01}^{(0)}$, $\omega_{10}^{(0)}$ and $\omega_{11}^{(0)}$ are all identified (as opposed to their sums, $\omega_{0+}^{(0)}$ and $\omega_{1+}^{(0)}$).

Finally, we can show that

$$\beta_1 = \operatorname{logit} P(Y_2 = 1|Y_1 = 1) - \beta_0 = \operatorname{logit} \frac{\omega_{11}^{(1)} + \omega_{11}^{(0)}}{\omega_{11}^{(1)} + \omega_{10}^{(1)} + \omega_{1+}^{(0)}} - \beta_0$$

and

$$\phi_0 = \text{logit}\, P(R_2 = 1|Y_2 = 0) = \text{logit}\, \frac{\omega_{00}^{(1)} + \omega_{10}^{(1)}}{\omega_{00}^{(0)} + \omega_{10}^{(0)} + \omega_{00}^{(1)} + \omega_{10}^{(1)}}.$$

2.2 In Bayesian semiparametric selection models

The factorization of a selection model provides a transparent way to understand the missing data mechanism. In Bayesian selection models, an intuitive prior specification assumes independence between the parameters of the missing data mechanism ($\boldsymbol{\phi}$) and the full data response ($\boldsymbol{\beta}$)(Scharfstein et al., 2003) .

However, in a Bayesian model under this prior specification, sensitivity parameters in a selection model, denoted by τ, can be (weakly) identified by the observed data, i.e. $p(\tau|\boldsymbol{y}_{\text{obs}}, \boldsymbol{r}) \neq p_\tau(\tau)$, even though the observed data likelihood contains no information about the sensitivity parameters (Daniels and Hogan, 2008). We outline how this occurs in the following.

In general, a semi-parametric selection model might specify the full data response distribution nonparametrically (or saturated if a categorical response), p($\boldsymbol{y}$; $\boldsymbol{\beta}$) with a missing data mechanism given as as follows:

$$\text{logit}\, P(R_j = 1|R_{j-1} = 0, \boldsymbol{Y}) = h_j(\overline{Y}_{j-1}; \boldsymbol{\phi}) + q_j(\boldsymbol{Y}; \tau)$$

for $j = 1, \ldots, J$, where h_j is an arbitrary smooth function, and q_j is a user specified function that encodes assumptions about how the MDM depends on missing data and it parameters are sensitivity parameters. Note that $q_j(\boldsymbol{Y}) = 0$ implies MAR and $q_j(\boldsymbol{Y}) = q_j(Y_j)$ implies non-future dependence.

To see the cause of the weak identification, let $\boldsymbol{\theta} = \{\boldsymbol{\phi}, \boldsymbol{\beta}\}$ and $\boldsymbol{\omega_I}$ be the identified parameters. By re-parameterizing the model, we may find a mapping, indexed by τ, between $\boldsymbol{\theta}$ and $\boldsymbol{\omega_I}$,

$$\boldsymbol{\omega}_I = \eta_\tau(\boldsymbol{\theta}).$$

Due to the mapping, even a priori independence between τ and $\boldsymbol{\theta}$ will yield a priori dependence between τ and $\boldsymbol{\omega_I}$, since

$$p(\boldsymbol{\omega_I}, \tau) = p_\theta\left(\eta_\tau^{-1}(\boldsymbol{\omega_I})\right) p_\tau(\tau) \left|\frac{d\eta_\tau^{-1}(\boldsymbol{\omega_I})}{d\boldsymbol{\omega_I}}\right|. \tag{4}$$

The Jacobian introduces the dependence.

The posterior for the sensitivity parameters $\boldsymbol{\tau}$ can be expressed as

$$\begin{aligned} p(\tau|\boldsymbol{y}_{\text{obs}}, \boldsymbol{r}) &\propto \int p(\boldsymbol{y}_{\text{obs}}, \boldsymbol{r}|\boldsymbol{\omega_I}, \tau) p(\boldsymbol{\omega_I}, \tau) d\boldsymbol{\omega_I} \\ &= \int p(\boldsymbol{y}_{\text{obs}}, \boldsymbol{r}|\boldsymbol{\omega_I}) p(\boldsymbol{\omega_I}|\tau) p_\tau(\tau) d\boldsymbol{\omega_I} \\ &= p_\tau(\tau) \int p(\boldsymbol{y}_{\text{obs}}, \boldsymbol{r}|\boldsymbol{\omega_I}) p(\boldsymbol{\omega_I}|\boldsymbol{\tau}) d\boldsymbol{\omega_I}. \end{aligned}$$

Thus from (4), $p(\tau|\boldsymbol{y}_{\text{obs}}, \boldsymbol{r}) \neq p_\tau(\tau)$.

As a concrete example, consider a bivariate binary response with missing data only in Y_2 from the previous section. A saturated selection model can be specified as

$$\begin{aligned} \text{logit}\, P(Y_1 = 1) &= \beta_0 \\ \text{logit}\, P(Y_2 = 1|Y_1) &= \beta_1 + \beta_2 Y_1 \\ \text{logit}(R_2 = 1|Y_1, Y_2) &= \phi_0 + \phi_1 Y_1 + \tau Y_2 \end{aligned}$$

and $\boldsymbol{\theta} = \{\beta_0, \beta_1, \beta_2, \phi_0, \phi_1\}$. MAR holds when $\tau = 0$. Note τ is not identified by the observed data. It can be shown that for any Δ_τ, there exists $\boldsymbol{\Delta}_\theta$, such that

$$\eta_\tau(\boldsymbol{\theta}) = \eta_{\tau+\Delta_\tau}(\boldsymbol{\theta} + \boldsymbol{\Delta}_\theta),$$

i.e $(\tau, \boldsymbol{\theta})$ and $(\tau + \Delta_\tau, \boldsymbol{\theta} + \boldsymbol{\Delta}_\theta)$ will yield the same law of observed data.

Let $\boldsymbol{\theta}^* = \{e^{\alpha_0}, e^{\alpha_1}, e^{\alpha_2}, e^{\phi_0}, e^{\phi_1}\}$ and $\tau^* = e^\tau$. We can derive that

$$\left|\frac{d\boldsymbol{\theta}^*}{d\boldsymbol{\omega_I}}\right| \propto \frac{\tau^*(\omega_{11}^{(0)} + \omega_{1+}^{(1)}) + \omega_{10}^{(0)}}{(\omega_{10}^{(0)} + \tau^*\omega_{11}^{(0)})(\omega_{0+}^{(1)} + \omega_{1+}^{(1)} + \tau^*\omega_{11}^{(0)})}.$$

The a priori dependence of $p(\boldsymbol{\omega_I}|\tau)$ is thus introduced by $\left|\frac{d\boldsymbol{\theta}^*}{d\boldsymbol{\omega_I}}\right|$. This has been pointed out in Scharfstein et al. (2003) and explored further in Wang et al. (working paper).

References

A. Azzalini. Logistic regression for autocorrelated data with application to repeated measures. *Biometrika*, 81(4):767–775, 1994.

M. J. Daniels and J. W. Hogan. *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Chapman & Hall/CRC, 2008.

G.M. Fitzmaurice, N.M. Laird, and L. Shneyer. An Alternative Parameterization of the General Linear Mixture Model for Longitudinal Data with Non-ignorable Drop-outs. *Statistics in Medicine*, 20(7):1009–1021, 2001.

P.J. Heagerty. Marginally Specified Logistic-Normal Models for Longitudinal Binary Data. *Biometrics*, 55(3):688–698, 1999.

P.J. Heagerty. Marginalized Transition Models and Likelihood Inference for Longitudinal Categorical Data. *Biometrics*, 58(2):342–351, 2002.

R.J.A. Little. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88(421):125–134, 1993.

R.J.A. Little. A class of pattern-mixture models for normal incomplete data. *Biometrika*, 81(3):471–483, 1994.

G. Molenberghs, B. Michiels, MG Kenward, and PJ Diggle. Monotone missing data and pattern-mixture models. *Statistica Neerlandica*, 52(2):153–161, 1998.

J. Roy and M. J. Daniels. A general class of pattern mixture models for nonignorable dropout with many possible dropout times. *Biometrics*, 64:538–545, 2008.

D.O. Scharfstein, M.J. Daniels, and J.M. Robins. Incorporating prior beliefs about selection bias into the analysis of randomized trials with missing outcomes. *Biostatistics*, 4(4):495, 2003.

A.B. Troxel, G. Ma, and D.F. Heitjan. An Index of Local Sensitivity to Nonignorability. *Statistica Sinica*, 14(4):1221–1238, 2004.

C. Wang and M. J. Daniels. A note on identifying restriction in normal mixture models with and without covariates for incomplete data. working paper.

C. Wang, M. J. Daniels, and D. O. Scharfstein. Bayesian semiparametric selection model with application to a breast cancer prevention trial. working paper.

K.J. Wilkins and G.M. Fitzmaurice. A Hybrid Model for Nonignorable Dropout in Longitudinal Binary Responses. *Biometrics*, 62(1):168–176, 2006.

J. Zhang and D.F. Heitjan. A Simple Local Sensitivity Analysis Tool for Nonignorable Coarsening: Application to Dependent Censoring. *Biometrics*, 62(4):1260–1268, 2006.