



Published in final edited form as:

*Stat Neerl.* 2010 May 1; 64(2): 133–156. doi:10.1111/j.1467-9574.2009.00435.x.

## Semiparametric regression models and sensitivity analysis of longitudinal data with nonrandom dropouts

**David Todem,**

Division of Biostatistics, Department of Epidemiology, Michigan State University, B601 West Fee Hall, East Lansing, MI 48824, U.S.A.

**KyungMann Kim,**

Departments of Statistics and Biostatistics & Medical Informatics, University of Wisconsin-Madison, 600 Highland Ave., Madison, WI 53792, U.S.A.

**Jason Fine,** and

Departments of Statistics and Biostatistics & Medical Informatics, University of Wisconsin-Madison, 600 Highland Ave., Madison, WI 53792, U.S.A.

**Limin Peng**

Department of Biostatistics, Rollins School of Public Health, Emory University, 1518 Clifton Rd NE, Atlanta, GA 30322, U.S.A.

David Todem: todem@msu.edu

### Abstract

We propose a family of regression models to adjust for nonrandom dropouts in the analysis of longitudinal outcomes with fully observed covariates. The approach conceptually focuses on generalized linear models with random effects. A novel formulation of a shared random effects model is presented and shown to provide a dropout selection parameter with a meaningful interpretation. The proposed semiparametric and parametric models are made part of a sensitivity analysis to delineate the range of inferences consistent with observed data. Concerns about model identifiability are addressed by fixing some model parameters to construct functional estimators that are used as the basis of a global sensitivity test for parameter contrasts. Our simulation studies demonstrate a large reduction of bias for the semiparametric model relatively to the parametric model at times where the dropout rate is high or the dropout model is misspecified. The methodology's practical utility is illustrated in a data analysis.

### Key words and Phrases

Exponential family distribution; Functional estimators; Global sensitivity analysis; Informative dropout; Infimum/Supremum statistic; Nonparametric mixture; Uniform convergence; non-identifiable models

### 1. Introduction

Proper handling of missing data and dropouts in particular is critical in statistical analyses of longitudinal studies. It is well documented that improper handling of missing values may lead to misleading inferences (see, for example, Little and Rubin, 1987; Scharfstein et al., 1999; Rotnitzky et al., 2001; Kenward et al., 2001; and Hogan et al., 2004). Proper treatment

of missing data in statistical analysis depends primarily on the underlying missingness mechanism for which Little and Rubin (1987) provide a helpful terminology. Data are classified as missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR), if missingness is allowed to depend on (1) none of the outcomes, (2) the observed outcomes only, and (3) unobserved outcomes as well, respectively. From a modelling standpoint, most statistical approaches for handling missing data rely on the stronger MCAR or the less restrictive MAR assumptions (Hogan et al., 2004). Examples of such models in the context of longitudinal data are the generalized estimating equations-GEE models in their original formulation (Liang and Zeger, 1986) for the MCAR mechanism, and both the weighted GEE-based (Robins et al., 1995) and likelihood-based models (e.g., Laird and Ware, 1982) for the MAR mechanism. These models are commonly employed in practical data analytic settings, owing to their conceptual simplicity and ease of implementation. They are also known to generate valid inferences under the associated missing data mechanisms (Verbeke and Molenberghs, 2000). However, when the missingness mechanism depends on the unobserved outcomes, these procedures are known to produce biased inferences. To overcome this difficulty, Diggle and Kenward (1994) and Molenberghs et al. (1997) among others have proposed a model that incorporates both the information from the measurement process and the missing data process into a unified estimating function. This has provoked a large debate about the role for such models in understanding the true data generating mechanism. The original enthusiasm was followed by skepticism about the strong and untestable assumptions on which this type of models rests (Verbeke et al., 2001). Despite these limitations, more researchers recognize that these models should not be rejected but should be made part of a sensitivity analysis. Next to this issue, models that incorporate the measurement and dropout processes are usually not identifiable from observed data (see, for example, Troxel et al., 1998; Scharfstein et al., 1999; and Kenward et al., 2001). One then has to impose quantitative restrictions to recover identifiability. Conventional restrictions result from considering a minimal set of parameters, called sensitivity parameters, conditional upon which the remaining parameters are assumed identifiable. This method therefore produces a range of models which forms the basis of sensitivity analysis (Vach and Blettner, 1995). Over the years, numerous authors have proposed a local sensitivity approach to assess the impacts of uncertainties on inferences when the data are partially observed (Copas and Eguchi, 2001; Verbeke et al., 2001; Troxel et al., 2002; and Todem et al., 2006). The idea stems from assessing the effects of small perturbations of the MAR model in the direction of MNAR models. Where such methods have appeared in the literature, the sensitivity of identifiable parameters with respect to some fixed parameters is estimated via partial derivatives calculated in the neighborhood of a specified solution, typically at the MAR location. This local approach is important but a limitation is that only selected values of the sensitivity parameters are considered. In practice, it is often impossible to know the true value of the sensitivity parameters, making the global analysis conducted across all values of the sensitivity parameters the most conservative analytic strategy. Another limitation is that inference regarding the identifiable model parameters is ad hoc, ignoring that multiple tests are conducted and that the assumed values of the sensitivity parameters may be incorrect.

We consider a class of likelihood models for the measurement and dropout processes. Specifically, we formulate a generalized linear mixed effects model to describe the joint distribution of the measurement outcomes in time and then extend this basic model to allow for nonrandom dropouts. A novel formulation of a shared random effects model is introduced to capture the dropout dependence on the outcome process. Two forms of sensitivity analysis are considered. A qualitative sensitivity analysis is conducted to address concerns about the unverifiable nature of any working MNAR model using observed data. Specifically, we consider a nonparametric distribution as well as a fully parametric distribution for the random effects. Another line of sensitivity, essentially quantitative, is

performed to address concerns about model identifiability. We profile the model across a fixed parameter to delineate the range of inferences consistent with observed data. We propose a supremum test to conservatively evaluate the sensitivity of some parameter contrasts across all plausible values of the sensitivity parameter. This is particularly important when the quantity being tested does not increase or decrease monotonically as the sensitivity parameters are increased or decreased. Under monotonicity, it is only necessary to perform the tests at the limits of the sensitivity parameter space. However, in practice, missing data models are often quite complicated and it may not be clear whether monotonicity holds. In these set-ups, the analysis should be undertaken globally across all values of the sensitivity parameter.

The rest of this article is organized as follows. In section 2, we develop a class of shared latent variable models for longitudinal data with nonrandom dropouts, and discuss some key features of the models with related theoretical and asymptotic results. In Section 3, we propose a global sensitivity test based on a supremum hypothesis to assess simultaneously any perturbations of the MAR model in the direction of MNAR models. An illustration of the methodology using a psychiatric dataset is given in section 4. Section 5 presents a simulation study. Some remaining issues are discussed in section 6.

## 2. Shared latent variable models

### 2.1. Setup and notations

For each subject  $i = 1, \dots, n$ , there are  $q$  underlying outcomes represented by the vector  $Y_i(\mathcal{C}) = (Y_i(t_1), \dots, Y_i(t_q))'$ , measured respectively at discrete time points in the set  $\mathcal{C} = \{t_1, \dots, t_q\}$ , where  $t_1 < t_2 < \dots < t_q$ . These outcomes, however, are not always fully observed and therefore are coupled with a missingness indicator vector  $R_i(\mathcal{C}) = (R_i(t_1), \dots, R_i(t_q))'$ , where  $R_i(t) = 1$  if  $Y_i(t)$  is unobserved and 0 if otherwise. For a monotone missing data process,  $R_i(\mathcal{C})$  has the property that  $R_i(s) = 0$  whenever  $R_i(t) = 0$  for  $s \leq t$  and  $R_i(s) = 1$  whenever  $R_i(t) = 1$  for  $s \geq t$ . It is then intuitive to represent the series of indicators  $\{R_i(t) : t \in \mathcal{C}\}$  for subject  $i$  by the random variable  $D_i = 1 + \sum_{t \in \mathcal{C}} (1 - R_i(t))$ . We assume that for all  $i$ ,  $\text{pr}(D_i \geq 2) = 1$  which implies that all subjects are present at first time point. The random variable  $D_i$  then takes values  $2, \dots, q + 1$ , with  $q + 1$  being the realized value for a complete sequence.

### 2.2. Model formulation

We use a shared random parameter  $b_i$  coupled with the conditional independence assumption to specify the joint distribution of the measurement outcome and dropout response for subject  $i$ , denoted by  $\mathcal{L}(Y_i(\mathcal{C}), D_i) = \int \mathcal{L}(Y_i(\mathcal{C})|b_i)\mathcal{L}(D_i|b_i)d\mathcal{L}(b_i)$ . In this formulation, the generic label  $\mathcal{L}(u)$  denotes the law of  $u$  and  $\mathcal{L}(u|v)$  the conditional law of  $u$  given  $v$ .

The law of  $Y_i(\mathcal{C})$  for fixed  $b_i$  could be very general, but we restrict our formulation to the conditional independence assumption  $\mathcal{L}(Y_i(\mathcal{C})|b_i) = \prod_{j=1}^q \mathcal{L}(Y_i(t_j)|b_i)$ . The conditional law for the time point outcome is assumed to be generated from a generalized linear model with random effects. Specifically, we assume that the conditional mean  $\mu_i(t) = E\{Y_i(t)|b_i, x_i(t)\}$  of an hypothetical (observed or unobserved) response  $Y_i(t)$  at time point  $t$  for subject  $i$  is modeled as,

$$\mu_i(t) = g^{-1}\{x_i'(t)\beta + b_i\} \tag{1}$$

where  $\beta = (\beta_0, \dots, \beta_{p-1})'$  is the slope vector associated with covariates  $x_i(t)$  of dimension  $p \times 1$  and  $g(\cdot)$  is a monotone, differentiable and invertible function. The function  $g(\cdot)$  is assumed

known and is typically chosen to be the identity function for continuous outcomes, whereas for binary outcomes, the logit link is the natural choice.

To specify the law  $\mathcal{L}(D_i|b_i)$ , we assume that there exists an observed covariate  $z_i(t_j)$  that describes the discrete dropout hazard  $h_i(t_j) = \text{pr}(D_i = j | D_i \geq j, b_i, z_i(t_j))$  at timepoint  $t_j$ . Specifically, we consider the hazard model given by,

$$\log \{h_i(t)/(1 - h_i(t))\} = \varepsilon(z_i(t))\alpha + \delta\phi(b_i), \quad 0 < \phi(b_i) < 1, \tag{2}$$

where  $\varepsilon(z_i(t))$  is a function of the covariate  $z_i(t)$  of dimension  $r$  with associated coefficient  $\alpha$ , and  $\delta$  is the slope associated with  $\phi(b_i)$ ,  $\phi(\cdot)$  being a map from  $\mathcal{R}$  to the unit length open interval  $(0, 1)$ . One simple example of the function  $\phi(\cdot)$  is the nondecreasing function,  $\phi(b_i) = (1 + e^{-\varpi(b_i)})^{-1}$ , where  $\varpi(\cdot)$  is also a nondecreasing function. When  $b_i$  has an infinite support, we have  $\lim_{b_i \rightarrow -\infty} \varpi(b_i) = -\infty$  and  $\lim_{b_i \rightarrow \infty} \varpi(b_i) = \infty$ , with suitable modifications when  $b_i$  has a finite support. Any such function  $\varpi(\cdot)$  that meets the conditions above defines a function  $\phi(\cdot)$  and vice versa. Expressing  $\varpi(\cdot)$  in terms of  $\phi(\cdot)$  is straightforward and given by,  $\varpi(b_i) = \log \{ \phi(b_i) / (1 - \phi(b_i)) \}$  and is referred to as the logit function associated with  $\phi(\cdot)$ .

Finally, to complete the model formulation, the law of the random variable  $b_i$  denoted  $\mathcal{L}_b = \mathcal{L}(b_i)$  can be assumed parametric or left completely unspecified with a discrete support. That is,

$$\mathcal{L}_b \sim \mathcal{N}(0, \tau^2) \text{ or } \mathcal{L}_b \sim \sum_{m=1}^M \pi_m \mathcal{D}(\zeta_m), \tag{3}$$

where  $\tau^2$  is the unknown variance and  $\mathcal{D}(\zeta_m)$  is the Dirac measure placing point mass  $\pi_m$  at the single point  $\zeta_m$ , with  $\sum_{m=1}^M \pi_m = 1$ . These random effects distributions are viewed as a sensitivity analysis to address distribution misspecification. An advantage of the discrete law is that we do not introduce possibly inappropriate and unverifiable assumptions about the distribution of random effects. With the maximum number of identifiable latent classes, the mixing distribution may be interpreted as a nonparametric distribution (Laird, 1978). The discreteness of this law forces the random effect  $b_i$  to be one of the  $M$  unknown latent points,  $\zeta_1, \dots, \zeta_M$ . An important difference with the parametric model is that each subject does not have her own intercept in a latent class model. Instead, it is assumed that each subject belongs to one of the  $M$  latent classes and that each latent class has its own intercept. In practice, the number  $M$  is increased until the model fit no longer improves.

In the proposed model, the random effects term  $b_i$  is shared between the measurement outcome and the dropout process. This idea of using a shared latent variable model for modelling longitudinal outcomes subject to nonrandom dropouts is not new (see, for example, Wu and Carroll, 1988; Albert and Follmann, 2000; and Ten Have et al., 2002). A more recent illustration of this methodology is given by Beunckens et al. (2008). These authors have proposed a latent class model for incomplete longitudinal gaussian data viewed as a candidate model in their sensitivity analysis. Our extension of this technique is the introduction of the monotonic transformation  $\phi(\cdot)$  on the unobserved latent variable, which gives several attractive features to the dropout model. The constraint on the transformation is similar to that imposed by Rosenbaum (2002, page 107) on the unobserved covariate used to explain the hidden bias in assigning different treatments to subjects with same observed covariates. Our constraint may be seen as a restriction on the scale of the transformed unobserved latent variable, a restriction needed for  $\delta$  to have a meaningful interpretation. It

can be shown that the odds ratio of dropout hazard for two subjects with the same observed covariates for the dropout model is at most  $\exp\{|\delta|\}$ . For a sketch of the proof, we consider two subjects  $i$  and  $i'$  who have the same observed covariates with respect to the dropout model, say  $z_i(t) = z_{i'}(t)$  (or  $\varepsilon(z_i(t)) = \varepsilon(z_{i'}(t))$ ) at time point  $t$ . It can easily be shown (proof given in Appendix) that the odds ratio of dropout hazard between subjects  $i$  and  $i'$  is given by,

$$\{h_i(t)/(1 - h_i(t))\}\{h_{i'}(t)/(1 - h_{i'}(t))\}^{-1} = \exp\{\delta(\phi(b_i) - \phi(b_{i'}))\}.$$

In other words, two subjects with the same observed covariates for the dropout model differ in their odds of dropout hazard by a factor that involves the parameter  $\delta$  and the difference in their transformed unobserved covariates  $\phi(b_i) - \phi(b_{i'})$ . Giving the constraint imposed on the function  $\phi(\cdot)$ , it follows that the ratio above is bounded as follows,

$$\exp\{-|\delta|\} \leq \{h_i(t)/(1 - h_i(t))\}\{h_{i'}(t)/(1 - h_{i'}(t))\}^{-1} \leq \exp\{|\delta|\}.$$

Hence, for  $\delta = 0$  we have a dropout at random mechanism where two subjects with same observed dropout covariates have equal dropout hazard odds. When  $\delta = \log(3)$ , two subjects with same observed dropout covariates could differ in their instantaneous dropout odds by a factor as much as 3.

### 2.3. Model identifiability

It is well known that nonrandom dropout models are typically not identifiable from observed data (see, for example, Scharfstein et al., 1999; and Kenward et al., 2001). Both simulation-based studies and analysis of actual data show that nonrandom models often have likelihood functions with flat surfaces and/or with multiple modes. We study some implications of these identifiability concerns for the shared parameter model. For this, let

$\mathcal{L}_b^o = \mathcal{L}(b|D > q)$  and  $\mathcal{L}_b^d = \mathcal{L}(b|D \leq q)$  be the laws of the random effects  $b$  for complete and incomplete data subjects, respectively. Here, we have suppressed the subject index for simplicity. The law  $\mathcal{L}_b$  of the random effects can be written as,

$$\mathcal{L}_b = (1 - p_d)\mathcal{L}_b^o + p_d\mathcal{L}_b^d, \tag{4}$$

where  $p_d = P(D \leq q)$ . If many independent copies of the indicator variable  $I(D \leq q)$  are available,  $p_d$  can be estimated very well. Although the random effects are not observed, the law  $\mathcal{L}_b^o$  can also be well estimated using the law of  $Y(\mathcal{C})$  among completers. It can be shown, using a simple probability argument, that

$\mathcal{L}_b^o = \int \mathcal{L}(b|Y(\mathcal{C}), D > q)\mathcal{L}(Y(\mathcal{C})|D > q)d\mathcal{L}(Y(\mathcal{C}))$ . The law  $\mathcal{L}(Y(\mathcal{C})|D > q)$  is easily identifiable from observed data. The law  $\mathcal{L}(b|Y(\mathcal{C}), D > q)$  is also identifiable using the fundamental assumption of conditional independence of measurement responses given random effects. A typical estimate of  $\mathcal{L}(b|Y(\mathcal{C}), D > q)$  is the nonparametric maximum likelihood estimate, a discrete distribution with at most  $n - n_d$  support points,  $n_d = \sum_i I(D_i \leq q)$  being the number of dropout subjects (Lindsay, 1983; and Davidian and Giltinan, 1998). The unidentifiable component of the model is clearly

$\mathcal{L}_b^d = \int \mathcal{L}(b|Y(\mathcal{C}), D \leq q)\mathcal{L}(Y(\mathcal{C})|D \leq q)d\mathcal{L}(Y(\mathcal{C}))$ , the law of  $b$  for dropout subjects, as the law  $\mathcal{L}(Y(\mathcal{C})|D \leq q)$  is unidentifiable from observed data. The theorem below gives some implications of this nonidentifiability in a setup where  $q = 2$ .

**THEOREM 1.** Assume that the study has only two planned time points,  $q = 2$ , and that the laws  $p_d$  and  $\mathcal{L}_b^c$  are all known. The fact that the law  $\mathcal{L}_b^d$  is not identifiable translates into  $\alpha_0$ , the intercept of the hazard model, and  $\delta$  being related by the deterministic relationship,

$$\exp\{\alpha_0\} = (1 - p_d)p_d^{-1} \int \exp\{-\delta\phi(b)\} d\mathcal{L}_b^o(b). \tag{5}$$

Here, the integration  $d\mathcal{L}_b^o(b)$  is with respect to the Lebesgue measure on the real line for the normal latent variable and to the counting measure for the discrete nonparametric distribution. The proof of this theorem (given in Appendix) is similar to that of Freedman (1999) in the context of dropout selection models. When  $q > 2$ , from our experience the joint model is often at best weakly identifiable especially when  $M$  gets large. Such overspecification of the model can be managed in a more general way by considering a minimal set of parameters, conditional upon which the other parameters are estimable. It is important, however, to realize that the choice of the sensitivity parameter is non-unique and can be a difficult task (Kenward et al., 2001). In general, sensitivity parameters are chosen from parameters that are not of primary interest. Ideally, the choice should be closely linked to the substantive problem under analysis. A natural choice for the sensitivity parameter in our joint model is  $\delta$ , the parameter that measures the extent of nonrandomness of the dropout process.

**2.4. Functional semiparametric inferences via EM algorithm**

Let  $\psi = (\beta', \alpha', \zeta', \pi')'$  be the identifiable parameter vector of the semiparametric model, with  $\zeta = (\zeta_1, \dots, \zeta_M)'$  and  $\pi = (\pi_2, \dots, \pi_M)'$ . Since  $M$  free parameters are considered for  $\zeta$ , no overall intercept is allowed in the vector  $\beta$  for the model in (1) to ensure identifiability. The maximum likelihood estimate (MLE) of the parameter vector  $\psi$  can be obtained by means of the EM algorithm introduced by Dempster, Rubin and Laird (1977), since direct maximization of integrated likelihood is difficult.

We denote by  $Y_i(\mathcal{O}_i)$ , where  $\mathcal{O}_i = \{t_1, \dots, t_{D_i-1}\}$ , the collection of all observed outcomes for subject  $i$  and by  $Y_i(\mathcal{M}_i)$ , where  $\mathcal{M}_i = \{t_{D_i}, \dots, t_q\}$  for  $D_i \leq q$ , the collection of all missing random outcomes. We also denote by  $x_i(\mathcal{O}_i)$  and  $z_i(\mathcal{O}_i)$ , respectively the collection of all fixed effects covariates for the measurement process and dropout hazard. Define the indicator random variables for the class membership as follows:  $\mathcal{U}_{im} = 1$  if the  $i$ th subject belongs to the  $m$ th class and 0 if otherwise. It follows from this definition that  $E(\mathcal{U}_{im}) = \pi_m$ . We consider the complete data  $(\mathcal{U}_i, Y_i(\mathcal{O}_i), D_i)$ , where  $\mathcal{U}_i = (\mathcal{U}_{i1}, \dots, \mathcal{U}_{iM})'$  and the observed data  $W_i = \{Y_i(\mathcal{O}_i), D_i, z_i(\mathcal{O}_i), x_i(\mathcal{O}_i)\}$ . Unlike other EM algorithms for non-response models, our approach does not involve the missing component  $Y_i(\mathcal{M}_i)$  when  $D_i \leq q$ . Indeed, by assuming that the possibly incomplete random vector  $Y_i(\mathcal{C})$  and the dropout variable  $D_i$  are conditional independent given the latent variable, the integration of the conditional likelihood with respect to the law of missing data results in a closed form.

We denote by  $\ell(\mathcal{U}_i, Y_i(\mathcal{O}_i), D_i)$  the contribution of subject  $i$  to the complete data likelihood function. Suppose that the current estimate, at step  $a$ , of the parameter vector is  $\psi^{(a)}(\delta)$  for a fixed  $\delta$ . The E-step of the EM algorithm involves computing the expectation of the complete data log-likelihood given observed data and the current estimate for a fixed  $\delta$ . The contribution of subject  $i$  to this conditional expectation is given by,



$$E \left\{ \log \{ \ell(\mathcal{Z}_i, Y_i(\mathcal{O}_i), D_i) | W_i, \delta, \psi^{(a)}(\delta) \} \right. \\ \left. = \sum_{m=1}^M \sum_{t \in \mathcal{O}_i} \pi_{im}(\psi^{(a)}(\delta)) \log \{ f_Y(Y_i(t) | \zeta_m, x_i(\mathcal{O}_i), \psi) \} + \sum_{m=1}^M \pi_{im}(\psi^{(a)}(\delta)) \log \{ f_D(D_i | \zeta_m, z_i(\mathcal{O}_i), \delta, \psi) \} + \sum_{m=1}^M \pi_{im}(\psi^{(a)}(\delta)) \log \{ \pi_m \} \right.$$

Here  $\pi_{im}(\psi^{(a)}(\delta)) = E \{ \mathcal{Z}_{im} | W_i, \delta, \psi^{(a)}(\delta) \}$  is the posterior probability for the  $i$ th subject to belong to the  $m$ th latent class and given by,

$$E \left\{ \mathcal{Z}_{im} | W_i, \delta, \psi^{(a)}(\delta) \right\} = \frac{\pi_m \prod_{t \in \mathcal{O}_i} f_Y(Y_i(t) | \zeta_m, x_i(\mathcal{O}_i), \psi^{(a)}(\delta)) f_D(D_i | \zeta_m, z_i(\mathcal{O}_i), \delta, \psi^{(a)}(\delta))}{\sum_{v=1}^M \pi_v \prod_{t \in \mathcal{O}_i} f_Y(Y_i(t) | \zeta_v, x_i(\mathcal{O}_i), \psi^{(a)}(\delta)) f_D(D_i | \zeta_v, z_i(\mathcal{O}_i), \delta, \psi^{(a)}(\delta))}$$

The term  $f_Y(Y_i(t) | \zeta_m, x_i(\mathcal{O}_i), \psi)$  is the time point conditional density of the measurement process. The quantity  $f_D(D_i | \zeta_m, z_i(\mathcal{O}_i), \delta, \psi) = [\prod_{t \in \mathcal{O}_i} (1 - h_{i|m}(t))] \{ h_{i|m}(t_{D_i}) \}^{I(D_i \leq q)}$ , with  $h_{i|m}(t) = \{ 1 + \exp(-\varepsilon(z_i(t))\alpha - \delta \phi(\zeta_m)) \}^{-1}$  and setting  $h_{i|m}(t_1) = 0$  and  $h_{i|m}(t_{q+1}) = 1$ , is the dropout probability when the  $i$ th subject belongs to class  $m$ . In the M-step of the algorithm, identifiable model parameters are updated so that the expected log-likelihood is maximized.

An update of the parameter  $\pi_m$ , is given by  $\pi_m^{(a+1)}(\delta) = \frac{1}{n} \sum_{i=1}^n \pi_{im}(\psi^{(a)}(\delta))$ . Maximization of the expected complete likelihood with respect to  $\beta$  can be done using standard algorithms for the maximum likelihood estimation of generalized linear models.

An important issue to be confronted is the choice of  $M$ , the number of random effects classes in the discrete model in (3). It is well known that this number is not an interior point of a convex space, which rules out the use of the likelihood ratio approach (McLachlan and Peel, 2000). To reduce complexity, together with  $\delta$  we held  $M$  fixed for each model and used goodness-of-fit tools such as the Bayesian information criterion to select the appropriate number of classes (see, for example, Beunckens et al., 2008).

### 2.5. Asymptotic behaviors of the estimator of $\psi$

We now study the asymptotic behavior of the estimator of  $\psi$  when  $\delta$  the parameter which describes the degree of selection bias in the model is predetermined. First, we show that it is necessarily for  $\delta$  to be bounded, otherwise the marginal log-likelihood function diverges.

**THEOREM 2.** *We assume that  $0 < \inf_{i,t} |\varepsilon(z_i(t))\alpha| \leq \sup_{i,t} |\varepsilon(z_i(t))\alpha| < \infty$ . When  $\delta \rightarrow \infty$ , the profile log-likelihood  $\log \ell(\psi, \delta) \rightarrow -\infty$  and a finite maximizer does not exist.*

Here  $\ell(\psi, \delta)$  denotes the likelihood function for a fixed  $\delta$ . Note here that the assumptions of the theorem are also verified under the stronger condition of a compact space for the parameter vector  $\alpha$  and bounded associated transformed covariates. The theorem result (proof given in Appendix) also holds for the expected complete data log-likelihood given observed data and the current estimate for a fixed  $\delta$ . This result implies that when  $\delta$  is unbounded, a finite MLE of  $\psi$  does not exist. This suggests that the parameter space of  $\delta$  needs to be bounded. In other words, one needs to assume that  $\delta \in \mathcal{H}$ , a bounded set. Without any loss of generality we will restrict  $\mathcal{H}$  to the closed interval  $[0, \Delta]$ , with  $0 \leq \Delta < \infty$ . In practice, subject matter expert should be consulted to decide on the choice of the largest plausible value of  $\delta$ . From a technical standpoint, this choice should be computationally feasible. This limiting result is different from that of Scharfstein et al. (1999) where the parameters of interest converge to fixed values as the sensitivity parameter becomes infinitely large.

Let  $\delta_0$  be the true value of  $\delta$  from the joint model. If  $\delta_0$  is known, the MLE of  $\psi_0$  (the true value of  $\psi$ ) denoted  $\hat{\psi}(\delta_0)$  can be obtained by maximizing the profile likelihood with  $\delta$  fixed at the true value  $\delta_0$ . As the sample size  $n$  gets large, the maximum likelihood theory (see for example, Casella and Berger, 1990) ensures that,  $\hat{\psi}(\delta_0)$  is consistent for  $\psi_0$ , that is  $\hat{\psi}(\delta_0) \rightarrow_p \psi_0$ , and  $n^{\frac{1}{2}} \{ \hat{\psi}(\delta_0) - \psi_0 \} \rightarrow_d \mathcal{N}(0, \Sigma_0)$ , with  $\Sigma_0$  being the associated asymptotic variance-covariance matrix. Under a misspecified dropout process, i.e.  $\delta \neq \delta_0$ , and for a large  $n$ , we have  $\hat{\psi}(\delta) \rightarrow_p \psi^*(\delta)$ , with  $\psi^*(\delta)$  not necessarily equal to  $\psi_0$ . Moreover, the profile score functions are roughly quadratic in the neighborhood of  $\psi^*(\delta)$  for fixed  $\delta$  and the limiting distribution of  $n^{\frac{1}{2}} \{ \hat{\psi}(\delta) - \psi^*(\delta) \}$  is  $\mathcal{N}(0, \Sigma^*(\delta))$ , with  $\Sigma^*(\delta)$  being the asymptotic variance-covariance matrix. These pointwise weak convergence results are supported by the theory on empirical processes (see for example van der Vaart and Wellner, 2000). They can be made uniform across the space of  $\delta$  under certain smoothness conditions (see details in Appendix). As the covariance function  $\Sigma^*(\delta)$  characterizing the limiting distribution of  $\hat{\psi}(\delta)$  is quite complicated, the bootstrap may be used to calculate the asymptotic standard errors of parameter estimates from a misspecified model (Efron and Tibshirani, 1993). The bootstrap is especially useful when simultaneous inference about  $\psi^*(\delta)$  for  $\delta \in \mathcal{H}$  is of interest and cannot be carried out analytically. In the next section, we discuss a global sensitivity analysis in which inferences are conducted simultaneously across the range of  $\delta$ .

### 3. Global sensitivity tests

We consider the contrast  $C\psi$ , where  $C$  is an  $l \times (p + r)$  matrix which defines a general framework for estimating single and multiple linear combinations of model parameters. As an example, in the special case of evaluating the effect of the  $j$ th covariate, one takes  $C$  to be a  $1 \times (p + r)$  vector with a one in the  $j$ th position and zeros elsewhere. Suppose we are interested in assessing the sensitivity of the contrast  $C\psi^*(\delta)$  relative to  $C\psi^*(0)$  across all values of  $\delta$ . Formally, this global sensitivity hypothesis can be formulated as follows,

$$H_0: \sup_{\delta \in \mathcal{H}} \|C\psi^*(\delta) - C\psi^*(0)\| = 0,$$

where  $\|\cdot\|$  represents the Euclidean norm. This hypothesis can be used to assess any perturbation of the MAR model in the direction of the MNAR models. When the supremum hypothesis is not rejected say at 5% significance level it can be concluded that there is no evidence from observed data to reject  $\|C\psi^*(\delta_0) - C\psi^*(0)\| = 0$ . This follows from the trivial inequality,

$$\|C\psi^*(\delta_0) - C\psi^*(0)\| \leq \sup_{\delta \in \mathcal{H}} \|C\psi^*(\delta) - C\psi^*(0)\|.$$

In this case,  $C\hat{\psi}(0)$  can be used to make rigorous inferences about  $C\psi^*(\delta_0)$ . However, when the supremum hypothesis is rejected, a sensitivity analysis is then carried out to identify values of  $\delta$  for which the hypotheses  $\|C\psi^*(\delta) - C\psi^*(0)\| = 0$  are rejected. Our global analysis then asks how would inferences about the contrasts  $C\psi^*(\delta) - C\psi^*(0)$  be altered by nonrandom dropouts of a magnitude defined by  $\delta$ ?

The supremum test can easily be performed using a nonparametric bootstrap approach. Let  $0 < \gamma < 1$ , the hypothesis  $H_0$  is rejected at level  $\gamma$  when the observed test value is greater than the  $(1 - \gamma)$  percentile of the bootstrap samples of the test statistic  $T_{sup} = \sup_{\delta \in \mathcal{H}} \|C\hat{\psi}(\delta) - C\hat{\psi}(0)\|$  under the null hypothesis. In the special case of estimating the effect of a one-dimensional covariate say  $C\psi = \beta_1$ , simultaneous confidence intervals of  $\beta_1^*(\delta) - \beta_1^*(0)$  across all values of  $\delta$  can be used to evaluate the sensitivity hypothesis. Let  $\tilde{\beta}_1^s(\delta)$  and  $\tilde{\beta}_1^s(0)$ ,



$s = 1, \dots, S$ , be bootstrap samples of  $\hat{\beta}_1(\delta)$  and  $\hat{\beta}_1(0)$ , respectively. A simultaneous confidence interval for  $\beta_1^*(\delta) - \beta_1^*(0)$ , for all  $\delta \in \mathcal{H}$ , takes the form

$$\{u(\delta): \mathcal{H} \rightarrow R; -\tilde{\vartheta}_\gamma + \hat{\beta}_1(\delta) - \hat{\beta}_1(0) < u(\delta) < \tilde{\vartheta}_\gamma + \hat{\beta}_1(\delta) - \hat{\beta}_1(0)\},$$

where  $\tilde{\vartheta}_\gamma$  is the  $(1 - \gamma)$ th empirical percentile of

$\{\sup_{\delta \in \mathcal{H}} |(\hat{\beta}_1^s(\delta) - \hat{\beta}_1^s(0)) - (\hat{\beta}_1(\delta) - \hat{\beta}_1(0))|\}_{s=1}^S$ . When  $H_0$  is rejected, the interval  $(\inf_{\delta \in \mathcal{H}} \hat{\beta}_1(\delta), \sup_{\delta \in \mathcal{H}} \hat{\beta}_1(\delta))$  gives the minimum and the maximum sizes of the estimate of  $\beta_1$  over all values of the selection bias due to dropouts (see, for example, Kenward et al., 2001). This interval due to dropout uncertainty can be extended to incorporate uncertainty due to sampling imprecision by computing a lower and an upper confidence bound for  $\inf_{\delta \in \mathcal{H}} \beta_1^*(\delta)$  and  $\sup_{\delta \in \mathcal{H}} \beta_1^*(\delta)$ , respectively.

#### 4. Analysis of psychiatric data

The Fluvoxamine (a serotonin reuptake inhibitor) trial is a multi-center non-comparative study, designed to reflect clinical practice closely with out-patients diagnosed with depression, obsessive-compulsive disorder or panic disorder. Accumulated experience in controlled trials has shown that Fluvoxamine is as effective as conventional tricyclic antidepressant drugs, and more effective than placebo in the treatment of depression (for a review, see Burton, 1991). However, many patients suffering from depression have concomitant morbidity associated with this condition. It was then decided to set up a post-marketing pharmaco-vigilance trial to study more accurately the profile of Fluvoxamine in ambulatory clinical psychiatric practice. A total of 315 patients with a diagnosis of either depression or obsessive-compulsive or panic disorder were enrolled in the study. All subjects were treated with Fluvoxamine in doses ranging from 100 to 300 mg/day and underwent clinical evaluations at baseline, 2, 4, 8 and 12 weeks. One primary endpoint comprised the side effects of the drug recorded on an ordinal scale. A side effect occurs if new symptoms appear. Several patient's baseline characteristics such as sex, age, initial severity of the disease on a 1 to 7 scale, and duration of the mental illness were recorded. One key objective of the study was to assess the within-subject evolution of side effects over time adjusted for baseline covariates. In particular, the drug company was interested in describing the side effects' profile for each study participants for a possible change in doses. A full description of the study is given by Molenberghs and Lesaffre (1994), Lesaffre et al. (1996) and Kenward et al. (1994).

Out of 315 patients, 224 patients had a full-sequence data resulting from the fact that 14 subjects were not observed after recruitment, 31, 26, and 18 patients dropped out, respectively, after the first, second and third visit and 2 patients had a non-monotone missing pattern. As our analysis focuses only on dropouts, we ignore the 2 cases that had a non-monotone missing pattern and ignore all 14 patients who dropped out before the first visit. Compared to the start of the study, a reduction of side effects is depicted throughout the study for both the completers and the non-completers (see Figure 1). Specifically, about 54% of patients who complete the study had some side effects at 2 months compared to 66% for subjects who were present at 2 months but dropped out before the next visit. At 4 months, these numbers were 44% for completers compared to 62% for noncompleters and at 8 months, 35% for completers compared to 44% for noncompleters. It is clear that study noncompleters are likely to be doing poorly with respect to side effects. This makes the MNAR process a plausible missingness mechanism. Therefore, a naive reasoning that ignores this selection process may lead to an incorrect and over optimistic conclusion. A question then emerges. How do side effects evolve over time accounting for this potentially

nonrandom dropout mechanism? To answer this question, we consider the shared latent variable model described in section 2 to assess the effect of time on the outcome variable for various values of the parameter which measures the extent of nonrandomness of the dropout process.

Our analysis of the within-subject evolution (captured by time effects) is based on a dichotomized version (presence/absence) of side effects. Considering an ordinal outcome rather than binary would only add unnecessary complexity to the analysis. For this, we consider a simple Bernoulli model for which the conditional mean is given by,

$$E\{Y_i(t)|b_i, x_i(t)\} = \{1 + \exp(-\beta x_i(t) - b_i)\}^{-1},$$

where  $Y_i(t) = 1$  if new symptoms occur and 0 if otherwise, and  $\beta$  is the slope vector associated with the design covariate vector  $x_i(t)$ . The vector  $x_i(t)$  contains the linear time variable and the baseline variables (sex, age at enrolment, initial severity of the disease on a 1 to 7 scale, and duration of the mental illness at enrolment). We extend the proposed model to account for nonrandom dropouts as follows,

$$\log \{h_i(t)/(1 - h_i(t))\} = \alpha x_i(t) + \delta \phi(b_i),$$

where  $\alpha$  is the slope associated with  $x_i(t)$  and  $\phi(b_i) = 1/(1 + e^{-b_i})$ . Here the function  $\omega(\cdot)$  is set to identity.

We are interested in the parameter vector  $\beta$ , specifically the slope parameter say  $\beta_1$  of the linear time variable. We attempted to fit the parametric and the semiparametric models to the data without fixing  $\delta$  but the computations were unstable, especially for large  $M$ . Multiple starting values were tried. In some cases, the algorithm diverged, while in cases where it did converge, multiple local maxima were obtained for the full likelihood. This suggests that these models are at best weakly identified for our data. To address these identifiability concerns, we then fixed  $\delta$  and performed the supremum test described in section 3 to assess sensitivity of the parameter  $\beta_1$  across all values of  $\delta$ , under the parametric and the semiparametric models. For this, we computed estimates of  $\beta_1^*(\delta) - \beta_1^*(0)$  and associated 95% simultaneous confidence intervals across all values of  $\delta \in \mathcal{A}$ . These estimates and associated 95% simultaneous confidence intervals are graphically represented in Figures 2(a) and 2(b), respectively for the semiparametric and the parametric working models. These graphs are obtained by solving the profile score equations for fixed values of  $\delta$  on a grid and mid-points estimators are then interpolated via smoothing. The 95% critical points (.14 and .26 for the semiparametric and parametric models, respectively) to compute the simultaneous confidence intervals were obtained using 1000 bootstrap samples. Simultaneous confidence intervals do not contain 0 for  $\delta \leq 40$  under the semiparametric model and for  $\delta \leq 30$  under the parametric model. Hence, unless the true odds ratio of dropout hazard for two subjects with same fixed covariates is larger than  $\exp(40)$  and  $\exp(30)$  for the semiparametric and parametric models respectively, the linear time effect appears to be relatively stable under any perturbations of the MAR model in the direction of MNAR models. The evidence from our investigation speaks against dropout being at random in the Fluvoxamine data, but the sensitivity analysis results suggest that the adjusted slope of time effects are not much influenced by  $\delta$  in light of sampling variations. The MAR model is therefore preferable for its relative simplicity. From a clinical standpoint, our MAR analysis suggests that the intensity of side effects declines with time. For each unit increase in time, the estimated adjusted odds for not showing some side effects can increase by a factor (95% confidence interval in parentheses) of 2.071(1.648, 2.603) for the

semiparametric model and 2.109(1.671, 2.662) for the parametric model. This example illustrates how a global approach can be used to assess sensitivity of parameters of primary interest when some model characteristics are potentially unidentifiable from observed data. Moreover, it also illustrates how a sensitivity analysis can be conducted on the random effects distribution for model misspecification.

## 5. Simulations

In this section we report results of a numerical study conducted to evaluate the small sample performance of the functional estimators proposed in this paper when the dropout process is nonrandom. The simulations were conducted so to roughly approximate the binary outcomes from the Fluvoxamine study. In each Monte Carlo iteration, we simulated a sample of 100 subjects with four potential measurement time points ( $q = 4$ ). The measurement outcomes were simulated using the Bernoulli model,  $f_Y(Y_i(t)|b_i, x_i(t)) = \mu_i(t)^{Y_i(t)}\{1 - \mu_i(t)\}^{1-Y_i(t)}$ , where  $\mu_i(t) = \{1 + \exp(-1 + x_i(t) - b_i)\}^{-1}$ , with  $x_i(t) = t$  taking values in the set  $\{1, 2, 3, 4\}$ . To keep the simulation simple, the dropout observations were generated using a time independent dropout hazard model given by,  $\log \{h_i(t)/(1 - h_i(t))\} = \alpha_0 + \delta_0/(1 + e^{-b_i})$ . This simplified assumption may be seen as too restrictive as missingness is typically related to things which happen as the study is ongoing. Throughout the simulations, we also fixed  $\delta_0 = 1$ . We produced 20% and 50% dropout rates by setting the values of  $\alpha_0$  to 2.094 and .855, respectively. Two different distributions with a variance of 4.025 for the random effects  $b_i$  were considered: a Gaussian distribution,  $\mathcal{L}_b \sim \mathcal{N}(0, 4.025)$  and a bimodal mixture of two

Gaussian distributions,  $\mathcal{L}_b \sim \frac{1}{2} \mathcal{N}(-2, .025) + \frac{1}{2} \mathcal{N}(2, .025)$ . This process was repeated for 250 Monte Carlo replications. Our working measurement model is the Bernoulli model with the conditional mean,

$$\mu_i(t) = \{1 + \exp(-\beta_0 - \beta_1 x_i(t) - b_i)\}^{-1}. \tag{6}$$

The working dropout model is also given by,

$$\log \{h_i(t)/(1 - h_i(t))\} = \alpha + \delta \phi(b_i), \tag{7}$$

where  $\phi(b_i) = 1/(1 + e^{-b_i})$ . Both the semiparametric and the parametric models were then assessed for various values of  $\delta \in [0, 40]$ . In particular, we used the percentage bias relative to the truth and mean squared errors (MSEs) for the slope  $\beta_1$  of the time covariate.

Figure 3 shows the results of this simulation study when the measurement and dropout data are generated from a shared normal random effects model. Overall, higher dropout rates increase the bias as well as the MSEs of the estimates for both working models. Note however that these differences recede when  $\delta$  is close to the true value  $\delta_0 = 1$  and the two working models have similar performances in terms of bias and MSEs. However, for moderate to large deviations of  $\delta$  from  $\delta_0 = 1$ , the normal random effects approach gives higher values of percent bias and MSEs when the dropout rate is about 50%.

Also displayed in Figure 4 are simulation results from a Monte Carlo study where the measurement and dropout data are generated from a shared bimodal random effects model. As expected, higher dropout rates increase the bias as well as the MSEs of the two working models. We notice however that, even at the true value  $\delta_0 = 1$ , a wrongly assumed normal random effects model gives a slightly higher percent bias compared to the semiparametric model. This finding suggests that the misspecification of the random effects distribution by a parametric model will typically yield biased estimates even at the true value  $\delta_0$ . This is not

uncommon in the literature (See, for example, Tao et al., 1999). Note however that, this problem recedes when MSEs are compared. This is probably due to an increased variability which dominates the MSEs in the semiparametric model. When the dropout rate is high and  $\delta$  is much larger than  $\delta_0 = 1$ , this relative advantage of the normal random effects model in reducing the MSEs vanishes. In this case, the MSEs appear to be dominated by the bias.

While the ability to model the dropout hazard using the shared random effects approach is computationally convenient, we carried out a simulation to evaluate the estimation of  $\beta_1$  when miss-modelling  $h_i(t)$  with a shared random effects when the truth is actually a logistic model of the current possibly unobserved response  $Y_i(t)$ . Specifically, we simulate the measurement binary outcomes by reconsidering the Bernoulli model with the conditional mean  $\mu_i(t) = \{1 + \exp(-1 + x_i(t) - b_i)\}^{-1}$ , but dropout data are simulated using the process,

$$\log \{h_i(t)/(1 - h_i(t))\} = \alpha_0 + Y_i(t).$$

This selection model is similar to that of Diggle and Kenward (1994), and Molenberghs et al. (1997). We set  $\alpha_0$  to 2.094 and .855, which yield about 20% and 50% dropout rates, respectively. Our working joint model is given by (6) and (7). The results of this numerical investigation are displayed in Figures 5 and 6. It is clear that for normal random effects imposed on the mean of the measurement process and for a 20% dropout rate, both the parametric and the semiparametric models give small bias and small MSEs regardless of the values taken by  $\delta$ . However, for a dropout rate of about 50%, the two methods yield large bias and large MSEs. Moreover, the bias and MSEs from the semiparametric model are lower compared to those of the parametric model for large values of  $\delta$ , although the two models are indistinguishable for small values of  $\delta$ . These findings remain unchanged when a bimodal random effects is imposed on the mean of the measurement process, except that for a high dropout rate, the semiparametric model gives smaller bias and smaller MSEs for all values of  $\delta$ .

These simulation studies strongly indicate that routine modelling is not appropriate in the missing data context and sensitivity analysis should become much more customary. Especially, when substantial amounts of subjects drop out prematurely from the study, the only analysis that matters is the sensitivity analysis.

## 6. Discussion

We have proposed a class of semiparametric non-response models which requires the analyst to specify the joint distribution of the responses and the missing data mechanism. This class of models is made part of a sensitivity analysis to delineate the range of inferences consistent with observed data. Two types of sensitivity analyses are conducted. The first sensitivity analysis is performed on the grounds that the parametric form of the random effects distribution may be misspecified. This is done by relaxing the random effects parametric distribution to assume a discrete distribution with  $M$  support points. In certain situations, one may want to assume fully parametric random effects distributions, for instance, because they yield a more parsimonious description of the heterogeneity among subjects. For MNAR models, however, these parametric distributions must be cautiously interpreted because of their inability to be validated using observed data. Another line of sensitivity analysis results from the fact that the proposed models, both parametric and semiparametric, appear to be weakly identifiable for our data. The identifiability issue becomes acute as  $M$  the number of classes in the semiparametric models gets large. To address these identifiability concerns, we profile the likelihood function across a fixed sensitivity parameter. We discuss a global approach using a supremum test to assess

sensitivity of parameter contrasts resulting from a deviation of the MAR model in the direction of MNAR models. This approach can be used to quickly check if an hypothesis is rejected regardless of any misspecification of the sensitivity parameter. The methodology is therefore especially useful in situations where a worst case analysis is needed. In practice, MNAR models are often quite complicated and it may not be clear what is the true data generating mechanism. In these set-ups, a global sensitivity analysis appears to be the most conservative data analytic strategy.

An important question related to our working method is whether we actually believe in the shared parameter model or whether we merely use it as a device to accommodate potentially non-random dropouts. This question arises very often in practice where the researcher is confronted with the choice of a plausible dropout model. In this study, we have used random effects, that adequately describe a subject's response profile, to tie the dropout process to the measurement process. Although we have considered the simple case of a random intercept model, the method can easily be extended to incorporate a random slope with respect to time to account for each subject's deviation from the average linear trend. We have also assumed a linear relationship between the log-odds of dropout hazard and the observed and the unobserved covariates, but this is not necessary. For example, nonlinear relations and terms involving interactions between these covariates could be used with a suitable modification of the function  $\phi(\cdot)$ . Other model formulations and extensions are possible but the choice of any working model should be closely linked to the substantive problem under analysis. Another important issue is the choice of  $\Delta$ . Our recommendation is that a group of experts should help identify a plausible and meaningful value of  $\Delta$ . Of course, from a practical viewpoint, the chosen value should be computationally feasible. The dropout selection process is therefore said to be insensitive if only extreme and implausible values of the sensitivity parameter are required to alter inferences.

## Acknowledgments

The authors wish to thank Solvay Pharma B.V. for permission to use the data from the Fluvoxamine study. They also wish to thank the anonymous reviewer for his or her helpful comments.

This work was supported by the NCI/NIH K-award, 1K01 CA131259 and the Michigan State University Intramural Research Grant Program-IRGP.

## Appendix

### Appendix A

*Proof of the result that  $\exp(-|\delta|) \leq \frac{h_i(t)(1 - h_r(t))}{h_r(t)(1 - h_i(t))} \leq \exp(|\delta|)$  for any real  $\delta$  under the condition that  $z_i(t) = z_r(t)$*

The proof of this result is straightforward and relies entirely on the constraint imposed on  $\phi(\cdot)$ . If one assumes the covariate-response model in (1) and that  $0 < \phi(b_i), \phi(b_r) < 1$ , it can readily be shown that  $|\phi(b_i) - \phi(b_r)| < 1$ . Under the condition that  $z_i(t) = z_r(t)$ , this then leads

to the result that  $\exp(-\delta) < \frac{h_i(t)(1 - h_r(t))}{h_r(t)(1 - h_i(t))} < \exp(\delta)$  for  $\delta > 0$  and that  $\frac{h_i(t)(1 - h_r(t))}{h_r(t)(1 - h_i(t))} = 1$  for  $\delta = 0$ . We then have the desired result for  $\delta \geq 0$ . For  $\delta \leq 0$ , we have,

$$\exp(\delta) \leq \frac{h_i(t)(1 - h_r(t))}{h_r(t)(1 - h_i(t))} \leq \exp(-\delta), \text{ which is the desired result.}$$

## Appendix B

### Proof of Theorem 1

We prove that the intercept of the dropout process has a deterministic relationship with the parameter  $\delta$  as given by (5). First, it is clear that the integral on the right of (5) is finite as  $\mathcal{L}_b^o$  is a probability and  $\phi(\cdot)$  is bounded. We can assume without any loss of generality that the measures  $\mathcal{L}_b^o$  and  $\mathcal{L}_b^d$  have same support. If both distributions do not have the same support, a common support can be defined as the union of individual supports, where no mass is placed on points outside of the support of individual measures. Let  $B$  be any subset of this common support, we have,

$$\begin{aligned} (1 - p_d)\mathcal{L}_b^o(B) &= \text{pr}(D > 2, b \in B) \\ &= \int_{b \in B} \text{pr}(D > 2|b, z(t_2)) d\mathcal{L}_b(b) \\ &= \int_{b \in B} \text{pr}(D > 2|b, z(t_2)) \{(1 - p_d)d\mathcal{L}_b^o(b) + p_d d\mathcal{L}_b^d(b)\}, \end{aligned}$$

where the last equality is due to (4). We therefore have,

$$(1 - p_d) \int_B (1 - \text{pr}(D > 2|b, z(t_2))) d\mathcal{L}_b^o(b) = p_d \int_B \text{pr}(D > 2|b, z(t_2)) d\mathcal{L}_b^d(b), \text{ for any } B.$$

We then have,

$$\frac{d\mathcal{L}_b^d(b)}{d\mathcal{L}_b^o(b)} = \frac{1 - p_d}{p_d} \frac{1 - \text{pr}(D > 2|b, z(t_2))}{\text{pr}(D > 2|b, z(t_2))} \tag{8}$$

Note that under the assumption that no subject drops out at the first timepoint, we have  $\text{pr}(D \geq 2) = 1$  as  $q = 2$ . This then gives  $\text{pr}(D > 2|b, z(t_2)) = 1 - \text{pr}(D = 2|D \geq 2, b, z(t_2)) = h(t_2)$  from the definition of the hazard function. From the hazard model in 2, we have,

$$\frac{1 - \text{pr}(D > 2|D \geq 2, b, z(t_2))}{\text{pr}(D > 2|D \geq 2, b, z(t_2))} = \exp\{-\varepsilon(z(t_2))\alpha - \delta\phi(b)\}.$$

Replacing this expression in equation 8 and a basic algebra give,

$$d\mathcal{L}_b^d(b) = \frac{1 - p_d}{p_d} \exp\{-\varepsilon(z(t_2))\alpha - \delta\phi(b)\} d\mathcal{L}_b^o(b).$$

Let us denote by  $\alpha_j, j = 0, \dots, r - 1$  the slope coefficient associated with the  $j$ th component  $\varepsilon(z_i(t_2))_j$  of the covariate vector  $\varepsilon(z_i(t_2))$  for subject  $i$  at timepoint  $t$ . Here  $z_i(t_2)_0 = 1$  represents the covariate for the intercept term  $\alpha_0$  in the hazard model. Assume that the function  $\varepsilon(\cdot)$  centers all covariates  $z$ , that is  $\sum_i \varepsilon(z_i(t_2))_j = 0$ , for  $j = 1, \dots, r - 1$ . Because  $\mathcal{L}_b^d$  is a law, we have  $\int d\mathcal{L}_b^d(b) = 1$ , and then,

$$\exp\left\{\alpha_0 + \sum_{j=1}^{r-1} \varepsilon(z_i(t_2))_j \alpha_j\right\} = \frac{1 - p_d}{p_d} \int \exp\{-\delta\phi(b)\} d\mathcal{L}_b^o(b).$$



By multiplying the above expression over all subjects, we then get,

$$\exp \left\{ n\alpha_0 + \sum_{j=1}^{r-1} \left\{ \sum_{i=1}^n \varepsilon(z_i(t_2))_j \right\} \alpha_j \right\} = \left\{ \frac{1-p_d}{p_d} \int \exp\{-\delta\phi(b)\} d\mathcal{L}_b^o(b) \right\}^n.$$

Finally, given that  $\sum_i \varepsilon(z_i(t_2))_j = 0$ , for  $j = 1, \dots, r - 1$  the expression in (5) holds. Hence, for a fixed  $\delta$ , the equation in (5) can be used to compute  $\alpha_0$ , which we denote  $\alpha_0(\delta)$  to highlight this dependence. In particular, for  $\delta = 0$  the intercept term  $\alpha_0$  takes the value  $\alpha_0 = \log\{(1 - p_d)/p_d\}$ , when  $q = 2$ .

### Appendix C

Proof of Theorem 2

We recall that the probability of dropping out is given by,

$$f_D(D_i|b_i, z_i(t), \delta) = \begin{cases} h_i(t_2) & \text{if } D_i=2 \\ \left[ \prod_{t \in \mathcal{O}_i} (1 - h_i(t)) \right] \{h_i(t_{D_i})\}^{I(D_i < 1+q)} & \text{if } D_i=3, \dots, q+1. \end{cases}$$

It is clear that when  $\delta$  gets large and  $\sup_{i,t} |\varepsilon(z_i(t))\alpha|$  is bounded, the dropout hazard  $h_i(t) = \{1 + \exp(-\varepsilon(z_i(t))\alpha - \delta\phi(b_i))\}^{-1}$  converges to 1 as  $\phi(b_i) > 0$ . Hence,  $f_D(D_i|b_i, z_i(t), \delta)$  converges to 0 when  $2 < D_i \leq q + 1$  and to 1 only when  $D_i = 2$ .

Let denote by  $I_i^d$  and  $I_j^o$ , respectively, the contribution of noncompleter  $i$  and a completer  $j$  to the likelihood. This then gives  $\log l(\psi, \delta) = \sum_{i=1}^{n_d} \log I_i^d(\psi, \delta) + \sum_{i=n_d+1}^n \log I_i^o(\psi, \delta)$ , where noncompleters appear first in the dataset after a proper rearrangement. Let  $n_d^{(2)}$  be the number of subjects who dropped out right after the baseline, that is  $n_d^{(2)} = \sum_i I(D_i=2)$ . After a proper rearrangement of the subgroup of noncompleters in the dataset, we have for dropout subjects,

$$\sum_{i=1}^{n_d} \log I_i^d(\psi, \delta) = \sum_{i=1}^{n_d^{(2)}} \log I_{i,D_i=2}^d(\psi|\delta) + \sum_{i=n_d^{(2)}+1}^{n_d} \log I_{i,2 < D_i}^d(\psi|\delta),$$

where  $I_{i,D_i=2}^d(\psi, \delta)$  and  $I_{i,2 < D_i}^d(\psi|\delta)$  are respectively the likelihood contribution of a subject who drops out at the second time point and after the second time point. It is clear that when  $\delta \rightarrow \infty$ , we have  $\sum_{i=n_d^{(2)}+1}^{n_d} \log I_{i,2 < D_i}^d(\psi|\delta) \rightarrow -\infty$  and,

$$\sum_{i=1}^{n_d^{(2)}} \log I_{i,D_i=2}^d(\psi, \delta) \rightarrow \sum_{i=1}^{n_d^{(2)}} \log \left\{ \int \prod_{b_i \in \mathcal{O}_i} f_Y(Y_i(t)|b_i, x_i(t)) d\mathcal{L}^o(b_i) \right\} < 0 \text{ uniformly.}$$

Then, the marginal log-likelihood for dropout subjects  $\sum_{i=1}^{n_d} I_i^d(\psi, \delta) \rightarrow -\infty$ , when  $\delta \rightarrow \infty$ . Similarly for subjects with complete data, the marginal log-likelihood

$\sum_{i=n_d+1}^{n-n_d} \log l_i^o(\psi, \delta) \rightarrow -\infty$ , when  $\delta \rightarrow \infty$ . Hence, for large values of  $\delta$ , the log of the marginal likelihood given by,

$$\sum_{i=1}^{n_d} \log l_i^d(\psi, \delta) + \sum_{i=n_d+1}^n \log l_i^o(\psi, \delta) \rightarrow -\infty$$

It can also be shown that the expectation of the complete log-likelihood given observed data and the current parameter estimate  $\psi^{(a)}(\delta)$  for fixed  $\delta$  also diverges.

## Appendix D

*Uniform consistency and weak convergence of  $\hat{\psi}(\delta)$ ,  $0 \leq \delta \leq \Delta$*

Let  $\Psi$  denote the parameter space for  $\psi$ . We denote by  $l(\psi, \delta, W_i)$  the contribution of subject  $i$  to the log-likelihood function. Define

$$s(\psi, \delta, W_i) = \partial l(\psi, \delta, W_i) / \partial \psi, S(\psi, \delta) = n^{-1} \sum_{i=1}^n s(\psi, \delta, W_i),$$

$$c(\psi, \delta) = E\{s(\psi, \delta, W_1)\}, g(\psi, \delta, W_i) = \partial s(\psi, \delta, W_i) / \partial \psi^T,$$

$D_\psi(\psi, \delta) = n^{-1} \sum_{i=1}^n g(\psi, \delta, W_i)$ , and  $\tilde{D}_\psi(\psi, \delta) = E\{g(\psi, \delta, W_1)\}$ . For any given  $\delta \leq \Delta$ , let  $\hat{\psi}(\delta)$  denote the solution to  $S(\psi, \delta) = 0$ , i.e.  $S(\hat{\psi}(\delta), \delta) = 0$ , and define  $\psi^*(\delta) = \operatorname{argmax}_{\psi \in \Psi} E\{l(\psi, \delta, W_1)\}$ . Define  $\mathcal{G}_1 = \{s(\psi, \delta, W_i) : \psi \in \Psi, \delta \leq \Delta\}$  and  $\mathcal{G}_2 = \{g(\psi, \delta, W_i) : \psi \in \Psi, \delta \leq \Delta\}$ .

Assume that  $\Psi$  is compact and  $\psi^*(\delta)$  is an interior point in  $\Psi$  for any  $\delta \leq \Delta$ . In addition, assume the following regularity conditions:

- C1.** The function classes,  $\mathcal{G}_1$  and  $\mathcal{G}_2$ , are pointwise measurable and satisfy the uniform entropy condition; see van der Vaart and Wellner (2000) for the definitions. For example, functions which are uniformly bounded and uniformly Lipschitz of order  $> \{\dim(\psi) + \dim(\delta)\} / 2$  suffice the above conditions, where  $\dim(\cdot)$  denotes the dimension of a vector.
- C2.**  $\inf_{\psi \in \Psi, \delta \leq \Delta} \operatorname{eigmin}\{-\tilde{D}_\psi(\psi, \delta)\} > 0$ , where  $\operatorname{eigmin}(\cdot)$  denotes the minimum eigenvalue of a matrix.

Condition C1 implies that  $\mathcal{G}_1$  and  $\mathcal{G}_2$  are Donsker and hence Glivenko-Cantelli (van der Vaart and Wellner, 2000). Therefore,

$$\sup_{\psi \in \Psi, \delta \leq \Delta} \|S(\psi, \delta) - c(\psi, \delta)\| \rightarrow_p 0 \quad \text{and} \quad \sup_{\psi \in \Psi, \delta \leq \Delta} \|D_\psi(\psi, \delta) - \tilde{D}_\psi(\psi, \delta)\| \rightarrow_p 0, \tag{A1}$$

where  $\|\cdot\|$  denotes the Euclidean norm. The definitions of  $\hat{\psi}(\delta)$  and  $\psi^*(\delta)$  imply that

$$0 = S(\hat{\psi}(\delta), \delta) - c(\psi^*(\delta), \delta) = [S(\hat{\psi}(\delta), \delta) - S(\psi^*(\delta), \delta)] + [S(\psi^*(\delta), \delta) - c(\psi^*(\delta), \delta)]$$

$$= \tilde{D}_\psi(\check{\psi}(\delta), \delta) \cdot \{\hat{\psi}(\delta) - \psi^*(\delta)\} + v_{2,n}(\delta), \tag{A2}$$

where  $\check{\psi}(\delta)$  is on the line segment between  $\psi^*(\delta)$  and  $\hat{\psi}(\delta)$ , and  $\sup_{\delta \leq \Delta} \|v_{2,n}(\delta)\| \rightarrow_p 0$ . By C2, there exists a positive number  $\kappa_0$  not depending on  $\delta$  such that

$$\|\widehat{\psi}(\delta) - \psi^*(\delta)\| \leq \kappa_0^{-1} \|v_{2,n}(\delta)\|.$$

The uniform consistency of  $\widehat{\psi}(\delta)$  to  $\psi^*(\delta)$  follows from  $\sup_{\delta \leq \Delta} \|v_{2,n}(\delta)\| \rightarrow_P 0$ .

Based on the uniform consistency of  $\widehat{\psi}(\delta)$  and (A1–2), applying the Taylor expansion to  $S(\widehat{\psi}(\delta), \delta)$  around  $\{\psi^*(\delta), \delta\}$  gives that

$$\begin{aligned} n^{1/2}\{\widehat{\psi}(\delta) - \psi^*(\delta)\} &\approx -n^{-1/2} \sum_{i=1}^n \tilde{D}_{\psi}^{-1}(\psi^*(\delta), \delta) \{s(\psi^*(\delta), \delta, W_i) - E\{s(\psi^*(\delta), \delta, W_i)\}\} \\ &\equiv -n^{-1/2} \sum_{i=1}^n \iota_i(\delta), \end{aligned} \tag{A3}$$

where  $\approx$  denotes asymptotic equivalence uniformly in  $\delta \leq \Delta$ .

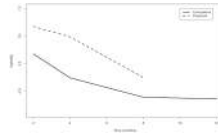
Because C1 implies that  $\mathcal{G}_1$  is Donsker and because C2 implies that  $\tilde{D}_{\psi}^{-1}(\psi^*(\delta), \delta)$  is uniformly bounded for  $\delta \leq \Delta$ , the function class  $\{\tilde{D}_{\psi}^{-1}(\psi^*(\delta), \delta) s(\psi^*(\delta), \delta, W_i), \delta \leq \Delta\}$  is Donsker. This permits the application of a functional central limit theory to establish the weak convergence of  $\widehat{\psi}(\delta)$ .

These general results for any  $\delta \leq \Delta$  also establishes asymptotic results of  $\widehat{\psi}(\delta_0)$ , where  $\delta = \delta_0$ .

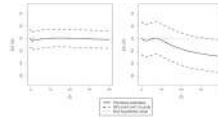
## References

- Albert PS, Follmann DA. Modeling repeated count data subject to informative dropout. *Biometrics* 2000;56:667–677. [PubMed: 10985201]
- Beunckens C, MGVG, Mallinckrodt C. A latent-class mixture model for incomplete longitudinal gaussian data. *Biometrics* 2008;64:96105.
- Burton SW. A review of the fluvoxamine and its uses in depression. *International Clinical Psychopharmacology* 1991;6 Supplement 3:1–17. [PubMed: 1806631]
- Casella, G.; Berger, RL. *Statistical inference*. Pacific Grove, CA: Wadsworth & Brooks/Cole; 1990.
- Copas J, Eguchi S. Local sensitivity approximations for selectivity bias. *Journal of the Royal Statistical Society, Series B: Methodological* 2001;63:871–895.
- Davidian, M.; Giltinan, DM. *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall Ltd; 1998.
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm (C/R: P22-37). *Journal of the Royal Statistical Society, Series B: Methodological* 1977;39:1–22.
- Diggle P, Kenward MG. Informative dropout in longitudinal data analysis (with discussion). *Applied Statistics* 1994;43:49–93.
- Efron, B.; Tibshirani, R. *An introduction to the bootstrap*. London: Chapman and Hall; 1993.
- Freedman DA. Adjusting for nonignorable drop-out using semiparametric nonresponse models: Comment. *Journal of the American Statistical Association* 1999;94:1121–1122.
- Hogan J, Roy J, Korkontzelou C. Biostatistics tutorial: handling dropout in longitudinal studies. *Statistics in Medicine* 2004;23:1455–1497. [PubMed: 15116353]
- Robins JM, AR, Zhao L. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* 1995;90:106–121.
- Kenward MG, Goetghebeur EJT, Molenberghs G. Sensitivity for incomplete categorical data. *Statistical Modelling* 2001;1:31–48.

- Kenward MG, Lesaffre E, Molenberghs G. An application of maximum likelihood and generalized estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random. *Biometrics* 1994;50:945–953. [PubMed: 7787007]
- Laird N, Ware J. Random-effects models for longitudinal data. *Biometrics* 1982;38:963–974. [PubMed: 7168798]
- Lesaffre E, Molenberghs G, Dewulf L. Effects of dropouts in a longitudinal study: An application of a repeated ordinal model. *Statistics in Medicine* 1996;15:1123–1141. [PubMed: 8804143]
- Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73:13–22.
- Lindsay BG. The geometry of mixture likelihoods: A general theory. *The Annals of Statistics* 1983;11:86–94.
- Little, R.; Rubin, D. *Statistical analysis with missing data*. New York: John Wiley and Sons; 1987.
- McLachlan, G.; Peel, D. *Finite Mixture Models*. New York: Wiley; 2000.
- Molenberghs G, Kenward MG, Lesaffre E. The analysis of longitudinal ordinal data with non-random dropout. *Biometrika* 1997;84:33–44.
- Molenberghs G, Lesaffre E. Marginal modeling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association* 1994;89:633–644.
- Rosenbaum, PR. *Observational Studies*. New York: Springer; 2002.
- Rotnitzky A, Scharfstein D, Su T-L, Robins J. Methods for conducting sensitivity analysis of trials with potentially nonignorable competing causes of censoring. *Biometrics* 2001;57:103–113. [PubMed: 11252584]
- Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for nonignorable drop-out using semiparametric nonresponse models (c/r: P1121-1146). *Journal of the American Statistical Association* 1999;94:1096–1120.
- Tao H, Palta M, Yandell B, Newton M. An estimation method for the semiparametric mixed effects model. *Biometrics* 1999;55:102–110. [PubMed: 11318143]
- Ten Have TR, Reboussin BA, Miller ME, Kunselman A. Mixed effects logistic regression models for multiple longitudinal binary functional limitation responses with informative drop-out and confounding by baseline outcomes. *Biometrics* 2002;58:137–144. [PubMed: 11890309]
- Todem D, Kim K, Lesaffre E. A sensitivity approach to modeling longitudinal bivariate ordered data subject to informative dropouts. *Health services and outcomes research methodology* 2006;6:37–57.
- Troxel A, Ma G, Heitjan D. An index of local sensitivity to nonignorability. *Statistica Sinica* 2002;14:1221–1237.
- Vach W, Blettner M. Logistic regression with incompletely observed categorical covariates – Investigating the sensitivity against violation of the missing at random assumption. *Statistics in Medicine* 1995;14:1315–1329. [PubMed: 7569490]
- van der Vaart, AW.; Wellner, JA. Preservation theorems for Glivenko-Cantelli and uniform Glivenko-Cantelli theorems. *High Dimensional Probability II*. Giné, E.; Mason, DM.; Wellner, JA., editors. Birkhäuser, Boston: 2000a.
- van der Vaart, AW.; Wellner, JA. *Weak convergence and empirical processes*. New York: Springer; 2000b.
- Verbeke, G.; Molenberghs, G. *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag; 2000.
- Verbeke G, Molenberghs G, Thijs H, Lesaffre E, Kenward M. Sensitivity analysis for nonrandom dropout: A local influence approach. *Biometrics* 2001;57:7–14. [PubMed: 11252620]
- Wu MC, Carroll RJ. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* 1988;44:175–188.

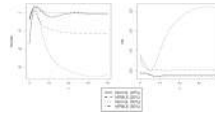


**Fig. 1.**  
Log odds of side effects across time for completers and dropouts

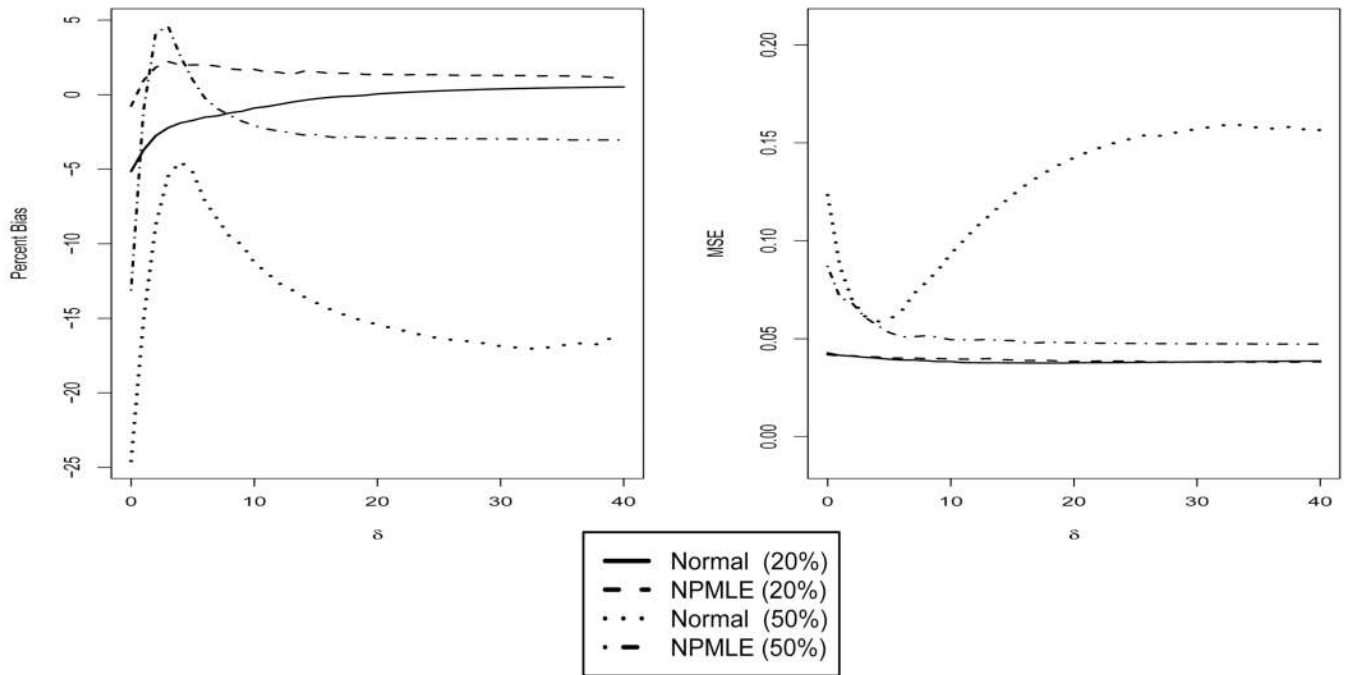


**Fig. 2.** Estimates and simultaneous 95% confidence intervals of  $\beta_1^*(\delta) - \beta_1^*(0)$  across values of the sensitivity parameter in the range  $0 \leq \delta \leq 40$  assuming (a) a semiparametric model and (b) a parametric model

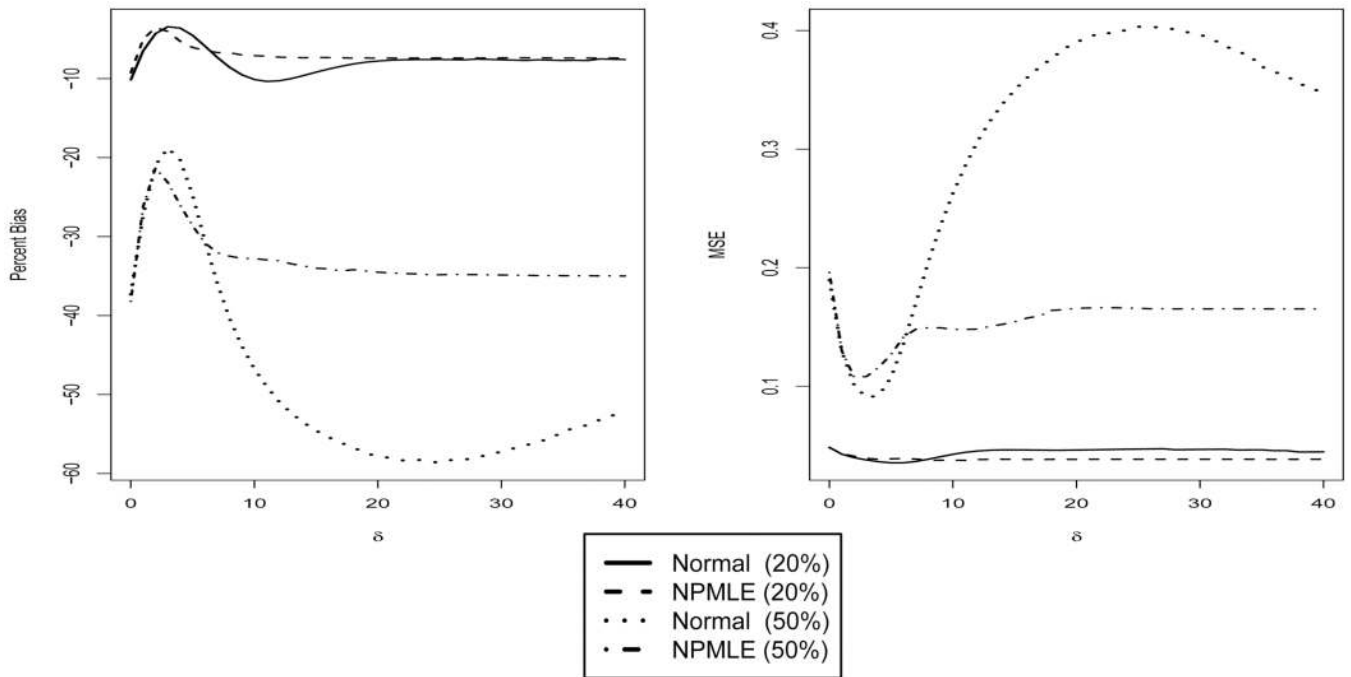




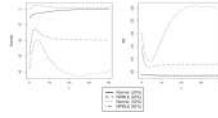
**Fig. 3.** Percent bias and MSEs of parameter  $\beta_1$  as a function of the sensitivity parameter  $\delta$  with true measurement model  $\mu_i(t) = \{1 + \exp(-1 + x_i(t) - b_i)\}^{-1}$ , true dropout process  $h_i(t) = \{1 + \exp(-\alpha_0 - \phi(b_i))\}^{-1}$ , with  $\alpha_0 = 2.094$  (20% dropouts) and  $\alpha_0 = .855$  (50% dropouts) and true shared random effects  $\mathcal{L}_b \sim \mathcal{N}(0, 4.025)$



**Fig. 4.** Percent bias and MSEs of parameter  $\beta_1$  as a function of the sensitivity parameter  $\delta$  with true measurement model  $\mu_i(t) = \{1 + \exp(-1 + x_i(t) - b_i)\}^{-1}$ , true dropout process  $h_i(t) = \{1 + \exp(-\alpha_0 - \phi(b_i))\}^{-1}$ , with  $\alpha_0 = 2.094$  (20% dropouts) and  $\alpha_0 = .855$  (50% dropouts) and true shared bimodal random effects  $\mathcal{L}_b \sim \frac{1}{2} \mathcal{N}(-2, .025) + \frac{1}{2} \mathcal{N}(2, .025)$



**Fig. 5.** Percent bias and MSEs of parameter  $\beta_1$  as a function of the sensitivity parameter  $\delta$  with true measurement model  $\mu_i(t) = \{1 + \exp(-1 + x_i(t) - b_i)\}^{-1}$ , true dropout process  $h_i(t) = \{1 + \exp(-\alpha_0 - Y_i(t))\}^{-1}$ , with  $\alpha_0 = 2.094$  (20% dropouts) and  $\alpha_0 = .855$  (50% dropouts) and true unshared random effects  $\mathcal{L}_b \sim \mathcal{N}(0, 4.025)$



**Fig. 6.** Percent bias and MSEs of parameter  $\beta_1$  as a function of the sensitivity parameter  $\delta$  with true measurement model  $\mu_i(t) = \{1 + \exp(-1 + x_i(t) - b_i)\}^{-1}$ , true dropout process  $h_i(t) = \{1 + \exp(-\alpha_0 - Y_i(t))\}^{-1}$ , with  $\alpha_0 = 2.094$  (20% dropouts) and  $\alpha_0 = .855$  (50% dropouts) and true unshared bimodal random effects  $\mathcal{L}_b \sim \frac{1}{2}\mathcal{N}(-2, .025) + \frac{1}{2}\mathcal{N}(2, .025)$