Bayesian Sensitivity Analysis of Incomplete Data Using Pattern-Mixture and Selection Models Through Equivalent Parameterization

Niko Kaciroti[*]   Trivellore E. Raghunathan[†]

Abstract

Pattern-mixture models (PMM) and selection models (SM) are two alternative approaches for statistical analysis with incomplete data and a nonignorable missing-data mechanism. Both models make empirically unverifiable assumptions and need additional constraints to identify the parameters. We introduce Bayesian parameterizations to identify the PMM for different types of outcome within the exponential distribution family and then translate these to their equivalent SM approach. This provides for a unified and robust parameterization that can be used for sensitivity analysis under either approach. The new parameterizations are easy-to-use and have intuitive interpretation from both PMM and SM perspectives. These models can be fitted using software implementing Gibbs sampling.

Key Words: Bayesian parameterization; Missing not at random; Mixture analysis; Ignorability index; Selection bias; Identifiability.

1. Introduction

Missing data is a common problem in statistical modeling both in cross-sectional design and more so in longitudinal design where subjects are often lost to follow-up. A wide range of statistical models for analyzing outcomes with missing data is available, but their validity will often depend on the nature of the missing-data mechanism as well as on the validity of any assumption used.

For analyzing data with missing values that are potentially missing not at random (MNAR) two widely used approaches are pattern-mixture models (PMM) and selection models (SM). Both PMM and SM derives their inferences based on the joint distribution $f(Y, R)$ of the outcome Y and the missing-data indicator R, but use different decomposition for $f(Y, R)$. Selection models partition $f(Y, R)$ as the product of $f(Y)$ and $f(R|Y)$ (Heckman (1974,1979)). They require explicit modeling of the missing-data mechanism where the probability that a subject is missing may depend on the observed and unobserved values. PMM, on the other hand, express the joint distribution as the product of $f(Y|R)$ and $f(R)$ where the data are stratified by the missing-data patterns with distinct parameters for each pattern (Rubin 1977, Little and Rubin 2002). The marginal estimates in PMM can be derived as a weighted average across pattern specific estimates (Little 1995) or by using imputation (Demirtas and Schafer (2003)). SM, on the another hand, estimate the marginal parameters directly. The choice between PMM and SM is based on different reasons, such as the analysis objective, how best to formulate and incorporate the assumption regarding the missing-data mechanism as well as on the flexibility of the sensitivity analysis that is followed. For example, Scharfstein *et al.* (1999, 2003) and Rotnitzky *et al.* (1998, 2001) used a selection model

---

[*]Center for Human Growth and Development, Department of Biostatistics and Bioinformatics. 300 N. Ingalls Bldg, 10-floor, Ann Arbor, MI 48109

[†]Department of Biostatistics, Survey Research Center, ISR. 109 Observatory St., 1700 SPH I Ann Arbor, MI 48109

approach; Little (1993, 1994) and Little and Rubin (2002) used a pattern-mixture design. Regardless of which model is used, assumptions that are not verifiable from the observed data, or additional data, are needed to identify the parameters in the joint distribution. These assumptions rely heavily on expert opinions about plausible ranges for non-identifiable parameters and are usually followed with sensitivity analysis (Troxel *et al.* (2004), Molenberghs and Verbeke (2005), Daniels and Hogan (2007)).

In this article we introduce Bayesian parameterizations to identify the PMM for different types of outcome within the exponential distribution family and then translate these to their equivalent SM approach. This provides for a unified and robust parameterization that can be used for sensitivity analysis under either approach. The new parameterizations are easy-to-use and have intuitive interpretation from both PMM and SM perspectives.

2. THE PROPOSED METHOD

In this section we propose new parameterizations within the pattern-mixture model framework for outcomes from the exponential distribution family. We focus on studies from randomized control trials, where subjects are randomized into two or more treatment groups (identified by Trt indicator) and have potentially MNAR dropouts. Using the generalized liner models approach (McGullagh and Nelder (1989)), the PMM for outcomes within the exponential family is:

$$g(E(Y|Trt, R = r)) = \beta_0^{(r)} + \beta_1^{(r)} Trt \tag{1}$$

where $g(.)$ function is the natural link function, and Trt is a covariate representing a group indicator. We will identify model (1) by relating $f(Y|R = 1)$ to $f(Y|R = 0)$ or by specifying $g(E(Y|Trt, R = 1)) - g(E(Y|Trt, R = 0)) = log(\tilde{\lambda})$ (slightly modified for normal outcomes) where $\tilde{\lambda}$ represents some intuitive Bayesian parameterization, which we refer to as ignorability index. This parameterization accommodates modeling of missing data generated from a potentially nonignorable missing-data mechanism and include the MAR mechanism as a special case. The parameterization is intuitive, easily used for sensitivity analysis and has an interpretation within both pattern-mixture framework and selection model framework. We consider here the most widely used distributions from the exponential family: binary, Poisson, multinomial (nominal and ordinal) and normal distributions.

2.1 Binary Outcomes with Missing Data

Binary outcomes are commonly used to indicate the presence or absence of an event or characteristic. For example, $Y = 1$ if an endpoint is observed (subject has hypertension) or 0 otherwise. When there are missing data we consider pattern-mixture models to model $E(Y)$ as a function of Trt stratified by the missing-data indicator R. Such models are often used when the missing data are missing not at random (MNAR). We assume Trt is a group indicator (0 for a placebo and 1 for treatment). Thus, we consider the following logistic model using pattern-mixture framework to model the effect of Trt on the odds of having an endpoint:

$$logit(p^{(r)}|Trt) = \beta_0^{(r)} + \beta_1^{(r)} Trt.$$

where $p^{(r)} = Pr(Y = 1|R = r) = E(Y|R = r)$. Here $R = 1$ indicates that Y is missing and $R = 0$ indicates Y is observed. The model is underidentified as

there are no data to estimate the parameters for missing-data pattern $R = 1$. We identify such model by using a Bayesian parameterization with informative prior distributions in the identifying parameters $\tilde{\lambda}_k$ (Kaciroti *et. al.* (2009)) where:

$$logit(E(Y|Trt = k, R = 1)) - logit(E(Y|Trt = k, R = 0)) = \log(\tilde{\lambda}_k)$$

or

$$\tilde{\lambda}_k = \frac{p_k^{(1)}/1 - p_k^{(1)}}{p_k^{(0)}/1 - p_k^{(0)}}. \tag{2}$$

Here $\tilde{\lambda}_k$ ($k = 0, 1$ for $Trt = 0, 1$) is a measure of the departure from the MAR and represent the odds ratio of having an endpoint between the missing-data pattern and the observed-data pattern for each group ($k = 0, 1$). The MAR is a special case where $\tilde{\lambda}_k$ equals 1. The $\tilde{\lambda}$ parameter is not estimateable from the data. We assume that $\tilde{\lambda}_k \sim Log - Normal$ with some mean l_k and variance $c^2 l_k^2$, c is the coefficient of variation and represents the degree of uncertainty where $c = 0$ translates to a deterministic constraint. The parameter l_k and c are the sensitivity parameters and are chosen to represent a range of the odds ratio (OR) of having an endpoint between subjects who dropped out and subjects who completed the study. Such parameterization is intuitive and easy to use for sensitivity analysis within a PMM framework. In addition, $\tilde{\lambda}_k$ also has an intuitive interpretation within a SM approach and next we derive the equivalent SM corresponding to the PMM identified by (2). Following Bayesian rule we obtain:

$$\tilde{\lambda}_k = \underbrace{\frac{Pr(Y = 1|R = 1, Trt = k)}{Pr(Y = 0|R = 1, Trt = k)} / \frac{Pr(Y = 1|R = 0, Trt = k)}{Pr(Y = 0|R = 0, Trt = k)}}_{PMM}$$
$$= \underbrace{\frac{Pr(R = 1|Y = 1, Trt = k)}{Pr(R = 0|Y = 1, Trt = k)} / \frac{Pr(R = 1|Y = 0, Trt = k)}{Pr(R = 0|Y = 0, Trt = k)}}_{SM}. \tag{3}$$

Thus, from (3) the sensitivity parameter $\tilde{\lambda}_k$ has an alternative interpretation from a SM perspective. It represents the odds ratio of dropping out between the subjects with endpoint versus subjects without endpoint for group k. In general in a SM approach there are several link functions used to model the missing-data mechanism, e.g. logistic, probit or log-log. However, for parameterization (2) the corresponding SM model is:

$$logit(Pr(R = 1|Y)) = \gamma_0 + \gamma_1 Trt + \gamma_2 Y + \gamma_3 Trt * Y$$

where $log(\tilde{\lambda}_0) = \gamma_2$ and $log(\tilde{\lambda}_1/\tilde{\lambda}_0) = \gamma_3$.

Thus, from relationship (2, 3) the sensitivity parameter $\tilde{\lambda}_k$ has a dual interpretation. In PMM, $\tilde{\lambda}_k$ represents the odds ratio of having an endpoint between the missing subjects and the observed subjects for group k. In SM it represents the odds ratio of dropping out between subjects with endpoint versus subjects without endpoint for group k. Both PMM and SM can be identified by giving a prior distribution to $log(\tilde{\lambda}_k)$. This parameterization and its dual interpretation unifies the sensitivity analysis through PMM and SM, where both models can be fitted interchangeably. This provides flexibility and assurance on the assumptions about the missing data from either perspective. Next we will define similar parameterization to relate PMM and SM that are relevant to other distributions of the outcome variables within the exponential family.

2.2 Poisson Outcomes with Missing Data

In this section we consider count outcome measures. For example, Y is the number of times that a subject with a chronic condition (e.g. asthma) visits the emergency care over a time period. Poisson regression is suitable for fitting such type of outcomes. In the presence of missing data we consider the following pattern-mixture model for Y:

$$log(\mu^{(r)}|Trt, R = r) = \beta_0^{(r)} + \beta_1^{(r)} * Trt$$

where $\mu^{(r)} = E(Y|R = r)$. To identify this model we used the following parameterization (Kaciroti *et al.* (2008)):

$$log(E(Y)|Trt = k, R = 1) - log(E(Y)|Trt = k, R = 0) = \log(\tilde{\lambda}_k)$$

or

$$\tilde{\lambda}_k = \frac{\mu_k^{(1)}}{\mu_k^{(0)}} \tag{4}$$

where $\tilde{\lambda}_k$ is again a measure of the departure from the MAR and represents the risk ratio of having an event between the missing-data pattern and the observed-data pattern for group k. Following Bayesian rule we have:

$$\frac{Pr(R = 1|Y = y + 1, Trt)}{Pr(R = 0|Y = y + 1, Trt = k)} \bigg/ \frac{Pr(R = 1|Y = y, Trt = k)}{Pr(R = 0|Y = y, Trt = k)}$$
$$= \frac{f(Y = y + 1|R = 1, Trt = k)}{f(Y = y|R = 1, Trt = k)} \bigg/ \frac{f(Y = y + 1|R = 0, Trt = k)}{f(Y = y|R = 0, Trt = k)}$$

where $f(y|R = r, Trt = k) = \frac{e^{-\mu_k^{(r)}} \mu_k^{(r)y}}{y!}$ is the Poisson probability density function, from which we obtain

$$\frac{Pr(R = 1|Y = y + 1, Trt = k)}{Pr(R = 0|Y = y, Trt = k)} \bigg/ \frac{Pr(R = 1|Y = y + 1, Trt = k)}{Pr(R = 0|Y = y, Trt = k)} = \frac{\mu_k^{(1)}}{y + 1} \bigg/ \frac{\mu_k^{(0)}}{y + 1} = \frac{\mu_k^{(1)}}{\mu_k^{(0)}}$$

or

$$\tilde{\lambda}_k = \underbrace{\frac{Pr(R = 1|Y = y + 1, Trt = k)}{Pr(R = 0|Y = y, Trt = k)} \bigg/ \frac{Pr(R = 1|Y = y + 1, Trt = k)}{Pr(R = 0|Y = y, Trt = k)}}_{SM} = \underbrace{\frac{\mu_k^{(1)}}{\mu_k^{(0)}}}_{PMM} \tag{5}$$

Then, from (5) the sensitivity parameter $\tilde{\lambda}_k$ has an alternative interpretation from a SM perspective. It represents the odds ratio of dropping out for one unit increase in y for group k. Similarly as in the Binomial case the corresponding SM model for PMM with parameterization (4) is:

$$logit(Pr(R = 1)|Y) = \gamma_0 + \gamma_1 Trt + \gamma_2 Y + \gamma_3 Trt * Y$$

where $log(\tilde{\lambda}_0) = \gamma_2$ and $log(\tilde{\lambda}_1/\tilde{\lambda}_0) = \gamma_3$.

Thus, from relationship (4, 5) the sensitivity parameter $\tilde{\lambda}_k$ has a dual interpretation. In PMM, $\tilde{\lambda}_k$ represents the risk ratio of having an event between the missing subjects and the observed subjects for group k. In SM it represents the odds ratio of dropping for one unit increase in y for group k. Both PMM and SM can be identified by giving a prior distribution to $log(\tilde{\lambda}_k)$. This parameterization and its dual interpretation unifies the sensitivity analysis through PMM and SM.

2.3 Multinomial Outcomes with Missing Data

Multinomial outcome variables are used to measure different categories. They can be in an ordinal scale where categories are ordered or in a nominal scale where no order among categories exists.

2.3.1 Nominal Outcomes

Nominal measures are often used to indicate group membership. For example Y indicates political affiliation, "1=Democrat", "2=Republican" or "3=Independent" or type of disease "1=None", "2=Cancer", "3=Cardiovascular" or "4=Other". Multinomial logistic regression modeling the odds of being in one category versus a reference category is commonly used for analyzing this type of measures (Agresti (2002)). In the presence of dropouts we consider the following pattern-mixture model for nominal outcome:

$$log(\frac{p_j^{(r)}}{p_1^{(r)}}|Trt) = \alpha_j^{(r)} + \beta_j^{(r)}Trt$$

where $p_j^{(r)} = Pr(Y = j|R = r)$ for $j = 1, ..., J$ with $j = 1$ being the reference group. Similarly, as in previous cases, we identify the model by relating the main parameters of interest between the missing-data and the complete-data patterns:

$$log(p_{kj}^{(1)}/p_{k1}^{(1)}|Trt = k, R = 1) - log(p_{kj}^{(0)}/p_{k1}^{(0)}|Trt = k, R = 0) = \log(\tilde{\lambda}_{kj})$$

or

$$\tilde{\lambda}_{kj} = \frac{p_{kj}^{(1)}/p_{k1}^{(1)}}{p_{kj}^{(0)}/p_{k1}^{(0)}}. \qquad (6)$$

Here $\tilde{\lambda}_{kj}$ is the odds ratio of being in category j versus the reference category 1 between the subjects who dropped out and subjects who completed the study for group k ($k = 0, 1$). Following Bayesian rule we relate the proposed PMM parameterization to a SM.

$$\tilde{\lambda}_{kj} = \underbrace{\frac{Pr(Y = j|Trt = k, R = 1)}{Pr(Y = 1|Trt = k, R = 1)} / \frac{Pr(Y = j|Trt = k, R = 0)}{Pr(Y = 1|trt = k, R = 0)}}_{PMM}$$

$$= \underbrace{\frac{Pr(R = 1|Trt = k, Y = j)}{Pr(R = 0|Trt = k, Y = j)} / \frac{Pr(R = 1|Trt = k, Y = 1)}{Pr(R = 0|Trt = k, Y = 1)}}_{SM} \qquad (7)$$

The corresponding SM here, as an extension of the binary case is:

$$logit(Pr(R = 1|Trt = k, Y)) = \gamma_1 + \gamma_2 I_{(Y=2)} + ... + \gamma_J I_{(Y=J)}$$

$$+\delta_2 Trt * I_{(Y=2)} + ... + \delta_J Trt * I_{(Y=J)}$$

using $J-1$ dummy variables for categories $2, ..., J$ as predictors with category $Y = 1$ being the reference group with $\gamma_j = \log(\tilde{\lambda}_{0j})$ and $\delta_j = \log(\tilde{\lambda}_{1j}/\tilde{\lambda}_{0j})$. Thus,

$$\tilde{\lambda}_{0j} = e^{\gamma_j} = \frac{p_{0j}^{(1)}/p_{01}^{(1)}}{p_{0j}^{(0)}/p_{01}^{(0)}}$$

and

$$\tilde{\lambda}_{1j} = e^{\gamma_j + \delta_j} = \frac{p_{1j}^{(1)}/p_{11}^{(1)}}{p_{1j}^{(0)}/p_{11}^{(0)}}$$

from where $\tilde{\lambda}_{kj}$ is the odds ratio of having a missing value between subjects in group $Y = j$ versus subjects in reference group $Y = 1$ for group k. Under PMM approach it is the odds ratio of being in group j compared to being in group 1 between the missing subjects and the observed subjects. Both PMM and SM are identified by giving a prior distribution on $\tilde{\lambda}_{kj}$ and can be fitted interchangeably.

2.3.2 Ordinal Outcomes

Ordinal outcomes are a special case of the multinomial case, where the categories are ordered. For example Y measures the quality of life "1=Poor", "2=Average" and "3=Good" which are ordered with higher scores indicating a better quality of life. For such outcomes the model used for the nominal case still applies. Thus, we consider the same pattern-mixture model for ordinal outcomes with dropout:

$$\log(\frac{p_j^{(r)}}{p_1^{(r)}}|Trt, R = r) = \alpha_j^{(r)} + \beta_j^{(r)}Trt.$$

Taking advantage of the order structure, a simplified model is often fitted for SM. A special case is the SM based on a logistic regression linear in Y:

$$logit(Pr(R = 1|Y, Trt)) = \gamma_0 + \gamma_1 Trt + \gamma_2 Y + \gamma_3 Trt * Y \qquad (8)$$

with $\gamma_2 = \log(\tilde{\lambda}_0)$ and $\gamma_3 = \log(\tilde{\lambda}_1/\tilde{\lambda}_0)$. Based on SM (8) $\tilde{\lambda}_0 = e^{\gamma_2}$ and $\tilde{\lambda}_1 = e^{\gamma_2+\gamma_3}$ represent the OR of dropping out for one unit increase in Y, for group k=0,1. Then from (7)

$$\tilde{\lambda}_{kj} = \frac{p_{kj}^{(1)}/p_{k1}^{(1)}}{p_{kj}^{(0)}/p_{k1}^{(0)}} = e^{\gamma_2+\gamma_3*k(j-1)} = \tilde{\lambda}_k^{j-1}.$$

Thus, following the SM (8), $\tilde{\lambda}_k$ for the corresponding PMM represents the odds ratio of being in category 2 versus category 1 (or category $j+1$ versus category j) between the subjects who dropped out and those who completed the study ($k = 0, 1$).

Another commonly used model for ordinal outcomes is based on cumulative logistic regression. The PMM for such model follows:

$$logit(q_l^{(r)}|Trt, R = r) = \beta_{0l}^{(r)} + \beta_{1l}^{(r)}Trt$$

where $q_{kl}^{(r)} = Pr(Y \leq l|Trt = k, R = r)$. We have identified such model (Kaciroti et al. (2006)) following:

$$\frac{q_{kl}^{(1)}/(1 - q_{kl}^{(1)})}{q_{kl}^{(0)}/(1 - q_{kl}^{(0)})} = \tilde{\lambda}_{kl}$$

from where

$$\frac{Pr(R=1|Trt=k, Y \le l)/(1 - Pr(R=1|Trt=k, Y \le l))}{Pr(R=1|Trt=k, Y > l)/(1 - Pr(R=1|trt=k, Y > l))} = \tilde{\lambda}_{kl}$$

Thus, under PMM, $\tilde{\lambda}_{kl}$ represents the ratio of cumulative odds of having a low value on Y between the observed subjects and dropouts for $k = 0, 1$. Under the SM approach it represents the OR of dropping out between subjects with scores lower or equal than l versus the ones with scores higher than l. When the number of categories is small ($J = 3$) the model can be simplified by assuming $\tilde{\lambda}_{kj} = \tilde{\lambda}_k$.

2.4 Normal Outcomes with Missing Data

Extensive work has been done on analyzing normal outcomes with missing data using both PMM and SM. Little (1994) and Little and Wang (1996) uses PMM to analyze normal outcomes with potentially nonignorable missing-data mechanism and use restrictions to identify the model followed by sensitivity analysis. Daniels and Hogan (2000) uses similar approach but they identify the PMM following Molenberghs *et al.* (1998) framework by relating the distribution of the missing data to that of the complete data. Similarly, here we identify the PMM model by relating the distribution of the missing data to that of the observed data using a Bayesian parameterization, which is intuitive, easy-to-use and with dual interpretation from PMM and SM perspective.

We consider the following pattern-mixture model for the normal outcome variable Y with dropout:

$$E(Y|Trt, R = r) = \beta_0^{(r)} + \beta_1^{(r)} Trt$$

First we assume that distribution of Y, $f(Y|Trt = k, R = r)$, is normal $\mathcal{N}(\mu_k^{(r)}, \sigma^2)$ where the variance of Y for each missing-data pattern and group $k = 0, 1$ is the same, $\sigma_k^{(1)} = \sigma_k^{(0)} = \sigma$. The natural link function for normal outcomes is identity, thus to identify the model we relate the location parameters (means) between the missing subjects and the observed subjects as follows:

$$\mu_k^{(1)} - \mu_k^{(0)} = \log(\tilde{\lambda}_k) \tag{9}$$

where $\tilde{\lambda}_k$ is again a measure of the departure from the MAR and represents the mean difference in exponential scale between the missing-data pattern and the observed-data pattern for group k. Following Bayesian rule we have:

$$\frac{Pr(R=1|Y=y+\sigma^2, Trt=k)}{Pr(R=0|Y=y+\sigma^2, Trt=k)} \bigg/ \frac{Pr(R=1|Y=y, Trt=k)}{Pr(R=0|Y=y, Trt=k)}$$

$$= \frac{f(Y=y+\sigma^2|R=1, Trt=k)}{f(Y=y|R=1, Trt=k)} \bigg/ \frac{f(Y=y+\sigma^2|R=0, Trt=k)}{f(Y=y|R=0, Trt=k)}$$

here $f(y|R = r, Trt = k) = (2\pi\sigma^2)^{-1/2} e^{-\frac{1}{2\sigma^2}(y-\mu_k^{(r)})^2}$ is the normal probability density function, from which we obtain

$$\frac{f(Y=y+\sigma^2|R=1, Trt=k)}{f(Y=y|R=1, Trt=k)} \bigg/ \frac{f(Y=y+\sigma^2|R=0, Trt=k)}{f(Y=y|R=0, Trt=k)} = e^{\mu_k^{(1)} - \mu_k^{(0)}}$$

or

$$\tilde{\lambda}_k = \frac{Pr(R=1|Y=y+\sigma^2, Trt=k)}{1 - Pr(R=1|Y=y, Trt=k)} \bigg/ \frac{Pr(R=1|Y=y+\sigma^2, Trt=k)}{1 - Pr(R=1|Y=y, Trt=k)}. \tag{10}$$

Then, from (10) the sensitivity parameter $\tilde{\lambda}_k$ has an alternative interpretation from a SM perspective. It represents the odds ratio of dropping out for σ^2 unit increase in y for group k. The corresponding SM model for PMM with parameterization (9) is:

$$logit(Pr(R = 1|Y, Trt)) = \gamma_0 + \gamma_1 Trt + \gamma_2 Y + \gamma_3 Trt * Y$$

where $\gamma_2 = \log(\tilde{\lambda}_0)/\sigma^2 = \frac{\mu_0^{(1)} - \mu_0^{(0)}}{\sigma^2}$ and $\gamma_3 = log(\tilde{\lambda}_1/\tilde{\lambda}_0)/\sigma^2 = \frac{\mu_1^{(1)} - \mu_1^{(0)}}{\sigma^2} - \frac{\mu_0^{(1)} - \mu_0^{(0)}}{\sigma^2}$.

Thus, from (9) and (10) the sensitivity parameter $\tilde{\lambda}_k$ has a dual interpretation. In PMM, $\tilde{\lambda}_k$ represents the mean difference in exponential scale between the missing subjects and the observed subjects for group k. In SM it represents the odds ratio of dropping out for one σ^2 increase in y for group k. Both PMM and SM can be identified by giving a prior distribution to $\log(\tilde{\lambda}_k)$. This parameterization and its dual interpretation unifies the sensitivity analysis through PMM and SM.

Next, let us consider the case where $\sigma_k^{(1)} \neq \sigma_k^{(0)}$. Thus, to identify the model we relate the location parameters (means) and the scale parameters between the missing subjects and the observed subjects as follows:

$$\mu_k^{(1)} - \mu_k^{(0)} = \log(\tilde{\lambda}_k) \tag{11}$$

and

$$\sigma_k^{(1)} = \tilde{\psi}_k \sigma_k^{(0)}. \tag{12}$$

The $\tilde{\psi}_k$ is the ratio of σ_k parameters between the missing-data pattern and the observed-data pattern for group k. We assume that it follows a log-normal distribution with mean ψ and variance $c^2\psi^2$. Following Bayesian rule as above we get:

$$\frac{Pr(R = 1|Y = y + 1, Trt = k)}{Pr(R = 0|Y = y + 1, Trt = k)} \Big/ \frac{Pr(R = 1|Y = y, Trt = k)}{Pr(R = 0|Y = y, Trt = k)}$$

$$= \frac{f(Y = y + 1|R = 1, Trt = k)}{f(Y = y|R = 1, Trt = k)} \Big/ \frac{f(Y = y + 1|R = 0, Trt = k)}{f(Y = y|R = 0, Trt = k)}$$

here $f(y|R = r, Trt) = (2\pi\sigma_k^{(r)2})^{-1/2}e^{-\frac{1}{2\sigma_k^{(r)2}}(y-\mu_k^{(r)})^2}$ is the normal probability density function for pattern r. The corresponding SM model for PMM with parameterization (11, 12) is:

$$logit(Pr(R = 1|Y, Trt)) = \gamma_0 + \gamma_1 Trt + \gamma_2(1 - Trt) * Y + \gamma_3 Trt * Y$$

$$+\gamma_4(1 - Trt) * Y^2 + \gamma_5 Trt * Y^2 \tag{13}$$

where $\gamma_{2+k} = \frac{\mu_k^{(1)}}{\sigma_k^{(1)2}} - \frac{\mu_k^{(0)}}{\sigma_k^{(0)2}} = \frac{\log(\tilde{\lambda}_k) + \mu_k^{(0)}(1 - \tilde{\psi}_k^2)}{\sigma_k^{(0)2}\tilde{\psi}_k^2}$ and $\gamma_{4+k} = \frac{1}{2\sigma_k^{(0)2}} - \frac{1}{2\sigma_k^{(1)2}} = \frac{1}{2\sigma_k^{(0)2}}\frac{\tilde{\psi}_k^2 - 1}{\tilde{\psi}_k^2}$. Thus, when the distribution of the missing data has different location and scale parameters compared to that of the observed data the corresponding SM is no longer linear. A quadratic term (with the coefficient expressing the difference of the inverse scale parameters between missing-data patterns) is part of the SM. In this situation, the identifiability of the model using a PMM approach is more meaningful and intuitive than the SM approach.

3. Conclusions

We propose here a Bayesian parameterization using a pattern-mixture model approach to analyze outcomes within the exponential distribution family with missing values that are potentially MNAR. Such parameterization is intuitive and easy-to-use. It measures the difference between the distributions of the missing data from the observed data and is then used for sensitivity analysis. We relate the identifying parameterization proposed within PMM framework to its corresponding parameterization using SM approach. Such equivalence results in a dual interpretation of the identifying parameter that is used for sensitivity analysis. Both PMM and SM represent the joint likelihood of the outcome and the missing-data indicator, but use alternative decomposition of such likelihood. When the parameter spaces of the models for the outcome and the missing-data indicator are independent, both PMM and SM will result in the same statistical inferences and therefore can be fitted interchangeably. The choice of the model is often decided based on the analysis objective, on how best to formulate and incorporate the identifying assumptions regarding the missing-data mechanism, as well as on the flexibility of the analysis that is followed. Here we provide a Bayesian parameterization to identify the model that has a dual interpretation and can be formulated in an intuitive way from an expect in the filed by either a PMM or a SM perspective. This provides flexibility and assurance to a subject expert for eliciting information that is used to construct prior distributions to identify the model. We use an informative prior to capture a range of different scenarios and apply sensitivity analysis to evaluate the robustness of the results towards different sensitivity assumptions. Finally, from a computational point, the proposed parameterization allow PMM and SM to be fitted interchangeably, for example when formulation is easier under one approach but implementation is easier under the other approach.

REFERENCES

Agresti, A. (2002), "Categorical data analysis, 2nd edition ", *New York: John Wiley.*

Amemiya, T. (1984), " Tobit models: a survey", *J. Econometrics*, 24, 3–61.

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

Daniels, M., Hogan, J. (2007), "Missing data in longitudinal studies: Strategies for Bayesian modeling and sensitivity analysis", *Chapman & Hall.*

Demirtas, H., Schafer, J. L. (2003), " On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out", *Statistics in Medicine*, 22, 1113–1121.

Gelfand, A. E., Smith, A. F. M. (1990), "Sampling based approaches to calculate marginal densities", *Journal of the American Statistical Association*, 85, 398–409.

Heckman J. J. (1974), "Shadow wages, market wages and labor supply", *Econometrica*, 42, 679-693.

Heckman, J. J. (1976). "The common structure of statistical models of truncation, sample selection, limited dependent variables and a simple estimator for such models", *Annals of Economic and Social Measurement*, 5, 475–492.

Heckman, J. J. (1979), "Sample selection bias as a specification error", *Econometrica*, 47, 153–162.

Kaciroti, N., Raghunathan, T. E., Schork, M. A., Clark, N. M., Gong, M. (2006), "A Bayesian approach for clustered longitudinal ordinal outcome with nonignorable missing data: Evaluation of an asthma education program", *Journal of the American Statistical Association*, 101, 435–446.

Kaciroti, N., Raghunathan, T. E., Schork, M. A., Clark, N. M. (2008), "A Bayesian model for longitudinal count data with non-ignorable dropout", *Journal of Royal Statistical Society Series C: Applied Statistics*, 57(5), 521–534.

Kaciroti, N., Schork, M. A., Raghunathan, T. E., Julius, S. (2009), "A Bayesian sensitivity model for intention-to-treat analysis on binary outcomes with dropouts", *Statistics in Medicine*, 28, 572–585.

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

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

Little, R. J. A. (1995), "Modeling the dropout mechanism in repeated measures studies", *Journal of the American Statistical Association*, 90, 1113–1121.

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

Little, R. J. A., Rubin, D. B. (2002), "Statistical analysis with missing data, 2nd edition ", *New York: John Wiley.*

McCullagh, P., Nelder, J. A. (1989), "Generalized linear models, 2nd edition", *London: Chapman & Hall.*

Molenberghs, G., Michiels, B., Kenward, M.G., Diggle, P.J. (1998), "Monotone missing data and pattern-mixture models", *Statistica Neerlandica*, 52, 153–161.

Molenberghs, G. and Verbeke, G. (2005) "Models for Discrete Longitudinal Data", *New York: Springer.*

Rotnitzky, A., Scharfstein, D. O., SU, T., Robins, J. M. (2001), " Methods for conducting sensitivity analysis of trials with potentially non-ignorable competing courses of censoring" *Biometrics*, 57, 103–113.

Rotnitzky, A., Robins, J. M., Scharfstein, D. O. (1998), "Semiparametric regression for repeated outcomes with nonignorable nonresponse", *Journal of the American Statistical Association*, 93, 1321–1339.

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

Scharfstein, D. O., Daniels, M. J., Robins, J.M. (2003), "Incorporating prior beliefs about selection bias into the analysis of randomized trials with missing outcomes", *Biostatistics*, 4(4), 495–512.

Scharfstein, D. O., Daniels, M. J., Robins, J.M. (1999), "Adjusting for non-ignorable drop-out using semiparametric nonresponse models" (with discussion), *Journal of the American Statistical Association*, 94, 1096–1146.

Troxel, A. B., Ma, G., Heitjan, D. (2004), "An index to local sensitivity of nonignorability", *Statistica Sinica*, 14, 1221-1237.