Bayesian sensitivity analysis of incomplete data: bridging pattern-mixture and selection models

Niko A. Kaciroti[a,b*†] and Trivellore Raghunathan[b,c]

Pattern-mixture models (PMM) and selection models (SM) are alternative approaches for statistical analysis when faced with incomplete data and a nonignorable missing-data mechanism. Both models make empirically unverifiable assumptions and need additional constraints to identify the parameters. Here, we first introduce intuitive parameterizations to identify PMM for different types of outcome with distribution in the exponential family; then we translate these to their equivalent SM approach. This provides a unified framework for performing sensitivity analysis under either setting. These new parameterizations are transparent, easy-to-use, and provide dual interpretation from both the PMM and SM perspectives. A Bayesian approach is used to perform sensitivity analysis, deriving inferences using informative prior distributions on the sensitivity parameters. These models can be fitted using software that implements Gibbs sampling. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: identifiability; ignorability index; missing not at random; mixture analysis; time-varying covariates

1. Introduction

Missing data remains a common problem in statistical modeling. A wide range of statistical models for analyzing outcomes with missing data is available, but their validity depends on the nature of the missing-data mechanism as well as on the assumptions used.

For analyzing data with missing values that are potentially missing not at random (MNAR), two widely used approaches are available: pattern-mixture models (PMMs) and selection models (SM). Both approaches derive their inferences based on the joint distribution $f(Y, R)$ of the outcome Y and the response data indicator R, but use a different decomposition for $f(Y, R)$. PMMs express the joint distribution as the product of $f(Y|R)$ and $f(R)$ where data are stratified by missing-data patterns, with distinct parameters for each pattern [1–3]. On the other hand, SM partition $f(Y, R)$ as the product of $f(Y)$ and $f(R|Y)$ [4, 5]. This requires explicit modeling of the missing-data mechanism where the probability that a subject is missing may depend on observed and unobserved values. In addition, SM requires correct specification of $f(y)$ for likelihood-based inferences. However, semiparametric approaches are commonly used where no likelihood function is specified for $f(y)$, while a parametric model is assumed for $f(r|y)$. The choice between using PMM or SM is based on different reasons, such as the analysis objective, the robustness of each model, how best to formulate and incorporate assumptions about the missing-data mechanism, as well as on the flexibility of the sensitivity analysis that is followed. For example, Scharfstein *et al.* [6, 7] and Rotnitzky *et al.* [8, 9] used an SM approach; Little and Rubin [3], Little [10, 11], and Hogan and Laird [12] used a pattern-mixture design. For additional comparisons of SM and PMM, see Glynn [2], Michiels *et al.* [13], and Little [14].

Regardless of which model is used, assumptions not verifiable from the observed data, or additional data, are needed to identify some of the parameters in the joint distribution. These assumptions rely

[a]*Center of Human Growth and Development, University of Michigan, Ann Arbor, MI, U.S.A.*
[b]*Department of Biostatistics, University of Michigan, Ann Arbor, MI, U.S.A.*
[c]*Institute for Social Research, University of Michigan, Ann Arbor, MI, U.S.A.*
Correspondence to: Niko Kaciroti, PhD Center for Human Growth and Development, University of Michigan, Ann Arbor, MI 48109.
[†]*E-mail: nicola@umich.edu*

heavily on expert opinions about a plausible range of non-identifiable parameters and are usually followed with sensitivity analysis [15–18]. A recent report by the National Research Council includes the following recommendation (Recommendation 15): *"Sensitivity analysis should be part of the primary reporting of clinical trials. Examining sensitivity to the assumptions about the missing data mechanism should be a mandatory component of reporting"* [19]. Although both frameworks can be used to do this, it may be more desirable, whenever possible, to choose parameters for sensitivity analysis that are easily interpretable from both PMM and SM settings.

In this article, we introduce Bayesian models for sensitivity analysis across a range of missing-data mechanisms, using missing at random (MAR) as the departure point of sensitivity. We propose intuitive parameterizations for sensitivity analysis focusing on a single outcome response. First, we introduce easy-to-use parameterizations to identify the PMM for different types of outcome with distribution from the exponential family. We then translate these to their equivalent SM approach. The new parameterizations thus provide a dual interpretation under both PMM and SM and are intuitive and easy to communicate to subject-matter experts using either a PMM or SM setting. This presents a unified framework of sensitivity analysis using transparent assumptions about the missing-data mechanism. The Bayesian approach also enables us to provide a probabilistic range to capture differences among observed and missing subjects by introducing an informative prior distribution (with mean l and coefficient of variation c) on the sensitivity parameters. Then, the l and c become second-order sensitivity parameters: l indicates the average difference from MAR, and c indicates our degree of uncertainty on the departure from MAR. A coefficient of $c = 0$ corresponds to a deterministic constrain.

In Section 2, we outline the general framework for outcomes within the exponential family and expand it in Section 3 to include covariates. In Section 4, we apply the proposed method for normal and binary outcomes. Conclusions are given in Section 5.

2. The proposed method

In this section, we introduce a general parameterization within PMM framework and relate it to the corresponding SM for single outcomes y in the exponential family distribution. That is, y has the density $f(y|\theta, \phi) = exp((\theta y - b(\theta))/a(\phi) + c(y, \phi))$, where θ, which can be a scalar or vector (e.g., multinomial distribution), is the main parameter of interest and ϕ is referred to as dispersion parameter. When ϕ is known (e.g., $\phi = 1$ for Poisson), we have an exponential-family model with canonical parameter θ. Examples with ϕ unknown include normal distribution where $\phi = \sigma^2$. We consider binary, multinomial, count, and normal outcomes.

We focus on randomized controlled studies where subjects are randomized into two or more treatment groups and the primary analysis is test for treatment effect. The statistical analysis plan usually follows the intention-to-treat principle where all subjects are included in the analysis regardless of whether they completed the study. Hence, it is important in such studies to use methods that deal with missing data, which are potentially MNAR. Our proposed method can also be applied to nonrandomized settings for group comparisons.

Using the generalized linear models approach [20], the PMM for outcomes within the exponential family is

$$h(E(Y|Trt = g, R = r)) = \theta_g^{(r)}, \tag{1}$$

where $h(.)$ is the link function, Trt is a group indicator ($Trt = 1$ for treatment, $Trt = 0$ for control), and R is the response indicator ($R = 1$ when y is observed and $R = 0$ when y is missing). Often we model $\theta_g^{(r)} = \beta_0^{(r)} + \beta_1^{(r)} Trt$, where $\beta_1^{(r)} = \theta_1^{(r)} - \theta_0^{(r)}$ measures the effect of treatment. Because no data are available to estimate parameters for pattern $R = 0$, model (1) is underidentified. Therefore, we identify such models by relating the distribution of the missing data to that of the observed data via some tilting. So, we assume that $f(y|Trt = g, R = r)$ are from the same distribution family but have different parameters, with the model identified by specifying $\theta_g^{(0)} - \theta_g^{(1)} = log(\tilde{\lambda}_g)$. This parameterization accommodates modeling of missing data generated from a potentially nonignorable missing-data mechanism and also includes the MAR mechanism as a special case when $\tilde{\lambda}_g = 1$. The $\tilde{\lambda}_g$ captures the departure from the MAR and is referred to as ignorability index. It is not estimable from the data, hence, we use the Bayesian framework to perform sensitivity analysis by assuming different informative prior distributions on $\tilde{\lambda}_g$, independent of the data, with some mean l_g, and variance $c^2 l_g^2$. Here, c is the coefficient of variation representing

the degree of uncertainty where $c = 0$ translates to a deterministic constraint. The parameters l_g and c are the sensitivity parameters and are chosen to represent the range of departure from MAR. We chose $\tilde{\lambda}_g \sim log-normal$ to allow for symmetry when capturing the difference between $\theta_g^{(0)}$ and $\theta_g^{(1)}$. However, other prior distributions are also possible.

Next, we show that the identifying parameter $\tilde{\lambda}_g$ allows for a dual interpretation from both PMM and SM framework, thus bridging the sensitivity analysis derived using PMM or SM. The identifying parameter is easy-to-understand and provides flexibility for a subject matter expert to elicit a prior distribution on $\tilde{\lambda}$. It also offers additional clarity on assumptions made about the missing data using either PMM or SM approach.

In PMM, the parameters $\beta^{(r)}$ of each pattern r may themselves be of interest. However, our main interest is on the marginal parameters, β, which are derived as a weighted average across pattern specific estimates [21]. Specifically, in a Bayesian approach, by using simulated draws of $\beta^{(r)}$, the inferences on β are derived via Monte Carlo (MC) simulations, following

$$E(Y|Trt) = (1 - \pi) * E(Y|Trt, R = 0) + \pi * E(Y|Trt, R = 1)$$

or

$$\theta_g = h\left\{(1 - \pi) * h^{-1}\left(\theta_g^{(0)}\right) + \pi * h^{-1}\left(\theta_g^{(1)}\right)\right\},$$

where $\pi = Pr(R = 1|Trt)$ is the proportion of the data observed and depends on Trt. Alternatively, in SM approach, the inferences on marginal parameters are derived directly.

Here, we show an overall connection between the proposed PMM and the corresponding SM for outcomes in the exponential family, then apply it to specific y.

Let $f_r(y|Trt) = f_r(y|Trt, R = r)$, then, following Bayes' rule, we have

$$\frac{Pr(R = 0|Trt, y)}{1 - Pr(R = 0|Trt, y)} = \kappa_{Trt} * \frac{f_0(y|Trt)}{f_1(y|Trt)}, \tag{2}$$

where $\kappa_g = Pr(R = 0|Trt = g)/Pr(R = 1|Trt = g)$. We assume that the distribution of the observed y is from exponential family:

$$f_1(y|Trt = g, \theta^{(1)}, \phi^{(1)}) = exp\left\{\frac{\theta_g^{(1)}y - b\left(\theta_g^{(1)}\right)}{a\left(\phi_g^{(1)}\right)} + c\left(y, \phi_g^{(1)}\right)\right\}.$$

Such an assumption is testable from the data, and if it is true, the following holds:

Proposition 1
The distribution of the missing data is from the same exponential family

$$f_0(y|Trt = g, \theta^{(0)}, \phi^{(0)}) = exp\left\{\frac{\theta_g^{(0)}y - b\left(\theta_g^{(0)}\right)}{a\left(\phi_g^{(0)}\right)} + c\left(y, \phi_g^{(0)}\right)\right\},$$

if, and only if, the selection model of the missing data mechanism follows the following logistic regression model:

$$logit(Pr(R = 0|Trt = g, y)) = \underbrace{log(\kappa_g) + \frac{b\left(\theta_g^{(1)}\right)}{a\left(\phi_g^{(1)}\right)} - \frac{b\left(\theta_g^{(0)}\right)}{a\left(\phi_g^{(0)}\right)}}_{\gamma_{g0}} + \underbrace{\left(\frac{\theta_g^{(0)}}{a\left(\phi_g^{(0)}\right)} - \frac{\theta_g^{(1)}}{a\left(\phi_g^{(1)}\right)}\right)}_{\gamma_{g1}} * y$$

$$+ c\left(y, \phi_g^{(0)}\right) - c\left(y, \phi_g^{(1)}\right). \tag{3}$$

The proof is straightforward by taking the log in both sides of (2).

Here, the PMMs assume a parametric distribution on f_0 and f_1. The parametric assumption about f_1 is testable, and we assume that it is correct. Then, the parametric distribution assumption on f_0 is equivalent to a parametric assumption about the missing-data mechanism. For example, when ψ is fixed (e.g., binomial, Poisson) or the same in each pattern (e.g., normal distribution where $\psi^{(0)} = \psi^{(1)} = \psi$), then the missing-data mechanism (3) follows a logistic regression model that is linear on y. When $\psi^{(0)} \neq \psi^{(1)}$ for normal distribution, the missing-data mechanism (3) is quadratic on y.

The proposed PMM parameterization and its connection with the corresponding SM (3) provide intuitive and easy-to-use tools to assess and understand the assumptions made about missing data using either PMM or SM framework. It allows for sensitivity analysis over a range of sensitivity parameters, which have dual interpretations. It also provides tools to fully understand the underlying parametric assumptions under both PMM and SM and to asess if they are reasonable. For example, if one makes assumptions about the distribution of the missing data in a PMM, but the corresponding selection model is unlikely to have the expected parametric form, then the assumed distribution on the missing data is incorrect. In addition, the equivalence between the PMM and the SM shown in Proposition 1 allows one to apply these approaches interchangeably – taking advantage of what both have to offer.

Next, we give details of the proposed model for commonly used outcomes: normal, count, binary, and multinomial.

2.1. Normal outcomes

The normal distribution is from the exponential family where $f(y) = exp(y\mu/\sigma^2 - \mu^2/2\sigma^2 - y^2/2\sigma^2 - log(\sigma) - log(2\pi)/2)$, and where the mean parameter μ is the main parameter for interest. We consider the following PMM in cases of a normal outcome variable Y with dropout:

$$E(Y|Trt = g, R = r) = \mu_g^{(r)},$$

where $f(Y|Trt = g, R = r)$ is $\mathcal{N}\left(\mu_g^{(r)}, \sigma_g^{2(r)}\right)$ and its natural link function is identity. To identify the PMM model, we relate the distribution of the missing data to that of the observed data. We use a parameterization, which is intuitive under PMM framework and also corresponds to a specific SM. That is, we relate the location parameters (means) and the scale parameters (variance) between the missing subjects and the observed subjects as follows:

$$\mu_g^{(0)} - \mu_g^{(1)} = \log(\tilde{\lambda}_g) \tag{4}$$

and

$$\sigma_g^{(0)2} = \tilde{\psi}_g \sigma_g^{(1)2}. \tag{5}$$

Here, $log(\tilde{\lambda}_g)$ is the mean difference between the missing-data pattern and the observed-data pattern for group g. The $\tilde{\psi}_g$ is the ratio of σ_g^2 parameters between the missing-data pattern and the observed-data pattern for group g. The $\tilde{\lambda}_g$ and $\tilde{\psi}_g$ measure the departure from the MAR. We assume they follow some prior distribution (e.g., log-normal) with mean l_g, ψ_g, and coefficient of variation c.

For PMM identified by (4, 5), the corresponding selection model following (3) is as follows:

$$logit(Pr(R = 0|Trt = g, y)) = \gamma_{g0} + \gamma_{g1}y + \gamma_{g2}y^2,$$

where the unidentified parameters γ_{g1} and γ_{g2} ($g = 0, 1$) are equal to

$$\gamma_{g1} = \mu_g^{(0)}/\sigma_g^{(0)2} - \mu_g^{(1)}/\sigma_g^{(1)2} = \left(\log(\tilde{\lambda}_g) + \mu_g^{(1)}(1 - \tilde{\psi}_g)\right)/\sigma_g^{(1)2}\tilde{\psi}_g \tag{6}$$

$$\gamma_{g2} = 1/2\sigma_g^{(1)2} - 1/2\sigma_g^{(0)2} = (\tilde{\psi}_g - 1)/2\sigma_g^{(1)2}\tilde{\psi}_g. \tag{7}$$

Thus, when the distributions of the missing data and observed data are normal, but with different location and scale parameters, the corresponding SM is a logistic regression that is quadratic on Y. In this situation, the identifiability of the model using a PMM approach is easier to understand than the SM

approach. When $\sigma_g^{(0)} = \sigma_g^{(1)}$ then $\gamma_{g2} = 0$ and the corresponding selection model is linear on Y, with $\gamma_{g1} = \log(\tilde{\lambda}_g)/\sigma_g^{(1)^2}$. Then $\log(\tilde{\lambda}_g)$ has an easy interpretation under SM, it represents the odds ratio of dropping out per $\sigma_g^{(1)^2}$ increase on Y. The converse is true, provided that the distribution of the observed data is normal: if SM is linear on Y, then the distribution of the missing data is also normal with the same variance as that of the observed data, but with a different location parameter.

When the observed data are normally distributed, using PMM and assuming normal distribution for missing data are equivalent to a missing-data mechanism that follows a logistic regression model of up to second order on Y. Such equivalence provides an assurance when PMM are implemented. That is, if a linear or quadratic logistic regression model for missing mechanism is not plausible, then one cannot run a PMM assuming normality. On the other hand, if a quadratic logistic regression for modeling missing-data mechanism is likely, fitting SM is equivalent to fitting a PMM model where the parameters for the missing data are identified using (6, 7).

2.2. Poisson outcomes

Poisson distribution, $f(y|\theta) = exp(\theta y - exp(\theta) - log(y!))$, is widely used for modeling count outcomes where data are not overdispersed. In the presence of missing data, we consider the following PMM for Y:

$$log\left(\mu_g^{(r)}|Trt = g, R = r\right) = \theta_g^{(r)}, \tag{8}$$

where $\mu_g^{(r)} = E(Y|Trt = g, R = r)$ and $log(.)$ is the natural link function. To identify this model, we use the following parameterization:

$$log\left(\mu_g^{(0)}\right) - log\left(\mu_g^{(1)}\right) = log(\tilde{\lambda}_g),$$

or following Bayes' rule

$$\tilde{\lambda}_g = \underbrace{\frac{\mu_g^{(0)}}{\mu_g^{(1)}}}_{PMM} = \underbrace{\frac{Pr(R = 0|Y = y + 1, Trt = g)}{Pr(R = 1|Y = y + 1, Trt = g)} \Big/ \frac{Pr(R = 0|Y = y, Trt = g)}{Pr(R = 1|Y = y, Trt = g)}}_{SM}. \tag{9}$$

Then the corresponding SM based on (3) following parameterization (9) is

$$logit(Pr(R = 0|Trt = g, y)) = \gamma_{g0} + \gamma_{g1}y,$$

where $\gamma_{g1} = log(\tilde{\lambda}_g)$.

The PMM is identified by giving a prior distribution to $\tilde{\lambda}_g$, which has a dual interpretation. In PMM, $\tilde{\lambda}_g$ represents the risk ratio of having an event between the missing subjects and the observed subjects for group g [22]. In SM, it represents the odds ratio of dropping out for a one unit increase in Y for group g.

2.3. Negative binomial outcomes

When count data are overdispersed, we use negative binomial distribution, $f(y) = exp(log(1 - p)y + mlog(p) + log(\Gamma(y + m) - log(y!) - log(\Gamma(m + 1)))$ with mean $\mu = (1 - p)m/p$, which is part of the exponential family when m is known.

For PMM (8) assuming a negative binomial distribution on Y and $m^{(0)} = m^{(1)} = m$, the corresponding selection model is

$$logit(Pr(R = 0)|Trt = g, y)) = \gamma_{g0} + \gamma_{g1}y,$$

where $\gamma_{g1} = log\left(\left(1 - p_g^{(0)}\right) \middle/ \left(1 - p_g^{(1)}\right)\right)$. Then, based on parameterization (9), we obtain

$$\gamma_{1g} = \log\left(\frac{\tilde{\lambda}_g\left(m + \mu_g^{(1)}\right)}{m + \tilde{\lambda}_g \mu_g^{(1)}}\right).$$

Then, for the PMM using a negative binomial distribution on Y, the identifying parameterization using $\tilde{\lambda}_g$ translates to the logistic SM with parameter γ_{g1} as defined earlier. That is, for one unit increase on Y, the odds of dropping out are increased by $\frac{\tilde{\lambda}_g\left(m + \mu_g^{(1)}\right)}{m + \tilde{\lambda}_g \mu_g^{(1)}}$ for group g.

2.4. Binary outcomes

Bernoulli distribution, $f(y|\theta) = exp(\theta y - log(1 + e^\theta))$, is widely used for modeling binary data, for example, the presence or absence of an endpoint (e.g., hypertension). In the presence of missing data, we consider the following PMM for Y:

$$logit\left(p_g^{(r)}|Trt = g, R = r\right) = \theta_g^{(r)},$$

where $p_g^{(r)} = E(Y|Trt = g, R = r) = Pr(Y = 1|Trt = g, R = r)$ and $logit(.)$ is the natural link function. We identify the model by using the following parameterization:

$$logit\left(p_g^{(0)}\right) - logit\left(p_g^{(1)}\right) = log(\tilde{\lambda}_g),$$

or following Bayes' rule

$$\tilde{\lambda}_g = \underbrace{\frac{p_g^{(0)}/1 - p_g^{(0)}}{p_g^{(1)}/1 - p_g^{(1)}}}_{PMM} = \underbrace{\frac{Pr(R = 0|Y = 1, Trt = g)}{Pr(R = 1|Y = 1, Trt = g)} \middle/ \frac{Pr(R = 0|Y = 0, Trt = g)}{Pr(R = 1|Y = 0, Trt = g)}}_{SM}. \tag{10}$$

Then the corresponding SM based on (3) following parameterization (10) is

$$logit(Pr(R = 0|Trt = g, y)) = \gamma_{g0} + \gamma_{g1}y,$$

where $\gamma_{g1} = \log(\tilde{\lambda}_g)$.

The PMM is identified by giving a prior distribution to $\tilde{\lambda}_g$, which allows a dual interpretation. In PMM, $\tilde{\lambda}_g$ represents the odds ratio of having an endpoint between the missing subjects and the observed subjects for group g [23]. In SM, it represents the odds ratio of dropping out between subjects with endpoint versus subjects without endpoint for group g.

2.5. Multinomial outcomes

Multinomial outcome variables are used to indicate various categories, $k = 0, 1, \cdots, K$. The distribution of y is $f(y|\theta) = exp\left(\theta_1 I(y = 1) + \cdots + \theta_K I(y = K) - log\left(1 + e^{\theta_1} + \cdots + e^{\theta_K}\right)\right)$. This multinomial distribution is from a vector exponential family, where $\theta = (\theta_1, \theta_2, \cdots, \theta_K)$ is a vector with K dimension. Multinomial logistic regression, which models the ratio of probabilities being in one category versus a reference category, is commonly used to analyze this type of measures [24]. In the presence of dropouts, we consider the following PMM for multinomial outcome, for $j = 1, 2, \cdots, K$:

$$log\left(p_{gj}^{(r)}/p_{g0}^{(r)}|Trt = g, R = r\right) = \theta_{gj}^{(r)},$$

where $p_{gj}^{(r)} = Pr(Y = j | Trt = g, R = r)$ for $j = 0, \ldots, K$, with $j = 0$ being the reference group. We identify the model by relating the main parameters of interest between the missing-data and the complete-data patterns using

$$\log\left(p_{gj}^{(0)}/p_{g0}^{(0)}\right) - log\left(p_{gj}^{(1)}/p_{g0}^{(1)}\right) = log(\tilde{\lambda}_{gj}),$$

or following Bayes' rule

$$\tilde{\lambda}_{gj} = \underbrace{\frac{p_{gj}^{(0)}/p_{g0}^{(0)}}{p_{gj}^{(1)}/p_{g0}^{(1)}}}_{PMM} = \underbrace{\frac{Pr(R = 0|Trt = g, Y = j)}{Pr(R = 1|Trt = g, Y = j)} / \frac{Pr(R = 0|Trt = g, Y = 0)}{Pr(R = 1|Trt = g, Y = 0)}}_{SM}. \tag{11}$$

Then the corresponding SM based on (3) following parameterization (11) is

$$logit(Pr(R = 0|Trt = g, Y = j)) = \gamma_{g0} + \gamma_{gj}$$

where $\gamma_{gj} = log(\tilde{\lambda}_{gj})$ for $j = 1, 2, \cdots, K$.

The PMM is identified by giving a prior distribution to $\tilde{\lambda}_{gj}$, which has the following dual interpretation. From a pattern-mixture perspective, $\tilde{\lambda}_{gj}$ is the relative risk ratio for $Y = j$ versus $Y = 0$ between the missing subjects and the observed subjects for group g. From an SM perspective, $\tilde{\lambda}_{gj}$ is the odds ratio of having a missing value between subjects in group $Y = j$ versus subjects in reference group $Y = 0$ for group g.

In Table I, we summarize the proposed parameterization using PMM followed by the corresponding SM for each outcome. For simplicity, the group index g is suppressed.

Table I. PMM parameterization and corresponding SM for different outcomes in exponential family.

Pattern-mixture model	PMM parameterization	Corresponding SM
Normal $Y^{(r)} \sim N(\mu^{(r)}, \sigma^{2(r)})$	$\mu^{(0)} = \mu^{(1)} + \log(\tilde{\lambda})$ $\log(\tilde{\lambda}) \sim N(l, c^2 l^2)$	$logit(Pr(R = 0|y)) = \gamma_0 + \gamma_1 y + \gamma_2 y^2$ $\gamma_1 = \frac{\log(\tilde{\lambda}) + \mu^{(1)}(1 - \tilde{\psi})}{\tilde{\psi}\sigma^{(1)2}}$
	$\sigma^{(0)2} = \tilde{\psi}\sigma^{(1)2}$ $\log(\tilde{\psi}) \sim N(\psi, c^2 \psi^2)$	$\gamma_2 = \frac{\tilde{\psi} - 1}{2\tilde{\psi}\sigma^{(1)2}}$
Poisson $Y^{(r)} \sim Poisson(\mu^{(r)})$	$\mu^{(0)} = \tilde{\lambda}\mu^{(1)}$ $\log(\tilde{\lambda}) \sim N(l, c^2 l^2)$	$logit(Pr(R = 0|y)) = \gamma_0 + \gamma_1 y$ $\gamma_1 = \log(\tilde{\lambda})$
Negative Binomial $Y^{(r)} \sim \text{Neg-Bin}(\mu^{(r)}, m)$	$\mu^{(0)} = \tilde{\lambda}\mu^{(1)}$ $\log(\tilde{\lambda}) \sim N(l, c^2 l^2)$	$logit(Pr(R = 0|Yy) = \gamma_0 + \gamma_1 y$ $\gamma_1 = \log\left(\frac{\tilde{\lambda}(m + \mu^{(1)})}{m + \tilde{\lambda}\mu^{(1)}}\right)$
Binary $Y^{(r)} \sim Bernoullie(p^{(r)})$	$\frac{p^{(0)}}{1 - p^{(0)}} = \tilde{\lambda}\frac{p^{(1)}}{1 - p^{(1)}}$ $\log(\tilde{\lambda}) \sim N(l, c^2 l^2)$	$logit(Pr(R = 0|y)) = \gamma_0 + \gamma_1 y$ $\gamma_1 = \log(\tilde{\lambda})$
Multinomial $Y^{(r)} \sim Mult\left(p_0^{(r)}, ..., p_K^{(r)}\right)$	$\frac{p_j^{(0)}}{p_0^{(0)}} = \tilde{\lambda}_j \frac{p_j^{(1)}}{p_0^{(1)}}$ $\log(\tilde{\lambda}_j) \sim N(l_j, c^2 l_j^2)$	$logit(Pr(R = 0|Y = j)) = \gamma_0 + \gamma_j$ $\gamma_j = \log(\tilde{\lambda}_j)$

PMM, pattern-mixture model; SM, selection models.

3. Including covariates

In this section, we expand the proposed PMM to adjust for baseline or time-varying covariates.

3.1. Adjusting for baseline covariates

Baseline covariates are often included in the model for adjusted analysis. In addition, including covariates can reduce the potential selection bias due to dropouts, resulting in a missing-data mechanism that is MAR, or closer to it, and therefore increases the validity of local sensitivity analysis. In this section, we expand our proposed parameterization to the PMM that include baseline covariates x. That is, we consider the following PMM with covariates:

$$h(E(Y|Trt = g, x, R = r)) = \theta_g^{(r)}(x) = \beta_g^{(r)} + \alpha_g^{(r)}x, \tag{12}$$

where x is centered to have mean zero. Without loss of generality, we assume x to be a scalar and identify model (12) by specifying

$$\beta_g^{(0)} = \beta_g^{(1)} + \log(\tilde{\lambda}_g)$$

$$\alpha_g^{(0)} = \alpha_g^{(1)} + \log(\tilde{\lambda}_{\alpha_g}).$$

Then $\log(\tilde{\lambda}_g)$ is the difference between $\theta_g^{(1)}(0) - \theta_g^{(0)}(0)$ for X equal to 0 (its mean), and $\log(\tilde{\lambda}_{\alpha_g})$ indicates how much the effect (slope) of x on y is different between missing and observed data patterns. We assume $\tilde{\lambda}_{\alpha_g}$ follows a log-normal distribution with mean l_{α_g} and coefficient of variation c. For normal outcome with different dispersion parameter between missing data patterns, the additional identifying constrained for the variance is used:

$$\sigma_g^{(0)2} = \tilde{\psi}_g \sigma_g^{(1)2}.$$

Under this parametrization, the corresponding SM for PMM (12) is

$$logit(Pr(R = 0|Trt = g, x, y)) = \underbrace{log(\kappa_g(x)) + \frac{b\left(\theta_g^{(1)}(x)\right)}{a\left(\phi_g^{(1)}\right)} - \frac{b\left(\theta_g^{(0)}(x)\right)}{a\left(\phi_g^{(0)}\right)}}_{\tilde{f}_g(x;\tilde{\psi}_g)} + \left(\frac{\theta_g^{(0)}(x)}{a\left(\phi_g^{(0)}\right)} - \frac{\theta_g^{(1)}(x)}{a\left(\phi_g^{(1)}\right)}\right)y$$

$$+ c\left(y, \phi_g^{(0)}\right) - c\left(y, \phi_g^{(1)}\right).$$

When the dispersion parameter is the same between the missing-data patterns $\left(\phi_g^{(0)} = \phi_g^{(1)} = \phi_g\right)$, the model is reduced to

$$logit(Pr(R = 0|Trt = g, x, y)) = \underbrace{log(\kappa_g(x)) + \frac{b\left(\theta_g^{(1)}(x)\right) - b\left(\theta_g^{(0)}(x)\right)}{a(\phi_g)}}_{\tilde{f}_g(x)} + \frac{\theta_g^{(0)}(x) - \theta_g^{(1)}(x)}{a(\phi_g)}y.$$

The specific SM for each outcome type is given in the succeeding text.

- Normal outcomes

$$logit(Pr(R = 0|Trt = g, x, y)) = \tilde{f}_N(x; \tilde{\psi}_g) + \frac{\log(\tilde{\lambda}_g) + \log(\tilde{\lambda}_{\alpha_g})x + (1 - \tilde{\psi}_g)\mu_g^{(1)}}{\tilde{\psi}_g \sigma_g^{(1)2}}y + \frac{\tilde{\psi}_g - 1}{2\sigma_g^{(1)2}\tilde{\psi}_g}y^2$$

where $\tilde{f}_N(x; \tilde{\psi}_g) = \log(\kappa_g(x)) + \frac{1}{2}\log(\tilde{\psi}_g) + \frac{\tilde{\psi}_g \mu_g^{(1)2} - \left(\mu_g^{(1)} + \log(\tilde{\lambda}_g) + \log(\tilde{\lambda}_{\alpha_g})x\right)^2}{2\tilde{\psi}_g \sigma_g^{(1)2}}$ and $\mu_g^{(1)} = \beta_g^{(1)} + \alpha_g^{(1)}x$.

For $\tilde{\psi} = 1$ the SM reduces to

$$logit(Pr(R = 0|Trt = g, x, y)) = \tilde{f}_N(x) + \frac{\log(\tilde{\lambda}_g) + \log(\tilde{\lambda}_{\alpha_g})x}{\sigma_g^2}y$$

and $\tilde{f}_N(x) = \log(\kappa_g(x)) + \frac{\mu_g^{(1)2} - \left(\mu_g^{(1)} + \log(\tilde{\lambda}_g) + \log(\tilde{\lambda}_{\alpha_g})x\right)^2}{2\sigma_g^2}$.

- Poisson outcomes

$$logit(Pr(R = 0|Trt = g, x, y)) = \tilde{f}_P(x) + (\log(\tilde{\lambda}_g) + \log(\tilde{\lambda}_{\alpha_g})x)y$$

where $\tilde{f}_P(x) = \log(\kappa_g(x)) + \left(1 - \tilde{\lambda}_g \tilde{\lambda}_{\alpha_g}^x\right)\mu_g^{(1)}$ and $\mu_g^{(1)} = e^{\beta_g^{(1)} + \alpha_g^{(1)}x}$.

- Negative binomial outcomes

$$logit(Pr(R = 0|Trt = g, x, y) = \tilde{f}_{NB}(x) + log\left(\frac{\tilde{\lambda}_g \tilde{\lambda}_{\alpha_g}^x \left(m + \mu_g^{(1)}\right)}{m + \tilde{\lambda}_g \tilde{\lambda}_{\alpha_g}^x \mu_g^{(1)}}\right)y$$

where $\tilde{f}_{NB}(x) = \log(\kappa_g(x)) + log\left(\frac{m + \mu_g^{(1)}}{m + \tilde{\lambda}_g \tilde{\lambda}_{\alpha_g}^x \mu_g^{(1)}}\right)$ and $\mu_g^{(1)} = e^{\beta_g^{(1)} + \alpha_g^{(1)}x}$.

- Binomial outcomes

$$logit(Pr(R = 0|Trt = g, x, y)) = \tilde{f}_B(x) + (\log(\tilde{\lambda}_g) + \log(\tilde{\lambda}_{\alpha_g})x)y$$

where $\tilde{f}_B(x) = \log(\kappa_g(x)) + log\left(\frac{1 + e^{\beta_g^{(1)} + \alpha_g^{(1)}x}}{1 + \tilde{\lambda}_g \tilde{\lambda}_{\alpha_g}^x e^{\beta_g^{(1)} + \alpha_g^{(1)}x}}\right)$.

- Multinomial outcomes

$$logit(Pr(R = 0|Trt = g, x, Y = j)) = \tilde{f}_M(x) + \log(\tilde{\lambda}_{gj}) + \log\left(\tilde{\lambda}_{\alpha_{gj}}\right)x$$

where $\tilde{f}_M(x) = \log(\kappa_g(x)) + log\left(\frac{1 + \sum_{j=1}^{K} e^{\beta_{gj}^{(1)} + \alpha_{gj}^{(1)}x}}{1 + \sum_{j=1}^{K} \tilde{\lambda}_{gj} \tilde{\lambda}_{\alpha_{gj}}^x e^{\beta_{gj}^{(1)} + \alpha_{gj}^{(1)}x}}\right)$.

Note that when the effect of the covariate x on the outcome y in model (12) is the same for both missing-data patterns, that is, $\log(\tilde{\lambda}_{\alpha_g}) = 0$, then in corresponding SM, the effect of y on the dropout indicator R does not depend on x and is the same as in the case with no covariates. Hence, in such a case, the interpretation of $\log(\tilde{\lambda}_g)$ under PMM or SM will be the same in both models with or without covariates.

When $\log(\tilde{\lambda}_{\alpha_g}) \neq 0$, in PMM $\log(\tilde{\lambda}_{\alpha_g})$ indicates how much the difference $\theta_g^{(0)}(x) - \theta_g^{(1)}(x)$ changes for one unit increase on x, with $\log(\tilde{\lambda}_g)$ measuring the difference at $x = 0$ (its mean). Under SM, $\log(\tilde{\lambda}_{\alpha_g})$ indicates how much the effect of y on R is modified by x, with $\log(\tilde{\lambda}_g)$ measuring the effect at $x = 0$.

The marginal parameters β_g and α_g will be derived using draws of $\beta_g^{(r)}$ and $\alpha_g^{(r)}$ via Monte Carlo simulations following

$$\beta_g + \alpha_g x = h\left\{(1 - \pi_g) * h^{-1}\left(\beta_g^{(0)} + \alpha_g^{(0)}x\right) + \pi_g * h^{-1}\left(\beta_g^{(1)} + \alpha_g^{(1)}x\right)\right\}$$

or

$$\beta_g + \alpha_g x = \underbrace{h\left\{(1 - \pi_g) * h^{-1}\left(\beta_g^{(1)} + \log(\tilde{\lambda}_g) + \left(\alpha_g^{(1)} + \log\left(\tilde{\lambda}_{\alpha_g}\right)x\right)\right) + \pi_g * h^{-1}(\beta_g^{(1)} + \alpha_g^{(1)}x)\right\}}_{\mathcal{H}_g\left(x; \beta_g^{(1)}, \alpha_g^{(1)}, \tilde{\lambda}_g, \tilde{\lambda}_{\alpha_g}\right)},$$

where $\pi_g = Pr(R = 1|Trt, x)$. Then, we calculate $\beta_g = \mathcal{H}_g(0)$ and $\alpha_g = \mathcal{H}_g(1) - \mathcal{H}_g(0)$.

It is important to note that inferences on β_g will not depend on $\log(\tilde{\lambda}_{\alpha_g})$. Thus, if the focus is in inferences on β_g (and the treatment effect $\beta_1 - \beta_0$), one can use PMM (12) and assume $\alpha_g^{(0)} = \alpha_g^{(1)}$. Under this assumption, $\log(\tilde{\lambda}_{\alpha_g}) = 0$, the interpretation of the identifying parameters $\log(\tilde{\lambda}_g)$ under PMM and SM is the same as in the case of no covariates.

3.2. Adjusting for time-varying covariates

Often in randomized trials, time-varying covariates V are included for adjusted analysis. In such case, the treatment effect on Y, conditional on V is of interest. For simplicity, we do not adjust for baseline covariates x, although the extension applies in the same way as described in Section 3.1. We consider the case where V is missing if Y is missing. Then we fit the following PMM for Y and V:

$$h_y(E(Y|Trt = g, v, R = r)) = \theta_g^{(r)}(v) = \beta_g^{(r)} + \alpha^{(r)}v \tag{13}$$

$$h_v(E(V|Trt = g, R = r)) = v_g^{(r)}, \tag{14}$$

where $\beta_1^{(r)} - \beta_0^{(r)}$ is the treatment effect adjusted for V and $v_1^{(r)} - v_0^{(r)}$ is the treatment effect on V. Models (13) and (14) are fitted sequentially [25], where model (14) is fitted first, then model (13) is fitted next with missing values of V drawn based on model (14). Both models are identified as before by specifying:

$$\beta_g^{(0)} = \beta_g^{(1)} + \log(\tilde{\lambda}_g)$$

$$\alpha^{(0)} = \alpha^{(1)} + \log(\tilde{\lambda}_\alpha)$$

$$v_g^{(0)} = v_g^{(1)} + \log\left(\tilde{\lambda}_{v_g}\right).$$

Then $\log(\tilde{\lambda}_g)$ is the mean difference between patterns, $\beta_g^{(1)} - \beta_g^{(0)}$, controlling for v; $\log(\tilde{\lambda}_\alpha)$ indicates how much the slope of v on y (after controlling for Trt) is different between patterns $R = 0$ and $R = 1$; and $\log\left(\tilde{\lambda}_{v_g}\right)$ has the same distribution and interpretation as described in Table I. For normal outcome (y or v) with different dispersion parameter between missing data patterns, the additional identifying constrained for the variance is used:

$$\sigma_g^{(0)2} = \tilde{\psi}_g \sigma_g^{(1)2}.$$

The corresponding SM for PMMs (13) and (14) is derived following

$$\frac{Pr(R = 0|Trt, y, v)}{1 - Pr(R = 0|Trt, y, v)} = \kappa_{Trt} * \frac{f_0(y|Trt, v)f_0(v|Trt)}{f_1(y|Trt, v)f_1(v|trt)}$$

or

$$logit(Pr(R = 0|Trt = g, x, y)) = \log(\kappa_g) + \underbrace{\log\left(\frac{f_0(v|Trt)}{f_1(v|trt)}\right)}_{SM_1} + \underbrace{\log\left(\frac{f_0(y|Trt, v)}{f_1(y|Trt, v)}\right)}_{SM_2}. \tag{15}$$

Thus, the SM (15) comprises two parts: SM_1 is as shown in Table I, and SM_2 is as shown in Section 3.1 (but replacing x with v). For instance, if y is binary and v follows a normal distribution, the corresponding

SM (15) would be

$$logit(Pr(R = 0|Trt = g, v, y)) = \gamma_{g0} + \gamma_{g1}v + \gamma_{g2}v^2 + \gamma_{g3}y + \gamma_{g4}vy,$$

where the unidentifying parameters γ_{g1}, γ_{g2}, γ_{g3}, and γ_{g4} are equal to

$$\gamma_{g1} = \left(\log\left(\tilde{\lambda}_{v_g}\right) + v_g^{(1)}(1 - \tilde{\psi}_g)\right)/\sigma_g^{(1)2}\tilde{\psi}_g$$

$$\gamma_{g2} = (\tilde{\psi}_g - 1)/2\sigma_g^{(1)2}\tilde{\psi}_g$$

$$\gamma_{g3} = \log(\tilde{\lambda}_g)$$

$$\gamma_{g4} = \log(\tilde{\lambda}_\alpha).$$

The overall parameters β, α_g, and v_g are derived using MC simulations.

4. Application studies

The PMM proposed in Sections 2 and 3 are applied to the data from the TRial Of Preventing Hypertension (TROPHY) [26, 27]. The primary objective of TROPHY was to determine whether for patients with prehypertension, 2 years of treatment with candesartan would prevent an increase on blood pressure and consequently reduce the incidence of hypertension for up to 2 years after active treatment is discontinued. In this 4 year, multicenter, randomized study, a total of 772 subjects were randomly assigned to candesartan group (391) or placebo group (381). We focus on normal data (diastolic blood pressure or DBP) and binary data (presence of hypertension).

4.1. Normal outcome without covariates

In this section, we apply the proposed PMM for normal outcomes identified using (4, 5), to assess the effect of treatment on DBP at the end of the active treatment period (year 2). Data on the DBP are available for 772 subjects at baseline, with 286 subjects having missing values at year 2. We fit the following PMM to test for treatment effect:

$$E(Y|Trt = g, R = r) = \beta_g^{(r)}.$$

Table II. Sensitivity analyses. Monte Carlo simulation errors range $0.001 - 0.006$.

l_0/l_1:	1/1	8/1	8/0.125	1/8	0.125/8
			Posterior mean(SD)		
Model parameters	($\psi_g = 1$)				
β_0: C	81.6(.49)	82.6(.49)	82.6(.49)	81.6(.49)	80.5(.49)
β_1: Trt	77.6(.40)	77.6(.40)	77.2(.40)	78.1(.40)	78.1(.40)
$\beta_1 - \beta_0$: Trt − C	−3.94(.63)	−5.04(.63)	−5.49(.63)	−3.49(.63)	−2.39(.63)
Model parameters	($\psi_g = 2$)				
β_0: C	81.6(.60)	82.6(.60)	82.6(.60)	81.6(.60)	80.5(.60)
β_1: Trt	77.6(.46)	77.6(.46)	77.2(.46)	78.1(.46)	78.1(.46)
$\beta_1 - \beta_0$: Trt − C	−3.94(.75)	−5.04(.76)	−5.49(.76)	−3.49(.76)	−2.39(.76)
Model parameters	($\psi_g = 0.5$)				
β_0: C	81.6(.42)	82.6(.43)	82.6(.43)	81.6(.42)	80.5(.43)
β_1: Trt	77.6(.36)	77.6(.36)	77.2(.37)	78.1(.37)	78.1(.37)
$\beta_1 - \beta_0$: Trt − C	−3.94(.56)	−5.04(.56)	−5.49(.56)	−3.49(.56)	−2.39(.56)

We run sensitivity analysis for values of l_0 and l_1 set at 1 (no difference on means between the observed data and the missing data), 0.125, and 8 (a difference of $\pm.3SD$ between the means of the observed data and the missing data). The sensitivity parameter for the variance, $\psi_0 = \psi_1$, is set at 0.5, 1, and 2. The coefficient of the variation is fixed at $c = 0.1$. The results of sensitivity analysis on the marginal parameters are shown in Table II, where C stands for the control group and Trt for the treatment group. The results show that the effect of candesartan on DBP varies for different values of sensitivity parameters. However, the reduction in DBP is always significant ranging from 2.39 (95%CI = (0.91, 3.88)) to 5.49 (95%CI = (4.39, 6.59)) mmHg. The lowest reduction is when $l_0 = 0.125$, $l_1 = 8$, and $\psi = 2$, and the highest when $l_0 = 8$, $l_1 = 0.125$, and $\psi = 0.5$. In addition, the point estimate for each parameter is not affected by the sensitivity parameter ψ; however, the 95%CI increases as ψ increases.

Table III. Sensitivity analyses. Monte Carlo simulation errors range 0.001–0.006.

			Posterior mean(SD)		
l_0/l_1:	1/1	8/1	8/0.125	1/8	0.125/8
Logistic SM1 ($\psi_g = 1$)					
$\hat{\gamma}_{01}: C \times Y$	$-0.00(.18)$	$0.35(.19)$	$0.35(.19)$	$-0.00(.18)$	$-0.36(.19)$
$\gamma_{01} = \frac{\log(\tilde{\lambda}_0)+\mu_0^{(1)}(1-\tilde{\psi}_g)}{\sigma_0^{(1)^2}\tilde{\psi}_g}$	$0.00(.04)$	$0.35(.09)$	$0.35(.09)$	$0.00(.04)$	$-0.35(.04)$
$\hat{\gamma}_{11}: Trt \times Y$	$-0.01(.20)$	$-0.01(.20)$	$-0.34(.22)$	$0.32(.20)$	$0.32(.20)$
$\gamma_{11} = \frac{\log(\tilde{\lambda}_1)+\mu_1^{(1)}(1-\tilde{\psi}_g)}{\sigma_1^{(1)^2}\tilde{\psi}_g}$	$-0.00(.02)$	$-0.00(.02)$	$-0.31(.06)$	$0.30(.03)$	$0.30(.03)$
$\hat{\gamma}_{02}: C \times Y^2$	$-0.00(.13)$	$0.01(.12)$	$0.01(.12)$	$-0.00(.13)$	$-0.02(.14)$
$\gamma_{02} = \frac{\tilde{\psi}_g-1}{2\sigma_0^{(1)^2}\tilde{\psi}_g}$	$-0.01(.06)$	$-0.01(.06)$	$-0.01(.06)$	$-0.01(.06)$	$-0.01(.06)$
$\hat{\gamma}_{12}: Trt \times Y^2$	$-0.03(.13)$	$-0.03(.13)$	$-0.04(.12)$	$-0.03(.13)$	$-0.03(.13)$
$\gamma_{12} = \frac{\tilde{\psi}_g-1}{2\sigma_1^{(1)^2}\tilde{\psi}_g}$	$-0.01(.05)$	$-0.01(.05)$	$-0.01(.05)$	$-0.01(.05)$	$-0.01(.05)$
Logistic SM2 ($\psi_g = 2$)					
$\hat{\gamma}_2: (1 - Trt) \times Y$	$-0.21(.14)$	$-0.04(.15)$	$-0.04(.15)$	$-0.21(.14)$	$-0.39(.15)$
$\gamma_2 = \frac{\log(\tilde{\lambda}_0)+\mu_0^{(0)}(1-\tilde{\psi}_g)}{\sigma_0^{(0)^2}\tilde{\psi}_g}$	$-0.21(.05)$	$-0.03(.03)$	$-0.03(.06)$	$-0.21(.05)$	$-0.38(.06)$
$\hat{\gamma}_3: Trt \times Y$	$0.12(.16)$	$0.12(.16)$	$-0.04(.16)$	$0.28(.15)$	$0.28(.15)$
$\gamma_3 = \frac{\log(\tilde{\lambda}_1)+\mu_1^{(0)}(1-\tilde{\psi}_g)}{\sigma_1^{(0)^2}\tilde{\psi}_g}$	$0.10(.03)$	$0.10(.03)$	$-0.05(.04)$	$0.26(.04)$	$0.26(.04)$
$\hat{\gamma}_4: (1 - Trt) \times Y^2$	$0.31(.09)$	$0.32(.09)$	$0.32(.09)$	$0.31(.09)$	$0.31(.09)$
$\gamma_4 = \frac{\tilde{\psi}_g-1}{2\sigma_0^{(0)^2}\tilde{\psi}_g}$	$0.29(.04)$	$0.29(.04)$	$0.29(.04)$	$0.29(.04)$	$0.29(.04)$
$\hat{\gamma}_5: Trt \times Y^2$	$0.24(.08)$	$0.24(.08)$	$0.24(.08)$	$0.24(.08)$	$0.24(.08)$
$\gamma_5 = \frac{\tilde{\psi}_g-1}{2\sigma_1^{(0)^2}\tilde{\psi}_g}$	$0.25(.03)$	$0.25(.03)$	$0.25(.03)$	$0.25(.03)$	$0.25(.03)$
Logistic SM3 ($\psi_g = 0.5$)					
$\hat{\gamma}_2: (1 - Trt) \times Y$	$0.40(.30)$	$1.05(.32)$	$1.05(.32)$	$0.40(.30)$	$-0.31(.29)$
$\gamma_2 = \frac{\log(\tilde{\lambda}_0)+\mu_0^{(0)}(1-\tilde{\psi}_g)}{\sigma_0^{(0)^2}\tilde{\psi}_g}$	$0.42(.13)$	$1.12(.21)$	$1.12(.21)$	$0.42(.13)$	$-0.28(.09)$
$\hat{\gamma}_3: Trt \times Y$	$-0.22(.28)$	$-0.22(.30)$	$-0.84(.33)$	$0.43(.30)$	$0.43(.30)$
$\gamma_3 = \frac{\log(\tilde{\lambda}_1)+\mu_1^{(0)}(1-\tilde{\psi}_g)}{\sigma_1^{(0)^2}\tilde{\psi}_g}$	$-0.22(.30)$	$-0.22(.08)$	$-0.82(.14)$	$0.39(.07)$	$0.39(.07)$
$\hat{\gamma}_4: (1 - Trt) \times Y^2$	$-0.65(.24)$	$-0.57(.21)$	$-0.57(.21)$	$-0.65(.24)$	$-0.72(.27)$
$\gamma_4 = \frac{\tilde{\psi}_g-1}{2\sigma_0^{(0)^2}\tilde{\psi}_g}$	$-0.60(.13)$	$-0.60(.13)$	$-0.60(.13)$	$-0.60(.13)$	$-0.60(.13)$
$\hat{\gamma}_5: Trt \times Y^2$	$-0.56(.25)$	$-0.56(.25)$	$-0.52(.22)$	$-0.62(.27)$	$-0.62(.27)$
$\gamma_5 = \frac{\tilde{\psi}_g-1}{2\sigma_1^{(0)^2}\tilde{\psi}_g}$	$-0.52(.11)$	$-0.52(.11)$	$-0.52(.11)$	$-0.52(.11)$	$-0.52(.11)$

Next, we fit the corresponding selection model in WinBUGS where the missing values are drawn based on PMM and the 'complete' data are used to fit

$$logit(Pr(R = 0|Trt = g, y)) = \gamma_{g0} + \gamma_{g1}y + \gamma_{g2}y^2. \qquad (16)$$

Draws of $\hat{\gamma}_{g1}$ and $\hat{\gamma}_{g2}$ based on fitting SM (16) are then compared to draws of the expected γ_{g1} and γ_{g2} as given by (6, 7). The results are given in Table III, (presented per one σ unit increase in Y) and show that the SM model fitted based on the 'complete' data is similar to the expected SM as derived in Section 2.1. The posterior mean for each $\hat{\gamma}$ is the same or very close to the mean of the expected γ, but the posterior SD is larger as it includes the uncertainty related to draws/imputation of the missing values.

We use a Bayesian p value to compare $\hat{\gamma}$ with γ where $p = 2 * min\{Pr(\hat{\gamma} \geqslant \gamma), Pr(\hat{\gamma} \leqslant \gamma)\}$ and is calculated using MC simulation as the smallest of the proportion of $\hat{\gamma} \geqslant \gamma$ or $\hat{\gamma} \leqslant \gamma$ times 2 for two-sided test. Thus, if $H_0 : \hat{\gamma} = \gamma$ is true, then $Pr(\hat{\gamma} \geqslant \gamma) = Pr(\hat{\gamma} \leqslant \gamma) = 0.5$. Small values of $Pr(\hat{\gamma} \geqslant \gamma)$ or $Pr(\hat{\gamma} \leqslant \gamma)$ would indicate evidence against H_0. The histograms for the difference $\hat{\gamma} - \gamma$ for some of the results in Table III and the Bayesian p values are shown in Figure 1. All the histograms are symmetric
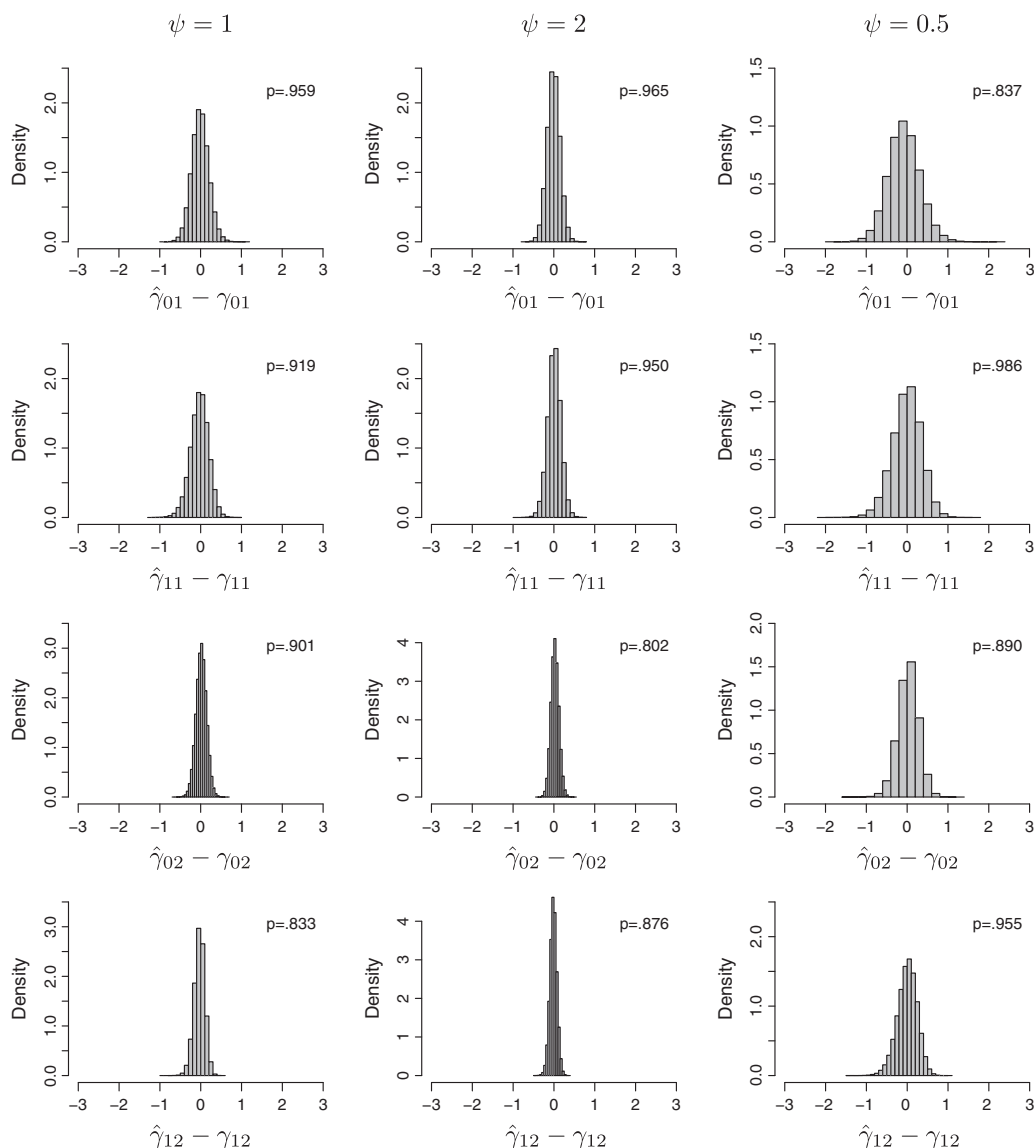

Figure 1. Histograms of the difference $\hat{\gamma} - \gamma$ for $l_0 = 8$, $l_1 = 0.125$, and $\psi \in \{0.5, 1, 2\}$.

around 0, and the p values for all the results in Table III are very high, ranging from 0.660 to 0.997, showing no evidence against H_0.

4.2. Normal outcomes with covariates

In this section, we apply the PMM to the DBP data controlling for baseline *DBP*

$$E(Y|Trt = g, DBP, R = r) = \beta_g^{(r)} + \alpha_g^{(r)}DBP. \tag{17}$$

For simplicity, we fix $\tilde{\psi}_g = 1$, but this approach can be applied for $\tilde{\psi} \neq 1$ in the same way as shown in Section 4.1. The corresponding SM derived in Section 3 for normal outcome is

$$logit(Pr(R = 0|Trt = g, DBP, y)) = \tilde{f}_N(DBP) + (\gamma_g + \delta_g DBP)y. \tag{18}$$

		Posterior mean(SD)			
l_0/l_1:	1/1	8/1	8/0.125	1/8	0.125/8
Model Parameters ($l_\alpha = 1.15$)					
β_0: C	82.3(.46)	83.4(.46)	83.4(.46)	82.3(.46)	81.2(.46)
β_1: Trt	77.8(.39)	77.8(.39)	77.4(.39)	78.2(.39)	78.2(.39)
$\beta_1 - \beta_0$: Trt−C	−4.48(.60)	−5.57(.60)	−6.00(.60)	−4.04(.60)	−2.95(.60)
α_0: DBP	3.11(.47)	3.39(.47)	3.39(.47)	3.11(.47)	2.83(.47)
α_1: DBP	1.81(.42)	1.81(.42)	1.64(.42)	1.98(.42)	1.99(.42)
Logistic SM					
$\hat{\gamma}_0$: Y	0.00(.03)	0.43(.06)	0.43(.06)	0.00(.03)	−0.43(.05)
$\gamma_0 = \log(\tilde{\lambda}_0)/\sigma_0^2$	0.00(.02)	0.43(.05)	0.43(.05)	0.00(.02)	−0.43(.05)
$\hat{\gamma}_1$: Y	0.00(.02)	0.00(.02)	−0.32(.04)	0.31(.04)	0.31(.04)
$\gamma_1 = \log(\tilde{\lambda}_1)/\sigma_1^2$	0.00(.02)	0.00(.02)	−0.32(.03)	0.32(.03)	0.32(.03)
$\hat{\delta}_0$: X × Y	0.03(.03)	0.04(.03)	0.04(.03)	0.03(.03)	0.04(.03)
$\delta_0 = \log(\tilde{\lambda}_{\alpha_0})/\sigma_0^2$	0.03(.02)	0.03(.02)	0.03.(.02)	0.03(.02)	0.03(.02)
$\hat{\delta}_1$: X × Y	0.02(.02)	0.02(.03)	0.03(.03)	0.02(.02)	0.02(.02)
$\delta_1 = \log(\tilde{\lambda}_{\alpha_1})/\sigma_1^2$	0.02(.02)	0.02(.02)	0.02(.02)	0.02(.02)	0.02(.02)
Model Parameters ($l_\alpha = 0.85$)					
β_0: C	82.3(.46)	83.4(.46)	83.4(.46)	82.3(.46)	81.2(.46)
β_1: Trt	77.8(.39)	77.8(.39)	77.4(.39)	78.2(.39)	78.2(.39)
$\beta_1 - \beta_0$: Trt−C	−4.48(.60)	−5.57(.60)	−6.00(.60)	−4.04(.60)	−2.95(.60)
α_0: DBP	2.91(.47)	3.19(.47)	3.19(.47)	2.91(.47)	2.63(.47)
α_1: DBP	1.73(.42)	1.73(.42)	1.56(.42)	1.90(.42)	1.89(.42)
Logistic SM					
$\hat{\gamma}_0$: Y	0.00(.03)	0.43(.05)	0.43(.05)	0.00(.03)	−0.43(.05)
$\gamma_0 = \log(\tilde{\lambda}_0)/\sigma_0^2$	0.00(.02)	0.43(.05)	0.43(.05)	0.00(.02)	−0.43(.05)
$\hat{\gamma}_1$: Y	−0.00(.02)	0.00(.02)	−0.32(.04)	0.31(.04)	0.31(.04)
$\gamma_1 = \log(\tilde{\lambda}_1)/\sigma_1^2$	−0.00(.02)	0.00(.02)	−0.32(.03)	0.32(.03)	0.32(.03)
$\hat{\delta}_0$: X × Y	−0.03(.03)	−0.02(.03)	−0.02(.03)	−0.03(.03)	−0.03(.03)
$\delta_0 = \log(\tilde{\lambda}_{\alpha_0})/\sigma_0^2$	−0.03(.03)	−0.03(.02)	−0.03(.02)	−0.03(.02)	−0.03(.02)
$\hat{\delta}_1$: X × Y	−0.02(.03)	−0.02(.03)	−0.02(.03)	−0.02(.03)	−0.02(.03)
$\delta_1 = \log(\tilde{\lambda}_{\alpha_1})/\sigma_1^2$	−0.03(.02)	−0.03(.02)	−0.03(.02)	−0.03(.02)	−0.03(.02)

Table IV. Sensitivity analyses. Monte Carlo simulation errors range $0.001 − 0.006$.

We run sensitivity analysis for model (17) with values of l_0 and l_1 ranging from 0.125 to 8, l_{α_0} and l_{α_1} set at 1.15 and 0.85, and $c = 0.1$. For each value of l_g, l_{α_g}, and c, missing data are drawn based on PMM and the 'complete' data are used to fit SM (18). Draws of $\hat{\gamma}_g$ and $\hat{\delta}_g$ based on fitting SM (18) are then compared to draws of $\gamma_g = \log(\tilde{\lambda}_g)/\sigma_g^2$ and $\delta_g = \log\left(\tilde{\lambda}_{\alpha_g}\right)/\sigma_g^2$ as given in Section 3.1.

The results in Table IV show a robust treatment effect over a range of MNAR missing-data mechanisms. The treatment is effective resulting in 2.95 (95%CI = (1.77, 4.13)) to 6.00 (95%CI = (4.82, 7.19)) mmHg reduction in DBP at 2 years. In addition, as expected, the estimates of β_0, β_1 and the treatment effect $\beta_1 - \beta_0$ are the same for $l_\alpha = 1.15$ and $l_\alpha = 0.85$.

The posterior mean for each $\hat{\gamma}$ and $\hat{\delta}$ (per one σ increase on Y) is the same or very close to the mean of the expected γ and δ, but the posterior SD is larger because it includes the uncertainty related to draws of the missing values. Bayesian p values for testing $H_0 : \hat{\gamma} = \gamma$ (or $H_0 : \hat{\delta} = \delta$) are very high, ranging from 0.781 to 0.994, showing no evidence against H_0.

4.3. Binary outcome with covariates

We apply the PMM for binary outcomes proposed in Section 3 to assess the effect of treatment on reducing the development of hypertension during the 4 year period of the study. The endpoint outcome is the presence of hypertension (yes/no) at year 4. At the end of the study, out of the 772 randomized subjects, 199 have a missing value on the endpoint outcome. We fit a PMM logistic regression model to test for a treatment effect on hypertension, controlling for baseline DBP. Initial analysis based on the observed data shows that the effect of DBP on predicting hypertension is the same in both groups; hence, we fit

Table V. Sensitivity analyses. Monte Carlo simulation errors range $0.001 - 0.006$.

		Posterior mean(SD)			
l_0/l_1:	1/1	1.5/1	1.5/0.5	1/1.5	0.5/1.5
Model Parameters ($l_\alpha = 1.05$)					
β_0: C	0.96(.14)	1.06(.14)	1.06(.14)	0.96(.14)	0.76(.14)
β_1: Trt	0.31(.13)	0.31(.13)	0.14(.13)	0.41(.13)	0.41(.13)
$\beta_1 - \beta_0$: Trt−C	−0.65(.19)	−0.75(.19)	−0.92(.19)	−0.55(.19)	−0.35(.19)
α: DBP	0.14(.03)	0.14(.03)	0.14(.03)	0.14(.03)	0.14(.03)
Logistic SM					
$\hat{\gamma}_0$: Y	−0.04(.32)	0.36(.30)	0.36(.30)	−0.05(.32)	−0.75(.37)
$\gamma_0 = \log(\tilde{\lambda}_0)$	−0.01(.10)	0.40(.10)	0.40(.10)	−0.01(.10)	−0.70(.10)
$\hat{\gamma}_1$: Y	0.02(.24)	0.02(.24)	−0.68(.27)	0.42(.23)	0.42(.23)
$\gamma_1 = \log(\tilde{\lambda}_1)$	0.00(.10)	0.00(.10)	−0.70(.10)	0.40(.10)	0.40(.10)
$\hat{\delta}$: X × Y	0.03(.09)	0.03(.09)	0.03(.09)	0.03(.09)	0.03(.09)
$\delta = \log(\tilde{\lambda}_\alpha)$	0.04(.10)	0.04(.10)	0.04(.10)	0.04(.10)	0.04(.10)
Model Parameters ($l_\alpha = 0.95$)					
β_0: C	0.96(.14)	1.06(.14)	1.06(.14)	0.96(.14)	0.76(.14)
β_1: Trt	0.31(.13)	0.31(.13)	0.14(.13)	0.41(.13)	0.41(.13)
$\beta_1 - \beta_0$: Trt−C	−0.65(.19)	−0.75(.19)	−0.92(.19)	−0.55(.19)	−0.35(.19)
α: DBP	0.11(.03)	0.11(.03)	0.11(.03)	0.11(.03)	0.11(.03)
Logistic SM					
$\hat{\gamma}_0$: Y	−0.05(.28)	0.35(.27)	0.36(.27)	−0.05(.28)	−0.75(.32)
$\gamma_0 = \log(\tilde{\lambda}_0)$	−0.01(.10)	0.40(.10)	0.40(.10)	−0.01(.10)	−0.70(.10)
$\hat{\gamma}_1$: Y	−0.01(.24)	−0.01(.29)	−0.71(.28)	0.40(.23)	0.40(.23)
$\gamma_1 = \log(\tilde{\lambda}_1)$	−0.01(.10)	−0.01(.10)	−0.70(.10)	0.40(.01)	0.40(.10)
$\hat{\delta}$: X × Y	−0.06(.10)	−0.06(.10)	−0.06(.10)	−0.06(.10)	−0.06(.10)
$\delta = \log(\tilde{\lambda}_\alpha)$	−0.06(.10)	−0.06(.10)	−0.06(.10)	−0.06(.10)	−0.06(.10)

the following PMM

$$logit(Pr(Y = 1|Trt = g, DBP, R = r)) = \beta_g^{(r)} + \alpha^{(r)}DBP. \qquad (19)$$

The corresponding SM model for binary outcome as derived in Section 3 is

$$logit(Pr(R = 0|Trt = g, DBP, y)) = \tilde{f}_B(DBP) + (\gamma_g + \delta DBP)y. \qquad (20)$$

Following PMM (19), we perform sensitivity analysis using log-normal prior distribution on $\tilde{\lambda}_g$ and $\tilde{\lambda}_\alpha$. We run sensitivity analysis for values of l_0 and l_1 set at 1 (no difference on odds of developing hypertension between the observed data and the missing data), 0.5, and 1.5; and l_α set at 0.95 and 1.05; and $c = 0.1$. For each value of l_0, l_1, l_α, and c, missing data are drawn based on PMM, and the 'complete' data are used to fit SM (20). Draws of $\hat{\gamma}_g$ and $\hat{\delta}_g$ based on fitting SM (20) are then compared to draws of the expected $\gamma_g = \log(\tilde{\lambda}_g)$ and $\delta_g = \log(\tilde{\lambda}_{\alpha_g})$ as given in Section 3.1.

The results given in Table V show a strong treatment effect over a range of MNAR missing-data mechanisms. The treatment is effective in reducing the odds of developing hypertension at 4 years with odds ratio (OR) ranging from OR = 0.40 (95%CI = (0.27, 0.57)) to OR = 0.70 (95%CI = (0.48, 1.01)). The lowest reduction is when $l_0 = 0.5$, $l_1 = 1.5$, and the highest when $l_0 = 1.5$, $l_1 = 0.5$. In addition, as expected, the estimates of β_0, β_1, and the treatment effect $\beta_1 - \beta_0$, are the same for $l_\alpha = 1.05$ and $l_\alpha = 0.95$.

Finally, the posterior mean for each $\hat{\gamma}$ and $\hat{\delta}$ is the same or very close to the mean of γ and δ, while the posterior SD is larger because it includes the uncertainty related to draws/imputations of the missing values. Bayesian p values for testing $H_0 : \hat{\gamma} = \gamma$ (or $H_0 : \hat{\delta} = \delta$) are very high, ranging from 0.906 to 0.998, showing no evidence against H_0.

5. Conclusions

In this article, we introduce a parameterization using PMM to analyze outcomes from the exponential distribution family with missing values that are potentially MNAR. We focus on randomized control trials (RCT) where the primary hypothesis is to test for group differences. For each outcome (binary, multinomial, count, and normal), we introduce easy-to-understand parameters that are used both to identify the model and for sensitivity analysis. We then relate the identifying parameters from the PMM to their equivalent corresponding missing-data mechanism in SM approach. Such equivalence results in a dual interpretation of the identifying parameters under both PMM and SM perspectives. This provides a unified framework for performing sensitivity analysis based on transparent assumptions about the missing-data mechanism. It also provides flexibility and assurance to a subject expert for eliciting information under both PMM and SM setting. This is used to construct informative prior distributions to identify the model. We use different informative priors for identifying parameters to capture a range of different missing data scenarios, and then apply sensitivity analysis to evaluate the range of the results towards different assumed prior distributions.

When baseline covariates x or time-varying covariates v are included in the model, additional assumptions are required to identify the parameters for x and v in the missing-data pattern. We extend PMM including covariates and derive the corresponding SM. We show that if the focus of the analysis is on the group comparisons, for example the overall treatment effect in a RCT, then the inferences on the treatment effect are robust to identifying assumptions about the parameters of baseline covariates x. In such cases, one could simplify the PMM (and consequently the corresponding SM) by constraining the parameters of x to be the same in both observed data and missing data patterns.

By relating PMM to an equivalent SM, we bridge the sensitivity analysis between two approaches and consequently reassure the validity of the PMM assumed. For example, if we assume a PMM using normal distribution with different means and variances between observed data and missing data patterns, but the SM is unlikely to be a logistic model quadratic in y, then the normal distribution assumed for the missing data in the PMM is not correct. In such cases, nonparametric PMM [28] or SM models may be more appropriate. Further, the equivalence between PMM and SM following proposition 1 allows PMM and SM to be fitted interchangeably. Thus, one could derive inferences based on the posterior distribution using the PMM approach while formulating the missing-data problem under either a PMM or an SM framework.

In future work, we plan to expand this work for data in a longitudinal setting, which is considerably more complicated [29–31]. We also plan to expand our models to outcomes with distributions in the double exponential family [32] for overdispersed data with missing values.

Acknowledgements

References

1. Rubin DB. Formalizing subjective notions about the effect of nonrespondents in sample surveys. *Journal of the American Statistical Association* 1977; **72**:538–543.
2. Glynn RJ, Liard NM, Rubin DB. Selection modelling versus mixture modelling with nonignorable nonresponse. In *Drawing Inference from Self-selected Samples*, Wainer H (ed.) Springer-Verlag: New York, 1986; 115–142.
3. Little RJA, Rubin DB. *Statistical Analysis with Missing Data* (2nd edn). John Wiley: New-York, 2002.
4. Heckman JJ. 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* 1979; **5**:475–492.
5. Diggle P, Kenward M G. Informative drop-out in longitudinal data analysis. *Applied Statistics* 1994; **43**:49–73.
6. Scharfstein DO, Daniels MJ, Robins JM. Adjusting for non-ignorable drop-out using semiparametric nonresponse models (with discussion). *Biostatistics* 1999; **4**(4):495–512.
7. Scharfstein DO, Daniels MJ, Robins JM. Incorporating prior beliefs about selection bias into the analysis of randomized trials with missing outcomes. *Journal of the American Statistical Association* 2003; **94**:1096–1146.
8. Rotnitzky A, Robins JM, Scharfstein DO. Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association* 1998; **93**:1321–1339.
9. Rotnitzky A, Scharfstein DO, Su T, Robins JM. Methods for conducting sensitivity analysis of trials with potentially non-ignorable competing courses of censoring. *Biometrics* 2001; **57**:103–113.
10. Little RJA. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 1993; **88**:125–134.
11. Little RJA. A class of pattern-mixture models for normal missing data multivariate. *Biometrika* 1994; **81**:471–483.
12. Hogan JWJ, Laird NM. Mixture models for joint distribution of repeated measures and event times. *Statistics in Medicine* 1997; **16**:239–257.
13. Michiels B, Molenberghs G, Lipstiz SR. Selection models and patter-mixture models for incomplete data with covariates. *Biometrics* 1999; **55**:978–983.
14. Little RJA. Selection and pattern-mixture models. In *Advances in Longitudinal Data Analysis*, Fitzmaurice G, Daviadian M, Verbeke G, Molenberghs G (eds). CRC Press: London, 2009; Chapter 18, 409–431.
15. Troxel AB, Ma G, Heitjan D. An index to local sensitivity of nonignorability. *Statistica Sinica* 2004; **14**:1221–1237.
16. Molenberghs G, Verbeke G. *Models for Discrete Longitudinal Data*. Springer: New York, 2005.
17. Vansteelandt S, Goetghebeur E, Kenward MG, Molenberghs G. Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica* 2006; **90**:953–979.
18. Daniels M, Hogan J. *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Chapman & Hall: Boca Raton FL, 2007.
19. National Research Council. The prevention and treatment of missing data in clinical trials. In *Panel on Handling Missing Data in Clinical Trials*, Committee on national Statistics, Division of Behavioral and Social Science and Education. The National Academies Press: Washington D.C., 2010; Chapter 5, 83–106.
20. McGullagh P, Nelder JA. *Generalized Linear Models* (2nd edn). Chapman & Hall: London, 1989.
21. Little RJA. Modeling the dropout mechanism in repeated measures studies. *Journal of the American Statistical Association* 1995; **90**:1113–1121.
22. Kaciroti N, Raghunathan TE, Schork MA, Clark NM. A Bayesian model for longitudinal count data with non-ignorable dropout. *Applied Statistics* 2006; **57**:521–534.
23. Kaciroti N, Schork MA, Raghunathan TE, Julius S. A Bayesian sensitivity model for intention-to-treat analysis on binary outcomes with dropouts. *Statistics in Medicine* 2009; **28**:572–585.
24. Agresti A. *Categorical Data Analysis* (2nd edn). John Wiley & Sons: New-York, 2002.
25. Kaciroti N, Raghunathan TE, Schork MA, Clark NM, Gong M. A Bayesian model for clustered longitudinal ordinal outcome with nonignorable missing data: evaluation of an asthma education program. *Journal of the American Statistical Association* 2006; **101**:435–446.
26. Julius S, Nesbitt S, Egan B, Weber MA, Michelson EL, Kaciroti N, Black HR, Grimm RH, Messerli FH, Oparil S, Schork MA, for the Trial of Prevention Hypertension (TROPHY) Study Investigators. Feasibility of treating prehypertension with an angiotensin-receptor blocker. *New England Journal of Medicine* 2006; **354**:1685–1697.
27. Julius S, Kaciroti N, Nesbitt S, Egan B, Michelson EL. TROPHY study: outcomes based on the JNC 7 definition of hypertension. *Journal of the American Society of Hypertension* 2008; **2**(1):39–43.
28. Kim JK, Yu CL. A semiparametric estimation of mean functionals with nonignorable missing data. *Journal of the American Statistical Association* 2011; **106**:157–165.
29. Kenward MG, Molenberghs G, Thijs H. Pattern-mixture models with proper time dependence. *Biometrika* 2003; **90**:53–71.
30. Kenward MG, Molenberghs G. Last observation carried forward: a crystal ball? *Journal of Biopharmaceutical Statistics* 2009; **19**:872–888.
31. Kaciroti N, Raghunathan TE, Taylor JMG, Julius S. A Bayesian model for discrete time-to-event data with informative censoring. *Biostatistics* 2012; **13**(2):341–554.
32. Effron B. Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association* 1986; **81**:709–721.