Measuring the Impact of Nonignorable Missingness Using the R Package isni

Hui Xie, Weihua Gao, Baodong Xing, Daniel F. Heitjan, Donald Hedeker, Chengbo Yuan

Abstract

The popular assumption of *ignorability* simplifies analyses with incomplete data, but if it is not satisfied, results may be incorrect. One can evaluate the dependence of inferences on this assumption by measuring their sensitivity to its violation. One tool for such an analysis is the *index of local sensitivity to nonignorability (ISNI)*, which evaluates the rate of change of model estimates in the neighborhood of an ignorable model. Computation of ISNI avoids the need to estimate any nonignorable model or to posit a specific magnitude of nonignorability. In this article we introduce a new R package, named **isni**, that implements the method for some common data structures and corresponding statistical models. Specifically, our package computes ISNI in the generalized linear model for independent data, and in the marginal multivariate Gaussian model and the linear mixed model for longitudinal/clustered data. **isni** allows for arbitrary patterns of missingness in the regression outcomes caused by dropout and/or intermittent missingness. Examples illustrate its use and features.

Keywords: Coarse data; exponential family; longitudinal data; missing not at random; selection model.

1. Introduction

Ignorability is the primary working assumption in the analysis of data with missing observations. When the data are missing at random (MAR) and the parameters of the ideal-data model and missing-data model are distinct (PD), the missingness mechanism is ignorable in the sense that one can generate valid inferences from straightforward likelihood/Bayesian analyses that do not require further modeling of it (Rubin 1976; Heitjan and Rubin 1991). This greatly simplifies analyses, because the missing data process can be difficult to model and is rarely of primary interest. Yet in practice we often suspect that the underlying missing data mechanism is nonignorable, typically because MAR does not hold. For example, if the absence of treatment effects, or the presence of excessive toxicity, gives rise to missing observations, the missingness may be related to the unobserved data values, even after conditioning on all available information. This constitutes a violation of MAR, and therefore calls ignorability into question. When ignorability does not hold, straightforward likelihood/Bayesian analyses can yield inferences that incorrectly summarize the information in the data.

Because we cannot test MAR robustly using only the observed data (Little, D'Agostino, Cohen, Dickersin, Emerson, Farrar, Frangakis, Hogan, Molenberghs, Murphy, Neaton, Rotnitzky, Scharfstein, Shih, Siegel, and Stern 2012), it is critical to have the means to conduct an analysis of sensitivity to departures from ignorability (Little *et al.* 2012; Daniels and Hogan 2008). Indeed, an expert panel convened to study the issue has declared that "[s]ensitivity analyses should be part of the primary reporting of findings from clinical trials. Examining sensitivity to the assumptions about the missing data mechanism should be a mandatory component of reporting." (on Handling Missing Data in Clinical Trials; National Research Council 2010) Moreover, "[s]ensitivity analysis is a relatively new area, and further research on the best methods is needed." (Little *et al.* 2012)

Estimating nonignorable models of missingness is conceptually and computationally challenging, thus limiting the number and types of sensitivity analyses that one can perform (Xie 2012; Xie and Qian 2012). One simple approach to evaluating sensitivity to nonignorable missingness is to compute the *index of local sensitivity to nonignorability*, or *ISNI*. Troxel, Ma, and Heitjan (2004) proposed ISNI as a tool to evaluate the potential effect of nonignorable missingness when the outcome follows a generalized linear model. The ISNI method utilizes the local sensitivity approach that examines the effect on inferences of minor departures from ignorability (Copas, JB, and Li, HG 1997; Copas, JB, and Eguchi, S 2001; Verbeke, Molenberghs, Thijs, Lesaffre, and Kenward 2001). The method has since been extended to a range of statistical models and missing data patterns (Xie and Heitjan 2004; Ma, Troxel, and Heitjan 2005; Xie and Heitjan 2009; Xie 2008, 2009, 2010, 2012; Xie, Qian, and Qu 2011; Xie and Qian 2012; Gao, Hedeker, Mermelstein, and Xie 2016). ISNI overcomes the computational problems of sensitivity analysis by requiring only the readily-available model computations under MAR, thereby avoiding the estimation of complicated nonignorable models.

The absence of general software for computing ISNI has hampered its widespread adoption. This article describes a new R package, denoted **isni**, that performs ISNI computations for some common models and data structures. **isni** currently computes sensitivity analyses for three types of popular statistical models: Generalized linear models for cross-sectional data, and marginal multivariate Gaussian models and linear mixed-effects models for longitudinal/clustered data.

We organize the article as follows: Section 2 describes the ISNI approach to sensitivity analysis. Section 3 describes the use of **isni**. Section 4 illustrates the application of **isni** and the interpretation of its outputs in real-data examples. Section 5 offers a summary and discussion. An Appendix presents further computational details.

2. Review of the ISNI Method

Let Y be a vector of outcomes, and G be the vector of missingness indicators with the same length as Y, where each element of G takes the value of 0(1) when the corresponding element of Y is observed(missing). We further define $Y = (Y^{\mathrm{obs}}, Y^{\mathrm{mis}})$, where $Y^{\mathrm{obs}}, Y^{\mathrm{mis}}$ denote the observed and missing elements in Y, respectively. We specify the joint distribution of Y and G with a selection model; that is, the marginal density of Y is $f_\theta(y)$, indexed by parameter θ, and the conditional probability mass function of G given $Y = y$ is $f_{\gamma_0,\gamma_1}(g|y)$, indexed by parameters γ_0 and γ_1. We further assume that we can write the conditional probability function of G given $Y = y$ as $f_{\gamma_0,\gamma_1}(g|y^{\text{obs}}, y^{\text{mis}})$, such that the parameter γ_0 associates the probability of missingness with a set of fully observed missingness predictors and the parameters γ_1 associates the probability of missingness with the unobserved outcomes Y^{mis}. We denote the conditional distribution of the missingness indicator to be the *missing data mechanism (MDM)*. Under these conditions, we define the observed data (y^{obs}, g) to be *missing at random (MAR)* if, for every possible value of the parameters, $f_{\gamma_0,\gamma_1}(g|y^{\text{obs}}, y^{\text{mis}})$ takes the same value for all y^{mis}. For example, if there are no missing observations, the observed data are MAR by default, even if the MDM stipulates a strong correlation of Y and G. The MDM is MAR if every possible data set generated under it is MAR.

Following common practice, we formulate the MDM in such a way that it reduces to MAR when $\gamma_1 = 0$. When $\gamma_1 \neq 0$, the missingness probability depends on unobserved data, y^{mis}, even after conditioning on the observed data. Because γ_1 captures the magnitude of nonignorability, we denote it the nonignorability parameter. Under the general selection model, the loglikelihood $L(\theta, \gamma_0, \gamma_1; y^{\text{obs}}, g)$ is

$$L(\theta, \gamma_0, \gamma_1; y^{\text{obs}}, g) \;=\; \ln \int_{\Omega_{Y^{\text{mis}}}} f_\theta(y^{\text{obs}}, y^{\text{mis}}) f_{\gamma_0,\gamma_1}(g|y^{\text{obs}}, y^{\text{mis}}) dy^{\text{mis}}, \tag{1}$$

where $\Omega_{Y^{\text{mis}}}$ denotes the sample space of Y^{mis}, and $f_\theta(y^{\text{obs}}, y^{\text{mis}})$ is $f_\theta(y)$ evaluated at the observed data y^{obs} and for posited values y^{mis}; the completeness indicator g dictates the identities of the observed and missing y values. Note that if we assume an MAR mechanism (i.e., that $\gamma_1 = 0$), we have $f_{\gamma_0,\gamma_1}(g|y^{\text{obs}}, y^{\text{mis}}) = f_{\gamma_0,\gamma_1=0}(g|y^{\text{obs}})$, because this conditional probability has no dependence on values of y^{mis}. Thus we can move this term out of the integration in Eqn (1), resulting in the simpler loglikelihood

$$\begin{aligned} L_I(\theta, \gamma_0, \gamma_1; y^{\text{obs}}, g) &= \ln \left[\int_{\Omega_{Y^{\text{mis}}}} f_\theta(y^{\text{obs}}, y^{\text{mis}}) dy^{\text{mis}} \right] f_{\gamma_0,\gamma_1=0}(g|y^{\text{obs}}) \\ &= \ln f_\theta(y^{\text{obs}}) + \ln f_{\gamma_0,\gamma_1=0}(g|y^{\text{obs}}). \end{aligned} \tag{2}$$

Under MAR and the additional assumption of *parameter distinctness* — i.e., that θ and γ are independent (for Bayesian inference) or have parameter spaces that factor (for likelihood inference) — $\ln f_\theta(y^{\text{obs}})$ contains the correct information on θ, and therefore it is unnecessary to to estimate the MDM. We denote such an analysis the *ignorability analysis*. The more general analysis that does not assume ignorability bases inferences on Eqn (1), which requires positing the MDM in detail. Because observed data alone provide no robust information to assess the dependence of the missingness probability on y^{mis}, such a nonignorable model is difficult to identify and estimate without additional data or other untestable assumptions.

Thus, practical analyses typically assume MAR and base inference on Eqn (2). To evaluate the robustness of such an inference, one can perform a sensitivity analysis; that is, one posits a range of values of γ_1 and determines the extent to which an estimate of θ, such as the MLE given γ_1, $\hat{\theta}(\gamma_1)$, depends on the values of γ_1. The ISNI approach is to execute this analysis in a neighborhood of the MAR model by determining the rate of change of $\hat{\theta}(\gamma_1)$ as a function of γ_1 at $\gamma_1 = 0$. That is, ISNI calculates the derivative of $\hat{\theta}(\gamma_1)$ with respect to γ_1, evaluated locally at the ignorable model (Troxel *et al.* 2004). As we show in Appendix A, a general formula for ISNI is

$$\text{ISNI} = \left. \frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_1^T} \right|_{\gamma_1=0} \;=\; -\nabla^2 L_{\theta,\theta}^{-1} \nabla^2 L_{\theta,\gamma_1},$$

where

$$\nabla^2 L_{\theta,\theta} = \frac{\partial^2 L(\theta,\gamma_0,\gamma_1)}{\partial\theta\partial\theta^T}\bigg|_{\hat\theta(0),\hat\gamma_0(0),\gamma_1=0} = \frac{\partial \ln f_\theta(y^{\text{obs}})}{\partial\theta\partial\theta^T}\bigg|_{\hat\theta(0),\hat\gamma_0(0),\gamma_1=0},$$

$$\nabla^2 L_{\theta,\gamma_1} = \frac{\partial^2 L(\theta,\gamma_0,\gamma_1)}{\partial\theta\partial\gamma_1^T}\bigg|_{\hat\theta(0),\hat\gamma_0(0),\gamma_1=0},$$

and $L(\theta,\gamma_0,\gamma_1)$ denotes the loglikelihood of Eqn (1). The computation of ISNI thus involves two parts: First one computes $\nabla^2 L_{\theta,\theta}^{-1}$, which is the inverse of the observed information matrix of θ under the MAR outcome model; this is readily available from standard statistical software. Second, one computes $\nabla^2 L_{\theta,\gamma_1}$, which measures the lack of orthogonality of θ and γ_1. Below we show ISNI formulas for some popular statistical models.

2.1. ISNI for Independent data

We first consider the case where Y follows a generalized linear model (GLM) (McCullagh and Nelder 1989) that assumes that scalar Y_i, given predictors $x_i, i = 1, \cdots, N$, are independent draws from the exponential family

$$f_\theta(y_i|x_i) = \exp\left\{\frac{y_i\Psi_i(\beta,x_i) - b(\Psi_i(\beta,x_i))}{a(\tau)} + c(y_i,\tau)\right\}, \tag{3}$$

where Ψ_i is the canonical parameter as a function of the regression coefficient parameter β; functions $b(\cdot)$ and $c(\cdot,\cdot)$ determine a particular distribution in the exponential family; and $a(\tau) = \tau/w_i$, with the dispersion parameter τ and a known weight w_i. We further assume that the MDM is a logistic regression

$$P(G_i = 1|s_i, y_i) \quad = \quad h(\gamma_0^T s_i + \gamma_1 y_i) = \frac{1}{1+\exp\left[-(\gamma_0^T s_i + \gamma_1 y_i)\right]}, \tag{4}$$

where $G_i = 0(1)$ if the ith observation is observed(missing), and s_i includes a set of observed predictors. Throughout this paper and in the **isni** package, we assume the inverse logit form for $h(\cdot)$, as it is popular and robust, and it simplifies interpretation (Xie and Heitjan 2009). Following Troxel *et al.* (2004), under independence over units i we have

$$\nabla^2 L_{\theta,\theta} \quad = \quad \frac{\partial^2 L(\theta,\gamma_0,\gamma_1)}{\partial\theta\partial\theta^T}\bigg|_{\hat\theta(0),\hat\gamma_0(0),\gamma_1=0} = \sum_{i:g_i=1} \frac{\partial \ln f_{\theta_y}(y_i^{\text{obs}}|x_i)}{\partial\theta\partial\theta^T}\bigg|_{\gamma_1=0},$$

$$\nabla^2 L_{\theta,\gamma_1} \quad = \quad \frac{\partial^2 L(\theta,\gamma_0,\gamma_1)}{\partial\theta\partial\gamma_1^T}\bigg|_{\hat\theta(0),\hat\gamma_0(0),\gamma_1=0} = \sum_{i:g_i=0} (1-h_i)\frac{\partial \text{E}(Y_i^{\text{mis}}|x_i)}{\partial\theta}\bigg|_{\gamma_1=0},$$

and ISNI for $\theta = (\beta,\tau)$ is

$$\text{ISNI} \quad = \quad \begin{pmatrix} -\left[\sum_i g_i\left(y_i\frac{\partial^2\Psi_i}{\partial\beta^2} - \frac{\partial^2 b}{\partial\beta^2}\right)\right]^{-1} a(\tau)\sum_i(1-g_i)(1-h_i)\frac{\partial^2 b}{\partial\Psi_i\partial\beta} \\ 0 \end{pmatrix}_{\hat\beta(0),\hat\tau(0)}, \tag{5}$$

where $h_i = h(\hat\gamma_0^T s_i)$ is the predicted probability of being missing under MAR, and $\hat\beta(0)$ and $\hat\tau(0)$ are MLEs under MAR. This formula is the same as that presented in Troxel *et al.* (2004) except that we reverse the signs to reflect the reversed role of the indicator G.

2.2. ISNIs for Longitudinal/clustered data

We now consider the case where Y_i, the datum for unit i, consists of a vector of potentially correlated observations $Y_{i1}, \cdots, Y_{in_i}$, and distinct units are independent. Several authors have described generalizations of ISNI to longitudinal/clustered data (Ma *et al.* 2005; Xie and Heitjan 2009; Xie 2008, 2009, 2012; Xie and Qian 2012). We describe below the ISNI method for the setting of longitudinal data with non-monotone missingness; one can adapt it to other types of clustered data in a straightforward way.

2.2.1. Models for the notional complete data

Let $Y_i = (Y_{i1}, ..., Y_{in_i})$ denote the notional complete data for subject i, where Y_{ij} is the outcome at measurement occasion j, $i = 1, ..., N$, $j = 1, ..., n_i$. We assume that the density function of Y_i is $f_\theta(y_i|x_i)$, where θ is the vector of the parameters of interest with length p_θ, and x_i is a matrix of fully observed predictors. We describe below the ISNI analysis of two popular classes of models for data of this form.

Marginal multivariate Gaussian model (MMGM). We define the MMGM for a continuous outcome as

$$Y_i|x_i \overset{\text{ind}}{\sim} \text{MVN}\left(x_i\theta_1, \Sigma_i(\theta_2)\right), \tag{6}$$

where θ_1 and θ_2 are parameters of the population mean model and the variance-covariance model, respectively. The matrix $\Sigma_i(\theta_2)$ must be symmetric and positive definite. Under ignorable missingness, one typically estimates this model by generalized least-squares, for example with the R function `gls()`.

Linear mixed model (LMM). The LMM for a continuous outcome is (Laird and Ware 1982)

$$Y_i|b_i, x_i, z_i \overset{\text{ind}}{\sim} \text{MVN}(x_i\beta + z_i b_i, \Lambda_i), \quad b_i \sim N(0, V_b). \tag{7}$$

Here β is a vector of p fixed population parameters; b_i is a vector of q random effects associated with individual i; x_i and z_i are predictor matrices for the fixed and random effects, respectively, where z_i is a subset of x_i; and V_b and Λ_i are variance-covariance matrices for the random effects and residuals, respectively. Λ_i depends on i only in that its size is $n_i \times n_i$. Marginally,

$$Y_i|x_i, z_i \overset{\text{ind}}{\sim} \text{MVN}\left(x_i\beta, \Lambda_i + z_i V_b z_i^T\right).$$

We set $\theta = (\beta, D)$ where D denotes the parameters in the variance-covariance matrices Λ_i and V_b. One can estimate the LMM using R function `lme()`.

2.2.2. An MDM for non-monotone missing data

In longitudinal studies, one typically experiences two types of missingness: Intermittently missing observations, for example from missed visits; and dropout, from subjects who leave the study permanently before completing follow-up. We therefore define an MDM that allows for both types of missingness by means of a general transitional model (Xie 2012). Let $G_i = (G_{i1}, \ldots, G_{in_i})$ denote the vector of missingness status variables for subject i, where $G_{ij}, j = 1, \ldots, n_i$ denotes the missingness status of subject i at occasion j, and

$$G_{ij} = \begin{cases} O & \text{if subject } i \text{ is observed at occasion } j, \\ I & \text{if subject } i \text{ is intermittently missing at occasion } j, \\ D & \text{if subject } i \text{ drops out at occasion } j. \end{cases}$$

Xie (2012) described an approach that writes the MDM as a product of transition probabilities:

$$f_\gamma(g_{i1},\ldots,g_{in_i}|y_i,x_i) = f(g_{i1}|y_i,x_i)\prod_{j=2}^{n_i} f_\gamma(g_{ij}|g_{i1},\ldots,g_{i,j-1},y_i,x_i).$$

We typically assume that all units are observed at baseline and so $f(O|y_i,x_i) = 1$. One can then model each of the remaining univariate conditional distributions separately. We first note that each conditional probability potentially includes all past missingness status variables, thereby naturally incorporating information on how past missingness behaviors affect current missingness status. Second, one needs to decide what variables in Y_i enter each conditional distribution. Let $s_{ij} = (y^{\text{obs}}_{i(j)}, x_i)$ be a matrix containing fully observed predictors for missingness up to visit j for subject i, where $y^{\text{obs}}_{i(j)}$ includes all observed outcome measurements prior to visit j for subject i. Because we condition on all the past missingness status variables within each past missingness pattern, $y^{\text{obs}}_{i(j)}$ can include all past observed outcomes. If any future outcome, y_{iJ}, where $J > j$, is observed for all the subjects with the same past missingness pattern, y_{iJ} can also be included in $y^{\text{obs}}_{i(j)}$ for those observations with that same past missingness pattern.

We further let a numerical variable u index the status of missingness with $u = 0, 1, 2$ representing O, I, D, respectively. Our MDM then assumes that the conditional distribution of G_{ij} is as follows:

$$\left[G_{ij}|S_{ij} = s_{ij}, Y_{ij} = y_{ij}, G_{i(j)} = g_{i,(j)}\right] \sim \text{Multinomial}\left(1, \left[P_{ij}^{0g_{i(j)}}, P_{ij}^{1g_{i(j)}}, P_{ij}^{2g_{i(j)}}\right]\right), \qquad (8)$$

where $g_{i(j)} = (g_{i,j-1}\cdots, g_{i1})$ denotes the past missingness pattern prior to visit j, and the cell probabilities $\left[P_{ij}^{0g_{i(j)}}, P_{ij}^{1g_{i(j)}}, P_{ij}^{2g_{i(j)}}\right]$ are specified as

$$\begin{aligned} P_{ij}^{ug_{i(j)}} &= \frac{\phi_{ij}^{ug_{(j)}}}{\sum_{U=0}^{2}\phi_{ij}^{Ug_{(j)}}}, \quad u = 0,1,2; \\ \text{and } \phi_{ij}^{ug_{(j)}}(s_{ij}, y_{ij}) &= \exp\left(\gamma_0^{ug_{(j)}} s_{ij} + \gamma_1^{ug_{(j)}} y_{ij}\right). \end{aligned} \qquad (9)$$

When $\gamma_1^{ug_{(j)}} = 0$ for all values of u and $g_{(j)}$, the MDM does not depend on the potentially unobserved outcome and thus is MAR. When $\gamma_1^{ug_{(j)}} \neq 0$ for some u and $g_{(j)}$, the model allows that, given the missingness pattern prior to time j and the other fully observed predictor s_{ij}, the missingness status depends on the potentially unobserved outcomes through the contemporaneous outcome Y_{ij}. An alternative MDM specification would let the missingness depend on both past and future unobserved outcomes. We chose the former model for two reasons: First, it reduces the number of parameters for nonignorable missingness, allowing us to more easily interpret the sensitivity analysis. As others (e.g., Vansteelandt, Rotnitzky, and Robins (2007)) have noted, parsimony is desirable in sensitivity analysis. Second, one can always take the integration of the latter model with respect to past and future unobserved outcomes so that the resulting selection model depends only on the outcome at the current visit. In this sense, our model can be viewed as an approximation to the model in which the probability of missingness depends also on past and future unobserved outcomes; see Xie (2012).

Xie (2012) applied the transitional MDM to a longitudinal dataset with non-monotone missingness arising from a design with $n_i = n$; this enables conditioning on the entire past missingness vector $g_{i(j)}$. In the setting with varying n_i, conditioning on $g_{i(j)}$ may be impossible. In this case, a convenient alternative is a first-order transition model, where one assumes that, conditional on $(s_{ij}, y_{ij}, g_{i,j-1})$, the missingness status G_{ij} at the current visit is independent of missingness status at all other prior visits; thus Eqn (8) reduces to

$$\begin{aligned} G_{ij}|S_{ij} = s_{ij}, Y_{ij} = y_{ij}, G_{i,j-1} = v &\sim \text{Multinomial}(1, [P_{ij}^{0v}, P_{ij}^{1v}, P_{ij}^{2v}]); \\ P_{ij}^{uv} &= \frac{\phi_{ij}^{uv}}{\sum_{U=0}^{2} \phi_{ij}^{Uv}}, \quad \text{where } u = 0,1,2; v = 0,1, \\ \text{and } \phi_{ij}^{uv}(s_{ij}, y_{ij}) &= \exp(\gamma_0^{uv} s_{ij} + \gamma_1^{uv} y_{ij}). \end{aligned} \tag{10}$$

By the definition of dropout, $G_{ij} = 2$ deterministically when $G_{i,j-1} = 2$ (the prior visit is a dropout); by the definition of intermittent missingness, $\phi_{ij}^{21} = 0$; and because the response probabilities must add up to unity, for identification purposes $\phi_{ij}^{00} = \phi_{ij}^{01} = 1$. Here s_{ij} is a vector of fully observed predictors for missingness at time j for subject i, which commonly includes the history of the predictors x in the ideal-data model up to and including time j as well as the history of the observed prior outcomes (*i.e.*, the observed elements in $(Y_{i1},...,Y_{i,j-1})$). The conditional probability $f_\gamma(g_{ij}|s_{ij}, y_{ij}, g_{i,j-1})$ then takes the form

$$f_\gamma(g_{ij}|s_{ij}, y_{ij}, g_{i,j-1}) = \begin{cases} \frac{1}{1+\exp(\gamma_0^{10}s_{ij}+\gamma_1^{10}y_{ij})+\exp(\gamma_0^{20}s_{ij}+\gamma_1^{20}y_{ij})} & \text{if } g_{i,j-1} = 0, g_{ij} = 0, \\ \frac{\exp(\gamma_0^{g_{ij}0}s_{ij}+\gamma_1^{g_{ij}0}y_{ij})}{(1+\exp(\gamma_0^{10}s_{ij}+\gamma_1^{10}y_{ij})+\exp(\gamma_0^{20}s_{ij}+\gamma_1^{20}y_{ij})} & \text{if } g_{i,j-1} = 0, g_{ij} \neq 0, \\ \frac{1}{1+\exp(\gamma_0^{11}s_{ij}+\gamma_1^{11}y_{ij})} & \text{if } g_{i,j-1} = 1, g_{ij} = 0, \\ \frac{\exp(\gamma_0^{11}s_{ij}+\gamma_1^{11}y_{ij})}{1+\exp(\gamma_0^{11}s_{ij}+\gamma_1^{11}y_{ij})} & \text{if } g_{i,j-1} = 1, g_{ij} = 1, \\ 0 & \text{if } g_{i,j-1} = 1, g_{ij} = 2, \\ 0 & \text{if } g_{i,j-1} = 2, g_{ij} \neq 2, \\ 1 & \text{if } g_{i,j-1} = 2, g_{ij} = 2. \end{cases} \tag{11}$$

The package **isni** allows computation of ISNI statistics under both the general transitional MDM and the simpler first-order transitional MDM.

2.2.3. *ISNI with the transitional MDM*

In general, we have

$$\text{ISNI}(\hat{\theta}) = \left. \frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_1^T} \right|_{\gamma_1=0} = -\nabla^2 L_{\theta,\theta}^{-1} \nabla^2 L_{\theta,\gamma_1}.$$

The term $\nabla^2 L_{\theta,\theta}$ is the observed Hessian matrix under the MAR model and can be readily obtained as a by-product of estimating the MAR model. The second term, $\nabla^2 L_{\theta,\gamma_1}$, measures the lack of orthogonality of θ and γ_1 and has been derived for longitudinal data with monotone missingness patterns (Ma *et al.* 2005) and with non-monotone missingness patterns (Xie 2012; Xie and Qian 2012). We include its derivation in Appendix B. For our longitudinal models with the first-order transitional MDM described above, we have $\gamma_1 = (\gamma_1^{10}, \gamma_1^{20}, \gamma_1^{11})$ and

$$\begin{aligned} \nabla^2 L_{\theta,\gamma_1} &= \left. \frac{\partial^2 L}{\partial\theta\partial\gamma_1^T} \right|_{\hat{\theta}(0),\hat{\gamma}_0(0),\gamma_1=0} \\ &= \left. \left(\frac{\partial^2 L}{\partial\theta\partial\gamma_1^{10}}, \frac{\partial^2 L}{\partial\theta\partial\gamma_1^{20}}, \frac{\partial^2 L}{\partial\theta\partial\gamma_1^{11}} \right) \right|_{\hat{\theta}(0),\hat{\gamma}_0(0),\gamma_1=0} \\ &= \sum_{i:K_i<n_i} \left. \frac{\partial \mathrm{E}((Y_i^{\mathrm{mis}})^T | y_i^{\mathrm{obs}})}{\partial\theta} \right|_{\gamma_1=0} \cdot \left[A_i^{10}, A_i^{20}, A_i^{11} \right], \end{aligned} \tag{12}$$

K_i is the length of y_i^{obs} and $\hat{\theta}(0)$ and $\hat{\gamma}_0(0)$ are MLEs of θ and γ_0 under the MAR assumption. In $\nabla^2 L_{\theta,\gamma_1}$, $\partial\mathrm{E}((Y_i^{\mathrm{mis}})^T|y_i^{\mathrm{obs}})/\partial\theta$ is a $p_\theta \times d_i$ matrix where d_i is the number of missing outcomes for subject i. Under MAR and for the complete-outcome models considered in Section 2.2.1, $(Y_i^{\mathrm{mis}})^T|y_i^{\mathrm{obs}}$ is a Gaussian distribution involving θ parameters only and $\partial\mathrm{E}((Y_i^{\mathrm{mis}})^T|y_i^{\mathrm{obs}})/\partial\theta$ can be derived in a closed form involving matrix multiplication and differentiation as shown in Appendix C.

Finally, $A_i = \left[A_i^{10}, A_i^{20}, A_i^{11}\right]$ is a $d_i \times 3$ matrix. For the lth $(l = 1, ..., d_i)$ component of y_i^{mis} that corresponds to the jth element of y_i, it can be shown that (Xie 2012; Xie and Qian 2012)

$$\begin{aligned} A_{il}^{10} &= I(g_{i,j-1} = 0) \left[I(g_{i,j} = 1) - P_{ij}^{10} \right]_{\hat{\gamma}_0(0),\gamma_1=0}, \\ A_{il}^{20} &= I(g_{i,j-1} = 0) \left[I(g_{i,j} = 2) - P_{ij}^{20} \right]_{\hat{\gamma}_0(0),\gamma_1=0}, \\ A_{il}^{11} &= I(g_{i,j-1} = 1) \left[P_{ij}^{01} \right]_{\hat{\gamma}_0(0),\gamma_1=0}. \end{aligned}$$

In the above, P_{ij}^{10}, P_{ij}^{20} and P_{ij}^{01} are the missingness status transitional probabilities defined in Eqn (10), calculated under MAR (i.e., setting $\gamma_1 = 0$). We note that because dropout and intermittent missingness are competing causes of missingness, the quantities, A_{il}^{10} and A_{il}^{20} are of opposite sign given that $g_{i,j-1} = 0$.

When γ_1 is a scalar, ISNI approximates the changes of the estimates when γ_1 is perturbed from 0 to 1. In our model, the nonignorability parameter γ_1 is the vector with $(\gamma_1^{10}, \gamma_1^{20}, \gamma_1^{11})$. Thus, ISNI is a vector of three elements where each approximates the changes in estimates when only the corresponding element in γ_1 is perturbed from 0 to 1. We suggest the following strategies to produce a parsimonious sensitivity analysis with multiple nonignorability parameters. To summarize the joint effects of all three nonignorable parameters, one can approximate the change in the MAR estimates, $\hat{\theta}(\gamma_1) - \hat{\theta}(0)$ by $\left.\frac{\partial\hat{\theta}(\gamma_1)}{\partial\gamma_1^T}\right|_{\gamma_1=0} \gamma_1$. When one is willing to assume that intermittent missingness and dropout have roughly the same nonignorability mechanism, we can fix $\gamma_1^{10}, \gamma_1^{20}$ and γ_1^{11} at the same largest perturbation value. In this scenario, γ_1 becomes a scalar and A_i become a vector where $A_i = A_i^{10} + A_i^{20} + A_i^{11}$ and the components of A_i become $P_{ij}^{0,g_{i,j-1}}$, the estimated probability of being observed for the missing observations predicted under the ignorability assumption. For its simplicity and ease of interpretation, this is the default method implemented in our package.

Alternatively, one could consider all perturbations of the elements of γ_1 that are within a hypercube of size 1 from the origin in the space. These perturbations include scenarios under which dropout and intermittent missingness can have different or even opposite nonignorable missingness mechanisms. For this strategy, we suggest the following way to examine the sensitivity using the ISNI vector associated with multiple nonignorability parameters:

$$\mathrm{MISNI}(\hat{\theta}) \quad = \quad \sum_{i=1}^{q} \left| \frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_{1i}} \right|_{\hat{\theta}(0),\hat{\gamma}_0(0),\gamma_1=0}, \tag{13}$$

where $q = 3$ is the length of γ_1. MISNI, as defined in the above, has the interpretation of maximum sensitivity, $\max_{|\gamma_{1i}|=1,i=1,\ldots,q}(\hat{\theta}(\gamma_1) - \hat{\theta}(0))$, when each element of γ_1 is perturbed between -1 and 1. Our package also produces MISNI. Xie and Heitjan (2004) consider a perturbation scheme when the nonignorability parameter lies on a hyperball of radius $\sqrt{q}$ around the origin, that is, $||\gamma_1|| = \sqrt{q}$. A potential issue with this scheme is that elements in the nonignorability parameter vector can have extreme and implausible values, especially when q, the dimension of the nonignorability parameter, is large. An alternative is to consider a hyperball of size that is independent of q (Gao *et al.* 2016). We note that a hypercube of size r as used here contains a hyperball of radius r.

One can extend these strategies to the general transitional MDM. Recall that under the general model of Eqn (9), the nonignorability parameter $\gamma_1 = \bigcup_{j,g_{i(j-1)}}(\gamma_1^{10g_{i(j-1)}}, \gamma_1^{20g_{i(j-1)}}, \gamma_1^{11g_{i(j-1)}})$, and the dimensionality of γ_1 depends on the number of unique past missing data patterns and can be large. For a more parsimonious sensitivity analysis, one sensible strategy is to reduce the number of nonignorability parameters such that $\gamma_1^{10g_{i(j-1)}} = \gamma_1^{10}, \gamma_1^{20g_{i(j-1)}} = \gamma_1^{20}$ and $\gamma_1^{11g_{i(j-1)}} = \gamma_1^{11}, \forall j$ and $g_{i(j-1)}$. This reduces γ_1 to $(\gamma_1^{10}, \gamma_1^{20}, \gamma_1^{11})$, and we can then use the strategies described for the first-order transitional model.

We note moreover that in the special case of no intermittent missingness, the equations further simplify: $A_i^{10} = A_i^{11} = 0$ and $\partial \mathrm{E}((Y_i^{\mathrm{mis}})^T | y_i^{\mathrm{obs}})/\partial\theta$ reduces to a $p_\theta \times 1$ vector. The calculation of $\nabla^2 L_{\theta,\gamma_1}$ then reduces to that of Ma *et al.* (2005) and Xie (2008).

2.3. Calibrating ISNI

With a logit link in the MDM, a scalar γ_1 is the log odds ratio in the probability of being missing associated with a one-unit change in y; when $\gamma_1 = 1$, a one-unit change in y is associated with an increase of 2.7-fold in the odds of being missing. For outcomes with a single natural scale, such as the Poisson and binomial, one can interpret ISNI directly in this manner. For continuous Y, this interpretation is inadequate because the value of ISNI depends on the scale of measurement, which may be arbitrary. We describe below two calibration approaches that facilitate interpretation by creating a scale-free index.

The first approach evaluates changes in estimates of θ for a magnitude of nonignorability where a one-SD change in Y is associated with an odds ratio of $e^1 = 2.7$ in the probability of being missing, i.e., when $\gamma_1 = \pm 1/\sigma_Y$, or one standardized unit of nonignorability.

The second approach is to approximate the minimum standardized magnitude of nonignorability that is needed for the change in $\hat{\theta}$ to equal one standard error (SE). One can then assess sensitivity by evaluating whether this level of nonignorability is plausible. Specifically, note that for θ^j (the jth element of θ) we have $\hat{\theta}^j(\gamma_1) - \hat{\theta}^j(0) \approx \gamma_1 \mathrm{ISNI}_Y(\hat{\theta}^j)$, where $\mathrm{ISNI}_Y(\cdot)$ refers to the ISNI for a parameter, computed with data on the Y scale. We compute the smallest absolute value of γ_1 that gives a 1-SE change as $\widetilde{\gamma}_1 = \frac{\mathrm{SE}}{\mathrm{ISNI}}$. To put the magnitude of nonignorability in the scale of standardized magnitude of nonignorability defined above, we define a sensitivity transformation c statistic as

$$c = |\widetilde{\gamma}_1 \times \sigma_Y| = \left| \frac{\sigma_Y \mathrm{SE}}{\mathrm{ISNI}} \right|. \tag{14}$$

The c statistic informs us that in order for selection bias to be as large as the sampling error, the magnitude of nonignorability needs to be at least as large as that with which $\frac{1}{c}$SD change in Y is associated with an odds ratio of 2.7 in the probability of being missing. For MISNI from Eqn (13), $c = \left|\frac{\sigma_Y \mathrm{SE}}{\mathrm{MISNI}}\right|$. The MISNI c statistic informs us that for selection bias to be as large as the sampling error, the vector nonignorability parameter γ_1 needs to lie on a hypercube with its size from the origin at least as large as c standardized unit of nonignorability as defined above; the hypercube of this size implies that $\frac{1}{c}$SD change in Y is associated with an odds ratio of 2.7 for dropout (or intermittent missingness) versus being observed.

When c is large, only extreme nonignorability can make the estimate change substantially, and consequently sensitivity to nonignorability is of little concern. For example, $c = 10$ implies that in order for the error in an MAR estimate to be the same size as its sampling error, the nonignorability needs to be strong enough that a 0.1-SD change in Y causes a significant change in the odds of being missing. When c is small, modest departure from MAR can cause the estimate to change substantially. For example, $c = 0.1$ implies that when even a 10-SD change in Y causes a significant change in the odds of being missing, the estimate may change substantially. As such a degree of nonignorability is plausible in many applications, this small c value signals sensitivity. Following (Troxel *et al.* 2004), we suggest using $c < 1$ as a rule of thumb to signal significant sensitivity.

Because ISNI is a derivative, we have in general that $\mathrm{ISNI}(A \cdot \theta) = A \cdot \mathrm{ISNI}(\theta)$ for any matrix A that has the same number of columns as there are elements in θ (where $\cdot$ here represents matrix multiplication). Thus, we can readily conduct a separate ISNI analysis (including computation of c) for any element of θ.

3. Program Description and Usage

The package **isni** implements the methods described above. In the following subsections, we describe the use of the three main functions: `isniglm()`, `isnimgm()`, and `isnilmm()`.

3.1. Interface of `ISNI` functions

The three primary ISNI functions share a common interface structure that permits specification of the complete-data model for y, the MDM, and the data.

We begin with function `isniglm()`, which computes ISNI for the GLM for a univariate y:

```
isniglm = function(formula,  family=gaussian, data, weights,  subset,
                   start=NULL,  offset)
```

formula an object of model formulas: at a minimum a two-sided formula that specifies the complete-data model using the variable names for the outcome y_i and the predictors x_i in Eqn (3). Alternatively one can specify a two-equation model formula that additionally specifies the MDM using the variable name for the missingness indicator g_i and the missingness predictor s_i in Eqn (4). More details of model specification are given under the subsection 3.2.

family a description of the error distribution to be used in the GLM for the outcome y in Eqn (3).

data the name of the data frame containing the model variables; the `y` element must include both the non-missing and missing observations, the latter indicated by `NA`.

weights an optional vector of "prior weights" to be used in the fitting process for the complete-data model and the missing data mechanism model.

subset an optional vector specifying a subset of observations to be used in the fitting process for the outcome model and the missing data mechanism model.

start starting values for the parameters in the linear predictor of the outcome model.

offset an optional vector to specify an a priori known component to be included in the linear predictor during fitting the GLM for the outcome. This should be NULL or a numeric vector of length equal to the number of observations.

Function `isnimgm()` computes ISNI for the MMGM:

```
isnimgm = function(formula, data, cortype="CS", id, subset, weights,
                   predprobobs, misni=FALSE)
```

formula an object of model formulas: at a minimum a two-sided formula that specifies the complete-data model using the variable names for the outcome y_i and the predictors x_i in Eqn (6). Alternatively one can specify a two-equation model formula that additionally specifies the MDM using the variable names for the missing status variable g_{ij} and the missingness predictor s_{ij} in Eqn (10). More details of model specification are given under the subsection 3.2.

data the name of the data frame containing all the variables in the model; the `y` element must include both the non-missing and missing observations, the latter indicated by `NA`.

cortype a description of the within-subject correlation structure of Σ_i in Eqn (6).

id the name of variable for the level-2 clustering variable.

subset an optional vector specifying a subset of observations to be used in the fitting process for the outcome model and the missing data mechanism model.

weights frequency weights to be assigned to each `id`. When supplied, indicates differential weights are used; otherwise each `id` is weighted equally.

predprobobs `NULL` if using the built-in multinomial first-order transitional logistic model to obtain predicted probabilities of being observed; otherwise, the user can supply the name of the variable in `data` that gives these probabilities *via* this argument.

misni `FALSE` if using the default approach to computing ISNI with a scalar nonignorability parameter; `TRUE` when computing ISNI with multiple nonignorability parameters.

Function `isnilmm()` computes ISNI for the LMM:

```
isnilmm = function(formula, data, random, id, weights, subset, predprobobs,
                   misni=FALSE)
```

formula an object of model formulas: at a minimum a two-sided formula that specifies the complete-data model using the variable names for the outcome y_i and the predictors x_i in Eqn (7). Alternatively one can specify a two-equation model formula that additionally specifies the MDM using the variable names for the missing status variable g_{ij} and the missingness predictor s_{ij} in Eqn (10). More details of model specification are given under the subsection 3.2.

data the name of the data frame containing all the variables in the model; the `y` element must include both the non-missing and missing observations, the latter indicated by `NA`.

random a one-sided formula that specifies the random-effects part of the linear mixed-effects complete-data model using the variables names for z_i in Eqn (7).

id the name of the level-2 clustering variable.

subset an optional vector specifying a subset of observations to be used in the fitting process for the complete-data model and the missing data mechanism model.

weights frequency weights to be assigned to each `id`. When supplied, indicates differential weights are used; otherwise each `id` is weighted equally.

predprobobs `NULL` if using the built-in multinomial first-order transitional logistic model to obtain predicted probabilities of being observed; otherwise, the user can supply the name of the variable in `data` that gives these probabilities *via* this argument.

misni `FALSE` if using the default approach to computing ISNI with a scalar nonignorability parameter; `TRUE` when computing ISNI with multiple nonignorability parameters.

3.2. Model Specification

The ISNI analysis is based on a joint selection model and requires specifying variables in two model equations: the complete-data model and the missing data mechanism model. The model specification is achieved primarily via the `formula` argument for the above three functions. At a minimum, the user should supply a single-equation for the complete-data model in the typical form: `response` $\sim$ `Xterms` where `response` is the (numeric or factor) vector for the outcome of interest and `Xterms` is a series of terms, separated by + operators, which specify a linear predictor for response. With the single-equation specification, the `isniglm` function will by default use (`is.na(response)`) as the missingness indicator g_i and `Xterms` as the missingness predictor s_i in Eqn (4). For longitudinal data setting, both `isnimgm` and `isnilmm` will by default use the utility function `definemissingstatus` provided in the package to generate the missingness status variables g_{ij} and then use `Xterms` as the predictors s_i for fitting a first-order transitional missing data mechanism model in Eqn (10). In this case, it is important to sort beforehand within-`id` observations by time so that the missingness status variable can be defined correctly. The ISNI functions then compute the MAR estimates and conduct ISNI computation to evaluate the rate of change of model estimates in the neighborhood of the MAR model and the missingness probability is allowed to depend on the unobserved value of `response`, even after conditioning on the other missingness predictors in s_i.

The above single-equation model specification assumes x_i in the complete-data model and s_i in the missing data mechanism model use the same set of predictors. To use different sets of predictors, one can explicitly specify a two-equation formula. For this purpose, the specification and processing of the "formula" argument in the above three ISNI functions make use of the R package "Formula" designed for handling model equations with multiple responses and multiple sets of predictors (Zeileis and Croissant 2010). For `isniglm`, this can be specified as: `response | is.na(response) ~ Xterms | Sterms`, which specifies the formula for the complete-data model as `response ~ Xterms` and that `is.na(response)` and `Sterms` are the missingness indicator g_i and the missingness predictor s_i in the missing data mechanism model as specified in Eqn (4), and `Xterms` and `Sterms` can be different. For both `isnimgm` and `isnilmm`, this can be specified as: `response | miss + missprior ~ Xterms | Sterms`, which specifies the formula for the outcome model as `response ~ Xterms` and that in the missing data mechanism model as specified in Eqn (10), `miss` and `missprior` are the variable names in `data` denoting the missingness status at the current visit and at the prior visit, respectively, and `Sterms` are the missingness predictor s_i, and `Xterms` and `Sterms` can be different,

For `isnilmm`, `response ~ Xterms` specifies the fixed-effect part of the linear mixed-effects model for the outcome. The random-effect part of the model is specified as a one-sided formula via the argument `random`.

3.3. Preparing Data for ISNI analysis: Format of `data`

The ISNI functions use the argument `data` to supply the input data frame. The user must observe two rules when preparing the data for input: First, except for the missingness status variables, the columns in the master dataset should include all the variables (i.e., response and explanatory variables) in both the complete-data model and MDM. For convenience, users can rely on the ISNI functions to generate the missingness status variables automatically inside the functions. For independent data, the missingness status variable is simply an indicator variable for missingness and is generated as `is.na(response)` inside `isniglm` if not provided via the `formula` argument. For longitudinal data, the missingness status variable G_{ij} can have three categories: "O" (being observed), "I" (intermittently missingness) and "D" (dropout). Users can define and supply the missingness status variables at the current and the prior visit as two separate columns in `data` and pass them into function via the `formula` argument to fit a first-order transitional MDM. The package **isni** provides a utility function `definemissingstatus` to generate these two missingness status variables for users. If users do not supply missing status variables, the `isnimgm` and `isnilmm` will call `definemissingstatus` inside these functions to generate them. Again in this case, it is important to sort beforehand within-`id` observations by time so that the missingness status variable can be defined correctly.

Second, as shown in the ISNI formula, observations with missing outcomes contribute to the sensitivity analysis through $\nabla^2 L_{\theta,\gamma_1}$. Thus, the master dataset must include places for all the planned observations, present or missing. This differs from standard ignorable analysis, in which one simply omits the missing observations. An exception to the second point occurs in the case of dropout with longitudinal data; we will explain this in the **Longitudinal/clustered data** section below.

Here we describe the data format in greater detail.

Independent data. We array the data in a rectangular matrix, where each row is an independent unit. The fields include the main outcome of interest with the missing values denoted as "NA", the predictors x for the outcome model, the predictors s in the MDM, and optionally the missingness indicator variable.

Longitudinal/clustered data. The data consist of multiple level-1 observations (e.g., repeated measures) within a level-2 unit. The name of the level-2 variable is passed into function via the argument `id`. The level-1 observations for a level-2 unit must appear on different rows and take up as many records as the number of planned measurements for that level-2 unit. Missing values in the outcome are denoted "`NA`".

As mentioned above, in principle all planned observations should appear in the master dataset. An exception occurs for dropouts in longitudinal data. When a subject leaves the study and never returns, the probability of observing outcomes becomes 0 for all subsequent times. Thus all the visits after the dropout visit are missing ignorably and do not contribute to the sensitivity analysis; one can therefore omit them from the database.

The fields in the master dataset for longitudinal/clustered data consist of the level-2 variable, the dependent variable y, the predictors x for the outcome model and s for the MDM, and optionally the variables for missingness status at the current and the prior visit; the order is irrelevant. Both models include an intercept by default; there is no need to include a column of ones.

If users supply the missingness status variable in the master dataset, it should be of a character or factor, with permitted values "O"(observed), "I" (intermittently missing), and "D" (dropout). Because the built-in MDM employs a first-order Markov model that depends on the previous missingness status, users should also supply a variable that denotes missingness status at the prior visit; this is a character/factor variable taking values of "O", "I", "D", or "U" for the baseline observation (which has no prior visit).

4. Examples

In this section we describe some applications of the **isni** package.

4.1. ISNI analysis of a GLM for a cross-sectional survey

Raab and Donelly (1999) analyzed a cross-sectional survey of sexual practices among students at the University of Edinburgh. The response variable is the students' answer to the question "Have you ever had sexual intercourse?". Because of the sensitivity of this question, many students declined to answer, leading to substantial missing data. We consider a simplified data set consisting of the answer to this question, with the student's sex and faculty as predictors.

```
# load the library and data set
> library(isni)
```

```
> data(sos)
> sos[sample(nrow(sos),10),]
     sexact gender faculty
2845   <NA>   male   other
1249    yes   male   other
5406   <NA> female   other
3132   <NA>   male     mdv
225     yes   male   other
5812   <NA> female   other
5115   <NA> female   other
769     yes   male   other
2366   <NA>   male   other
2622   <NA>   male   other
```

The `R` code above loads the library **isni** and the data frame `sos`, displaying a random subsample of 10 records. `sos` includes the following factor variables: `sexact` is the response to the question "Have you ever had sexual intercourse?" (two levels: no (reference level), yes); `gender` is the student's sex (two levels: male (reference level), female); `faculty` is the student's faculty (medical/dental/veterinary, all other faculty categories (reference level)).

Assuming ignorability, one can fit a logistic model (using responders only) to predict the outcome by sex, faculty and their interaction. We estimated the model with function `glm()`:

```
> ymodel= sexact  ~ gender*faculty
> summary(glm(ymodel,family=binomial, data=sos))

Call:
glm(formula = ymodel, family = binomial, data = sos)

Deviance Residuals:
Min       1Q   Median       3Q      Max
-1.6713  -1.3282   0.7540   0.7642   1.0338

Coefficients:
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)              1.08153    0.05561  19.448  < 2e-16 ***
genderfemale             0.03081    0.07958   0.387    0.699
facultymdv              -0.73389    0.14921  -4.918 8.73e-07 ***
genderfemale:facultymdv  0.10213    0.20670   0.494    0.621
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4450.3  on 3827  degrees of freedom
Residual deviance: 4408.2  on 3824  degrees of freedom
AIC: 4416.2
```

```
Number of Fisher Scoring iterations: 4
```

The estimates show that students in a medical faculty were less likely to report having had sexual intercourse. Because only 62.4% responded to the sexual practice question, there is concern that this analysis is sensitive to the assumption of ignorability. Thus Troxel *et al.* (2004) conducted an ISNI analysis for this model. Below we replicate their analysis with the function `isniglm()`. We posit a nonignorable selection in the form of Eqn (4) with the observed missingness predictor s_i including `gender`, `faculty` and their interaction and perform ISNI analysis as follows:

```
> sos.isni<-isniglm(ymodel, family=binomial, data=sos)
> sos.isni

Call:
isniglm(formula = ymodel, family = binomial, data = sos)

ISNIs:
(Intercept)        genderfemale      facultymdv  genderfemale:facultymdv
0.410141           -0.038983         -0.169859   0.027542

c statistics:
(Intercept)        genderfemale      facultymdv  genderfemale:facultymdv
0.13559            2.04146           0.87846     7.50482
```

The `summary` function in the package expresses the `isniglm()` object:

```
> summary(sos.isni)

Call:
isniglm(formula = ymodel, family = binomial, data = sos)

                         MAR Est.  Std. Err      ISNI      c
(Intercept)              1.081531  0.055611  0.410141 0.1356
genderfemale             0.030808  0.079583 -0.038983 2.0415
facultymdv              -0.733886  0.149215 -0.169859 0.8785
genderfemale:facultymdv  0.102133  0.206696  0.027542 7.5048
```

The columns "MAR Est." and "Std. Err" denote the logistic model estimates and their standard errors under MAR; "ISNI" and "c" denote ISNI values and c statistics. The ISNIs are equal in absolute values to those reported in Troxel *et al.* (2004), but with opposite signs because our package models the probability that an observation is missing rather than the probability that it is observed (Troxel *et al.* 2004). Recall that ISNI denotes the approximate change in the MLEs when γ_1 in the selection model is changed from 0 to 1. Under our nonignorable selection model, assuming that $\gamma_1 = 1$ means that a student whose answer is "yes" has an increase of 2.7-fold in the odds of nonresponse. Thus, subjects whose true value is "yes" would be more likely to have a missing value, and the naïve MAR estimate for `(Intercept)` should be less than the `(Intercept)` estimate under the correct nonignorable model. The positive sign of the ISNI value for `(Intercept)` is consistent with this prediction. The ISNI for the `faculty` predictor is -0.17, indicating that if, as is more plausible here, $\gamma_1 = 1$, the MLE for the estimate should change from -0.73 to -0.90. If $\gamma_1 = -1$, the estimate would change from -0.73 to -0.56. The c statistics for `(Intercept)` and `faculty` are both less than 1, suggesting that these coefficients are sensitive to nonignorability, confirming the original analyses (Raab and Donelly 1999). Raab and Donnelly also found that neither the `gender` nor the interaction term between `gender` and `faculty` should be sensitive, as our findings confirm.

In the above we do not explicitly specify an MDM via `formula` argument. The same analysis can be replicated by explicitly specifying an MDM model using a two-equation model formula. The two-equation formula `sexact | is.na(sexact) ~ gender*faculty | gender *faculty` uses the operator `|` to separately specify variables used in the complete-data model and MDM. The two-equation formula means that the complete-data model is `sexact` $\sim$ `gender*faculty` and that `is.na(sexact)` and `gender*faculty` are the missingness indicator g_i and the missingness predictor s_i in the missing data mechanism model as specified in Eqn (4),

```
> ygmodel= sexact | is.na(sexact)  ~ gender*faculty | gender *faculty
> summary(isniglm(ygmodel, family=binomial, data=sos))

Call:
isniglm(formula = ygmodel, family = binomial, data = sos)

                        MAR Est.  Std. Err      ISNI      c
(Intercept)             1.081531  0.055611  0.410141 0.1356
genderfemale            0.030808  0.079583 -0.038983 2.0415
facultymdv             -0.733886  0.149215 -0.169859 0.8785
genderfemale:facultymdv  0.102133  0.206696  0.027542 7.5048
```

Because all the covariates in `sos` are categorical variables, one can also analyze the data using a grouped binomial regression with the `weight` argument in `isniglm` as below.

```
> gender <- c(0,0,1,1,0,0,1,1)
> faculty <- c(0,0,0,0,1,1,1,1)
> gender <- factor(gender, levels = c(0, 1), labels =c("male", "female"))
> faculty <- factor(faculty, levels = c(0, 1), labels =c("other", "mdv"))
> SAcount <- c(NA, 1277, NA, 1247, NA, 126, NA, 152)
> total  <- c(1189,1710,978,1657,68,215,73,246)
> sosgrp <- data.frame(gender=gender, faculty=faculty, SAcount=SAcount, total=total)
> ymodel <- SAcount/total ~gender*faculty
> sosgrp.isni<-isniglm(ymodel, family=binomial, data=sosgrp, weights=total)
> summary(sosgrp.isni)

Call:
isniglm(formula = ymodel, family = binomial, data = sosgrp, weights = total)
```

```
                         MAR Est.  Std. Err      ISNI      c
(Intercept)               1.081531  0.055611  0.410141 0.1356
genderfemale              0.030808  0.079583 -0.038983 2.0415
facultymdv               -0.733886  0.149215 -0.169859 0.8785
genderfemale:facultymdv   0.102133  0.206696  0.027542 7.5048
```

4.2. ISNI analysis of an MMGM for longitudinal data

We consider here the qualify-of-life (QoL) example of Ma *et al.* (2005). The data are from a randomized trial comparing flutamide with a placebo in the treatment of prostate cancer. A sub-study collected QoL outcome data using a SWOG questionnaire at four time points (baseline, 1, 3, and 6 months after randomization). We focus here on the emotional functioning (EF) scale. Below is a sample of the data set `qolef`.

```
> qolef[1:12,]
       id         y time group perf sever        yp g gp     basey
1  117938  8.485281    0     0    0     1  0.000000 O  U  8.485281
2  117938  8.485281    1     0    0     1  8.485281 O  O  8.485281
3  117938        NA    3     0    0     1  8.485281 I  O  8.485281
4  117938  8.717798    6     0    0     1  8.485281 O  I  8.485281
5  124149 10.000000    0     1    0     1  0.000000 O  U 10.000000
6  124149 10.000000    1     1    0     1 10.000000 O  O 10.000000
7  124149  8.000000    3     1    0     1 10.000000 O  O 10.000000
8  124149  9.797959    6     1    0     1  8.000000 O  O 10.000000
9  124674  9.591663    0     0    0     1  0.000000 O  U  9.591663
10 124674  9.380832    1     0    0     1  9.591663 O  O  9.591663
11 124674        NA    3     0    0     1  9.380832 I  O  9.591663
12 124674  9.165151    6     0    0     1  9.380832 O  I  9.591663
```

The variables in `qolef` are as follows:

id — patient id

y — EF score

time — time in months since randomization

group — placebo (0) or flutamide (1).

perf — baseline performance score

sever — baseline disease severity

yp — most recently observed prior outcome

g — missingness status ("O"=observed, "D"=dropout, "I"=intermittently missing)

gp — missingness status in the prior visit ("O"=observed, "D"=dropout, "I"=intermittent missingness, "U"=undefined)

basey — EF at baseline

We seek to evaluate the drug effect on EF over time. The original EF is on a scale of 0 (worst) to 100 (best). Because preliminary analysis showed that normality is more nearly satisfied on the square-root scale, we transformed the data accordingly, and the EF values in the data lie in the range $[0, 10]$. Table 1 presents the mean and SD of the transformed EF variable, together with the fraction observed, by treatment arm and visit. We see that missingness percentages are comparable between arms but increase over time, with almost a quarter missing data by the end of study. Our analysis excludes a small number of patients ($\approx 3\%$) whose EF data were missing at baseline. Table 2 presents missing data patterns for the longitudinal EF outcomes (omitting subjects with missing baseline) and shows that there are both dropouts and intermittently missing items.

Table 1: Summary Statistics in the Prostate Cancer QoL Dataset.

	Placebo ($n = 367$)		Flutamide ($n = 370$)	
Month	Mean (SD)	n observed (%)	Mean (SD)	n observed (%)
0	8.31 (1.50)	352 (96)	8.28 (1.46)	363 (98)
1	8.76 (1.22)	315 (86)	8.54 (1.23)	325 (88)
3	8.83 (1.26)	301 (82)	8.57 (1.20)	313 (85)
6	8.76 (1.20)	274 (75)	8.43 (1.37)	291 (79)

Xie (2012) presents an ISNI analysis of the impact of nonignorable nonmonotone missingness on the MAR estimates for this dataset. Here we conduct an ISNI analysis for the marginal multivariate Gaussian model, using function `isnimgm()`. We take the predictor vector to be

$$\begin{aligned} x_i &= (\text{Intercept}, \text{perf}, \text{sever}, \\ &\quad \text{T1.0(pb)}, \text{T3.0(pb)}, \text{T6.0(pb)}, \\ &\quad \text{T0(fl)} - \text{T0(pb)}, \text{T1.0(fl)} - \text{T1.0(pb)}, \text{T3.0(fl)} - \text{T3.0(pb)}, \text{T6.0(fl)} - \text{T6.0(pb)})_i. \end{aligned}$$

The two predictors "perf" and "sever" are baseline covariates. The predictor T0(a) is an indicator for the baseline observation in arm a. The predictors Tt.0(a) are contrasts of time t *vs.* baseline in arm a, $t = 1, 3, 6$. Thus the third line of predictors gives contrasts of these contrasts between arms. The complete-data model can be specified in R as `y` $\sim$ `perf + sever+ as.factor(time) + group + as.factor(time):group`. To evaluate the robustness of the MAR analysis to nonignorable missingness, we assume the first-order transitional model of Eqn (10), where the missingness status variables at the current visit and at the prior visit are `g` and `gp` in `qolef`, respectively; the missingness predictors s_i is `as.factor(time) * group + yp + perf + sever` and `yp` is the most recently observed outcome prior to the current visit. We apply `isnimgm()` to perform the ISNI analysis:

```
> models= y | g+gp ~   perf + sever +  as.factor(time) + group +
    as.factor(time):group  |  group* as.factor(time)  + yp+ perf + sever


> qolef.isni=isnimgm(models, data=qolef, id=id)
# weights:  36 (22 variable)
```

Table 2: Missing Patterns in the Prostate Cancer QoL Dataset.

				Placebo ($n = 352$)		Flutamide ($n = 363$)	
EF0	EF1	EF2	EF3	Freq	Percent	Freq	Percent
P	P	P	P	239	67.8	258	71.1
P	P	P	A	38	10.8	37	10.2
P	P	A	A	19	5.4	14	3.8
P	A	A	A	20	5.6	22	6.1
P	P	A	P	13	3.7	13	3.6
P	A	P	P	11	3.1	15	4.1
P	A	P	A	7	2.0	1	0.3
P	A	A	P	5	1.4	3	0.8

Note: "P" indicates presence in the visit and "A" indicates absence in the visit.

```
initial  value 2144.491187
iter  10 value 797.136438
iter  20 value 780.985037
iter  30 value 779.568898
final  value 779.561688
converged
# weights:  12 (11 variable)
initial  value 52.679186
iter  10 value 19.792933
iter  20 value 19.572566
iter  30 value 19.570615
iter  40 value 19.570445
final  value 19.570432
converged
```

We summarize the results as follows:

```
> summary(qolef.isni)
Call:
isnimgm(formula = models, data = qolef, id = id)

                        MAR Est.  Std. Err       ISNI        c
(Intercept)             8.430605  0.103039 -0.0268853   5.1162
perf                   -0.271034  0.219967  0.0876716   3.3494
sever                  -0.146618  0.100374  0.0307667   4.3552
as.factor(time)1        0.452170  0.070674  0.1418210   0.6652
as.factor(time)3        0.493918  0.071813  0.1368454   0.7005
as.factor(time)6        0.362126  0.074198  0.1797166   0.5511
group                  -0.020183  0.099961 -0.0011857 112.5434
as.factor(time)1:group -0.220686  0.099022 -0.0144972   9.1182
as.factor(time)3:group -0.220806  0.100413 -0.0210812   6.3586
```

```
as.factor(time)6:group -0.226098  0.103316 -0.0310013   4.4489
sigma                   1.331936  0.025508 -0.0073885   4.6088
rho                     0.548913  0.019775 -0.0077351   3.4128
```

`isnimgm()` calls `gls()` in R package **nlme** to estimate the MMGM under MAR. The ignorable analysis suggests that placebo gives statistically significantly better EF at all three follow-up visits (i.e., the three coefficient estimates for `as.factor(time)1:group`, `as.factor(time)3:group`, and `as.factor(time)6:group` are all significant and negative), after adjusting for the baseline performance score and severity status.

The "ISNI" values quantify the potential change from the MAR estimates when setting $\gamma_1 = 1$. The scale-independent c statistics suggest that the time effect estimates `as.factor(time)1`, `as.factor(time)3`, and `as.factor(time)6` are sensitive to nonignorable missingness. The treatment comparisons at the follow-up visits, `as.factor(time)1:group`, `as.factor(time)3:group`, and `as.factor(time)6:group`, are insensitive, with $c > 1$.

If the data do not contain the missingness status variable, one can specify the model formula without specifying missingness status variable. In this case, users should first sort data by `id` and the variable(s) defining time so that the missingness status variables can be generated correctly. An example code is given below.

```
> qolef <- qolef[order(qolef$id, qolef$time), ]
> models1= y  ~  + perf + sever +  as.factor(time) +  group +
    as.factor(time):group  | as.factor(time) * group + yp+ perf + sever
> qolef.isni <- isnimgm(models1, data=qolef, id=id)
```

To posit an MDM other than the model of Eqn (10), one can supply predicted probabilities of being observed under the desired model through the optional argument `predprobobs`. The **R** code below uses function `tmdm()` to obtain the predicted probabilities of being observed for all observations and passes them as a vector through the argument `predprobobs`. The function `isnimgm()` then refrains from fitting the default first-order missing data model and instead uses `predprobobs` for the ISNI evaluation:

```
> predprobobs= tmdm(g+gp ~ as.factor(time) * group + yp+ perf + sever,
          data=qolef)$obsprob
> summary(isnimgm(models, data=qolef, id=id, predprobobs=predprobobs))

Call:
isnimgm(formula = models, data = qolef, id = id, predprobobs = predprobobs)

                        MAR Est.  Std. Err       ISNI        c
(Intercept)             8.430605  0.103039 -0.0268853   5.1162
perf                   -0.271034  0.219967  0.0876716   3.3494
sever                  -0.146618  0.100374  0.0307667   4.3552
as.factor(time)1        0.452170  0.070674  0.1418210   0.6652
as.factor(time)3        0.493918  0.071813  0.1368454   0.7005
as.factor(time)6        0.362126  0.074198  0.1797166   0.5511
group                  -0.020183  0.099961 -0.0011857 112.5434
as.factor(time)1:group -0.220686  0.099022 -0.0144972   9.1182
```

```
as.factor(time)3:group -0.220806  0.100413 -0.0210812   6.3586
as.factor(time)6:group -0.226098  0.103316 -0.0310013   4.4489
sigma                   1.331936  0.025508 -0.0073885   4.6088
rho                     0.548913  0.019775 -0.0077351   3.4128
```

By default, `isnimgm()` uses the compound symmetry correlation structure in the outcome model of Eqn (6). One can specify alternative correlation structures *via* the argument `cortype`. The current implementation permits two other correlation models: AR1 and unstructured. The code below illustrates the use of this optional argument in the QoL example.

```
> qolef.isni.ar=isnimgm(models, id=id, cortype='AR1', data=qolef)
> qolef.isni.un=isnimgm(models, id=id, cortype='UN', data=qolef)

> summary(qolef.isni.ar)
Call:
isnimgm(formula = models, data = qolef, cortype = "AR1", id = id)

                        MAR Est.  Std. Err       ISNI       c
(Intercept)             8.428061  0.102340 -0.0303458  4.5021
perf                   -0.233958  0.217404  0.0978705  2.9654
sever                  -0.154252  0.097506  0.0348082  3.7395
as.factor(time)1        0.455752  0.064620  0.1161464  0.7427
as.factor(time)3        0.488722  0.082711  0.1399436  0.7890
as.factor(time)6        0.384941  0.094432  0.2070506  0.6088
group                  -0.012404  0.101573 -0.0014319 94.6939
as.factor(time)1:group -0.227717  0.090564 -0.0075759 15.9583
as.factor(time)3:group -0.211955  0.115787 -0.0195638  7.9008
as.factor(time)6:group -0.243135  0.131756 -0.0327193  5.3756
sigma                   1.353818  0.024842 -0.0108687  3.0512
rho                     0.634755  0.015832 -0.0095759  2.2071
> summary(qolef.isni.un)
Call:
isnimgm(formula = models, data = qolef, cortype = "UN", id = id)

                        MAR Est.  Std. Err        ISNI        c
(Intercept)             8.434475  0.105291 -0.02615320   5.3744
perf                   -0.310572  0.224938  0.09172831   3.2736
sever                  -0.150371  0.102682  0.02929663   4.6789
as.factor(time)1        0.456745  0.074694  0.13558115   0.7354
as.factor(time)3        0.474511  0.077934  0.13015864   0.7993
as.factor(time)6        0.350500  0.081979  0.17355412   0.6306
group                  -0.019578  0.102050 -0.00099533 136.8699
as.factor(time)1:group -0.226505  0.104696 -0.00832823  16.7818
as.factor(time)3:group -0.203429  0.109134 -0.01529192   9.5271
as.factor(time)6:group -0.213808  0.114406 -0.02439221   6.2613
sigma                   1.359323  0.026155 -0.00680810   5.1285
cor(1,2)                0.509384  0.024820 -0.00664830   4.9837
```

```
cor(1,3)                0.473068  0.026213  0.00163081  21.4578
cor(1,4)                0.446263  0.028393 -0.00071445  53.0518
cor(2,3)                0.707996  0.018596 -0.00800573   3.1009
cor(2,4)                0.599994  0.028020 -0.00936507   3.9942
cor(3,4)                0.735079  0.016536 -0.00444708   4.9639
```

Our results suggest that the choice of correlation structure has little impact on either the MAR regression parameter estimates or the sensitivity to nonignorability.

The QoL dataset had both intermittent missingness (≈10%) and dropout (≈20%). In some applications, the missingness can only be of one type or the other, and thus the MDM reduces to a special case of the more general multinomial transitional MDM used in our package. To demonstrate the use of `isnimgm()` in these situations, we conduct ISNI analysis on two subsamples of QoL datasets; the first contains only complete observations and dropouts, and the second data only contains complete observations and those with intermittent missingness.

```
> ## Run ISNI analysis on  the subset  that excludes intermittent missingness
> qolefD.isni=isnimgm(models, id=id, data=qolef, subset= g !="I")
> summary(qolefD.isni)
Call:
isnimgm(formula = models, data = qolef, id = id, subset = g !=
"I")

                        MAR Est.  Std. Err        ISNI        c
(Intercept)             8.430605  0.103039 -0.02728549   5.0412
perf                   -0.271034  0.219967  0.10227827   2.8710
sever                  -0.146618  0.100374  0.03041236   4.4059
as.factor(time)1        0.452170  0.070674  0.07582711   1.2442
as.factor(time)3        0.493918  0.071813  0.08529092   1.1240
as.factor(time)6        0.362126  0.074198  0.18277508   0.5419
group                  -0.020183  0.099961 -0.00097846 136.3811
as.factor(time)1:group -0.220686  0.099022  0.00082897 159.4621
as.factor(time)3:group -0.220806  0.100413 -0.01436401   9.3321
as.factor(time)6:group -0.226098  0.103316 -0.02862151   4.8188
sigma                   1.331936  0.025508 -0.00836098   4.0727
rho                     0.548913  0.019775 -0.00876222   3.0128

> ## Run ISNI analysis  on the subset that excludes dropouts.
> qolefI.isni=isnimgm(models, id=id, data=qolef, subset= g != "D")
> summary(qolefI.isni)
Call:
isnimgm(formula = models, data = qolef, id = id, subset = g !=
"D")

                        MAR Est.  Std. Err        ISNI        c
(Intercept)             8.430605  0.103039 -0.00139330  98.7235
perf                   -0.271034  0.219967 -0.00978302  30.0157
sever                  -0.146618  0.100374  0.00246935  54.2629
```

```
as.factor(time)1        0.452170  0.070674  0.07502499   1.2575
as.factor(time)3        0.493918  0.071813  0.05856584   1.6369
as.factor(time)6        0.362126  0.074198 -0.00056741 174.5654
group                  -0.020183  0.099961 -0.00030367 439.4307
as.factor(time)1:group -0.220686  0.099022 -0.01653700   7.9935
as.factor(time)3:group -0.220806  0.100413 -0.00872364  15.3658
as.factor(time)6:group -0.226098  0.103316 -0.00322665  42.7446
sigma                   1.331936  0.025508  0.00030275 112.4753
rho                     0.548913  0.019775  0.00025840 102.1635
```

The results show that ISNIs from both `summary(qolefD.isni)` and `summary(qolefI.isni)` are generally smaller than those from the original dataset as shown in `summary(qolef.isni)`. This is expected because sensitivity tends to increase when the proportion of missingness in a dataset increases.

Last we illustrate the computation of MISNI. As compared with `summary(qolef.isni)`, which considers a scalar γ_1, MISNI shows a slight increase in sensitivity.

```
> qolef.misni=isnimgm(models, id=id, data=qolef, misni=T)
> summary(qolef.misni)
Call:
isnimgm(formula = models, data = qolef, id = id, misni = T)

                        MAR Est.  Std. Err      ISNI        c
(Intercept)             8.430605  0.103039 0.0270835   5.0788
perf                   -0.271034  0.219967 0.1159199   2.5332
sever                  -0.146618  0.100374 0.0307667   4.3552
as.factor(time)1        0.452170  0.070674 0.1418210   0.6652
as.factor(time)3        0.493918  0.071813 0.1368454   0.7005
as.factor(time)6        0.362126  0.074198 0.1860771   0.5323
group                  -0.020183  0.099961 0.0011857 112.5434
as.factor(time)1:group -0.220686  0.099022 0.0172965   7.6425
as.factor(time)3:group -0.220806  0.100413 0.0210812   6.3586
as.factor(time)6:group -0.226098  0.103316 0.0319254   4.3201
sigma                   1.331936  0.025508 0.0093732   3.6329
rho                     0.548913  0.019775 0.0098674   2.6753
```

4.3. ISNI analysis of an LMM for longitudinal data

We will illustrate the LMM analysis using the SWOG QoL data. We first consider a random intercept model with the output below. This model is equivalent to the marginal multivariate model with compound symmetry correlation structure illustrated above. Consequently they produce similar MAR inference and ISNI analysis results.

```
> data(qolef)
> models= y | g+gp ~ perf + sever +  as.factor(time) + group +
    as.factor(time):group | group* as.factor(time)  + yp+ perf + sever
```

```
## Random intercept model
> result=isnilmm(models, random=~1, id=id, data=qolef)

> summary(result)
Call:
isnilmm(formula = models, data = qolef, random = ~1, id = id)

                         MAR Est.  Std. Err       ISNI        c
(Intercept)              8.430611  0.102761 -0.0267777   5.1229
perf                    -0.270986  0.219351  0.0874180   3.3497
sever                   -0.146629  0.100085  0.0306377   4.3609
as.factor(time)1         0.452168  0.070555  0.1413020   0.6666
as.factor(time)3         0.493949  0.071709  0.1363900   0.7019
as.factor(time)6         0.362189  0.074101  0.1791559   0.5522
group                   -0.020181  0.099713 -0.0011793 112.8717
as.factor(time)1:group -0.220661  0.098839 -0.0144065   9.1587
as.factor(time)3:group -0.220821  0.100255 -0.0210332   6.3630
as.factor(time)6:group -0.226147  0.103170 -0.0309658   4.4477
sigmav                   0.983684  0.033229 -0.0135647   3.2702
sigmae                   0.893093  0.014957  0.0036156   5.5222
```

We next consider a model with random effects for both intercept and time slope. The ISNI analysis shows that the `time` estimate has $c < 1$, suggesting sensitivity to nonignorable missingness. It also shows that the `time:group` estimate for measuring the treatment group differences over time has a $c > 1$, suggesting that sensitivity to nonignorability is not of concern for this estimate. These results echo our findings under the MMGM.

```
## Random intercept and slope model
> models1=y |g+gp~   time*group + perf + sever
> result1=isnilmm(models1,random= ~ 1+time, id=id, data=qolef)

> summary(result1)
Call:
isnilmm(formula = models1, data = qolef, random = ~1 + time,
id = id)

             MAR Est.  Std. Err       ISNI       c
(Intercept)  8.633868  0.099125  0.0134140  9.8648
time         0.047974  0.013094  0.0294745  0.5930
group       -0.103960  0.092752 -0.0033128 37.3758
perf        -0.263945  0.221257  0.1001395  2.9495
sever       -0.152619  0.099815  0.0326422  4.0820
time:group  -0.032053  0.018209 -0.0064265  3.7824
sigmav1      1.043033  0.039745 -0.0124504  4.2614
sigmav2      0.118714  0.014123  0.0025849  7.2935
rho12       -0.327000  0.080491 -0.0344655  3.1176
sigmae       0.855890  0.017576  0.0045477  5.1592
```

5. Discussion

Because ignorability is a critical but unverifiable assumption in the analysis of incomplete data, there is a pressing need for general-purpose statistical software for sensitivity analysis. This article has described a new R package **isni** that can measure the impact of nonignorable missingness on the standard MAR analysis. The current version of the program provides functionality to compute the ISNI in the GLM for independent data, as well as the MMGM and LMM for longitudinal/clustered data. We intend that the availability of this software will make the analysis of sensitivity to nonignorability a routine part of statistical practice.

It is not always possible to distinguish intermittent missing observations from true dropouts. For example, if in the SWOG study a subject had missed the visit at Month 1 but returned for Month 3, then clearly Month 1 was an intermittent missing observation. But for a subject who missed the Month 1 visit with the full intention of returning for Month 3, but then ended up leaving the study before Month 3, we would (incorrectly) classify the Month 1 missing observation as the beginning of a dropout sequence. Accurate modeling of the MDM thus depends on correctly gleaning the reasons for missed visits. It is better to get this information directly from the subjects than to try to infer it from the pattern of missing observations (Xie 2012).

Currently, the package **isni** has functions that can compute ISNI for any GLM with independent, univariate outcomes, and for continuous longitudinal outcomes that are normally distributed with a linear mean structure. We plan to extend it to cover the case of the generalized linear mixed model, which includes several models for non-normal clustered and longitudinal data (Xie 2008). In this model, the computation of ISNI is more complex, as it requires numerical integration with respect to the random effects.

Another potential extension of the software would be to the analysis of local sensitivity to nonignorability under the general coarse-data model of Heitjan and Rubin (1991). Examples of such analyses for censored event-time outcomes appear in Zhang and Heitjan (2005, 2006, 2007); Liu and Heitjan (2012).

References

Copas, JB, and Eguchi, S (2001). "Local sensitivity approximations for selectivity bias." *Journal of the Royal Statistical Society B*, **63**, 871–895.

Copas, JB, and Li, HG (1997). "Inference for non-random samples (with discussion)." *Journal of the Royal Statistical Society B*, **59**, 55–95.

Daniels M, Hogan J (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis.* Chapman and Hall, Boca Raton.

Gao W, Hedeker D, Mermelstein R, Xie H (2016). "A scalable approach to measuring the impact of nonignorable nonresponse with an EMA application." *Statistics in Medicine*, **35**, 5579–5602.

Heitjan D, Rubin D (1991). "Ignorability and coarse data." *Annals of Statistics*, **19**, 2244–2253.

Laird N, Ware J (1982). "Random effects models for longitudinal data." *Biometrics*, **38**, 963–974.

Little R, D'Agostino R, Cohen M, Dickersin K, Emerson S, Farrar J, Frangakis C, Hogan J, Molenberghs G, Murphy S, Neaton J, Rotnitzky A, Scharfstein D, Shih W, Siegel J, Stern H (2012). "The prevention and treatment of missing data in clinical trials." *New England Journal of Medicine*, **367**, 1355–1360.

Liu T, Heitjan D (2012). "Sensitivity of the discrete-time Kaplan-Meier estimate to nonignorable censoring: Application in a clinical trial." *Statistics in Medicine*, **31**, 2998–3010.

Ma G, Troxel AB, Heitjan D (2005). "An index of local sensitivity to nonignorable dropout in longitudinal modeling." *Statistics in Medicine*, **24**, 2129–2150.

McCullagh P, Nelder J (1989). *Generalized Linear Models.* Chapman and Hall, New York.

on Handling Missing Data in Clinical Trials; National Research Council TP (2010). *The Prevention and Treatment of Missing Data in Clinical Trials.* National Academies Press, Washington, DC.

Raab G, Donelly C (1999). "Information on sexual behaviour when some data Are missing." *Applied Statistics*, **48**, 117–133.

Rubin DB (1976). "Inference and missing data (with discussion)." *Biometrika*, **63**, 581–592.

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

Vansteelandt S, Rotnitzky A, Robins J (2007). "Estimation of regression models for the mean of repeated outcomes under nonignorable nonmonotone nonresponse." *Biometrika*, **94**, 841–860.

Verbeke G, Molenberghs G, Thijs H, Lesaffre E, Kenward MG (2001). "Sensitivity analysis for nonrandom dropout: A local influence approach." *Biometrics*, **57**, 7–14.

Xie H (2008). "A local sensitivity analysis approach to longitudinal non-Gaussian data with non-ignorable dropout." *Statistics in Medicine*, **27**, 3155–3177.

Xie H (2009). "Bayesian inference from incomplete longitudinal data: A simple method to quantify sensitivity to nonignorable dropout." *Statistics in Medicine*, **28**, 2725–2747.

Xie H (2010). "Adjusting for nonignorable missingness when estimating generalized additive models." *Biometrical Journal*, **52**, 186–200.

Xie H (2012). "Analyzing longitudinal clinical trial data with nonignorable missingness and unknown missingness reasons." *Computational Statistics & Data Analysis*, **56**, 1287–1300.

Xie H, Heitjan DF (2004). "Sensitivity analysis of causal inference in a clinical trial subject to crossover." *Clinical Trials*, **1**, 21–30.

Xie H, Heitjan DF (2009). "Local sensitivity to nonignorability: Dependence on the assumed dropout mechanism." *Statistics in Biopharmaceutical Research*, **29**, 243–257.

Xie H, Qian Y (2012). "Measuring the impact of nonignorability in panel data with non-monotone nonresponse." *Journal of Applied Econometrics*, **27**, 129–159.

Xie H, Qian Y, Qu L (2011). "A semiparametric approach for analyzing nonignorable missing data." *Statistica Sinica*, **21**, 1881–1899.

Zeileis A, Croissant Y (2010). "Extended Model Formulas in R: Multiple Parts and Multiple Responses." *Journal of Statistical Software*, **34**(1), 1–13. doi:10.18637/jss.v034.i01.

Zhang J, Heitjan D (2005). "Nonignorable censoring in randomized clinical trials." *Clinical Trials*, **2**, 488–496.

Zhang J, Heitjan D (2006). "A simple local sensitivity analysis tool for nonignorable coarsening: Application to dependent censoring." *Biometrics*, **62**, 1260–1268.

Zhang J, Heitjan D (2007). "Impact of nonignorable coarsening on Bayesian inference." *Biostatistics*, **8**, 722–743.

Acknowledgements

The US National Cancer Institute supported our research under awards R01CA178061 and P01CA098262. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

Appendix A: Derivation of ISNI

For fixed γ_1, the conditional maximum likelihood estimates $\hat{\theta}(\gamma_1)$ and $\hat{\gamma}_0(\gamma_1)$ satisfy

$$\frac{\partial L(\hat{\theta}(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)}{\partial(\theta^T, \gamma_0^T)^T} = 0,$$

where $L(\theta, \gamma_0, \gamma_1)$ is the loglikelihood for the selection model (Eqn 1). Differentiating both sides with respect to γ_1 and noting that $\hat{\theta}(\gamma_1)$, and $\hat{\gamma}_0(\gamma_1)$ are implicit functions of γ_1, we have

$$\frac{\partial^2 L(\hat{\theta}(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)}{\partial(\theta^T, \gamma_0^T)^T \partial\gamma_1^T} + \frac{\partial^2 L(\hat{\theta}(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)}{\partial(\theta^T, \gamma_0^T)^T \partial(\theta^T, \gamma_0^T)} \frac{\partial(\hat{\theta}^T(\gamma_1), \hat{\gamma}_0^T(\gamma_1))^T}{\partial\gamma_1^T} = 0.$$

Thus for any γ_1, we have

$$\frac{\partial(\hat{\theta}^T(\gamma_1), \hat{\gamma}_0^T(\gamma_1))^T}{\partial\gamma_1^T} = -\left[\frac{\partial^2 L(\hat{\theta}(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)}{\partial(\theta^T, \gamma_0^T)^T \partial(\theta^T, \gamma_0^T)}\right]^{-1} \frac{\partial^2 L(\hat{\theta}(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)^T}{\partial(\theta^T, \gamma_0^T)^T \partial\gamma_1^T}. \tag{15}$$

In our local sensitivity analysis, the primary interest is to investigate sensitivity around the MAR model, i.e., $\gamma_1 = 0$. This local sensitivity can be captured by the derivatives at this point. In particular, we define the first derivative evaluated at $\gamma_1 = 0$ as ISNI and

$$\begin{aligned} \text{ISNI} &= \left.\frac{\partial\left(\hat{\theta}^T(\gamma_1), \hat{\gamma}_0^T(\gamma_1)\right)^T}{\partial\gamma_1^T}\right|_{\gamma_1=0} \\ &= -\begin{bmatrix} \nabla^2 L_{\theta,\theta} & \nabla^2 L_{\theta,\gamma_0} \\ \nabla^2 L_{\gamma_0,\theta} & \nabla^2 L_{\gamma_0,\gamma_0} \end{bmatrix}^{-1} \begin{bmatrix} \nabla^2 L_{\theta_y,\gamma_1} \\ \nabla^2 L_{\gamma_0,\gamma_1} \end{bmatrix}, \end{aligned}$$

where $\nabla^2 L_{a,b} = \left.\frac{\partial^2 L(\theta(\gamma_1), \gamma_0(\gamma_1), \gamma_1)}{\partial a \partial b}\right|_{\hat{\theta}(0), \hat{\gamma}_0(0), 0}$. Under MAR, we have $\nabla^2 L_{\theta,\gamma_0} = 0$, and thus the ISNI for $\hat{\theta}$, the parameter estimates of primary interest, is

$$\left.\frac{\partial\hat{\theta}(\gamma_1)}{\partial\gamma_1^T}\right|_{\gamma_1=0} = -\nabla^2 L_{\theta,\theta}^{-1} \nabla^2 L_{\theta,\gamma_1},$$

where

$$\nabla^2 L_{\theta,\theta} = \left.\frac{\partial^2 L(\theta, \gamma_0, \gamma_1)}{\partial\theta\partial\theta^T}\right|_{\hat{\theta}(0), \hat{\gamma}_0(0), \gamma_1=0} = \left.\frac{\partial \ln f_\theta(y^{\text{obs}})}{\partial\theta\partial\theta^T}\right|_{\hat{\theta}(0), \hat{\gamma}_0(0), \gamma_1=0}$$

and

$$\nabla^2 L_{\theta,\gamma_1} = \left.\frac{\partial^2 L(\theta, \gamma_0, \gamma_1)}{\partial\theta\partial\gamma_1^T}\right|_{\hat{\theta}(0), \hat{\gamma}_0(0), \gamma_1=0}.$$

Appendix B: Derivation of $\nabla^2 L_{\theta,\gamma_1}$

To derive the likelihood of the nonignorable selection model for longitudinal data, adopt the notation of $Y_i = (Y_i^{\text{obs}}, Y_i^{\text{mis}})$, where Y_i^{obs} refers to the components of Y_i that are observed, and Y_i^{mis} refers to the components of Y_i that are missing. Let K_i be the length of Y_i^{obs}. If subject i completed all the intended visits of the study, then $K_i = n_i$, and Y_i^{mis} vanishes; otherwise, $K_i < n_i$. Let L be the correct loglikelihood for $(\theta, \gamma_0, \gamma_1)$ under the nonignorable selection model specified in Sections 2.2.1 and 2.2.2. Then

$$\begin{aligned}
L(\theta,\, \gamma_0,\, \gamma_1) &= \sum_{i=1}^{N} L_i\left(\theta,\, \gamma_0,\, \gamma_1; y_i^{\text{obs}}, g_i\right) \\
&= \sum_{i=1}^{N} \ln\left(\int \prod_{j=1}^{n_i} f_\gamma\left(g_{ij}|s_{ij}, y_{ij}, g_{i,j-1}\right)\, f_\theta(y_i^{\text{obs}}, y_i^{\text{mis}}|x_i) dy_i^{\text{mis}}\right) \\
&= \sum_{i=1}^{N} \ln f_\theta(y_i^{\text{obs}}|x_i) + \\
&\quad \sum_{i=1}^{N} \sum_{j:g_{ij}=0} \ln f_\gamma(g_{ij}|s_{ij}, y_{ij}, g_{i,j-1}) + \\
&\quad \sum_{i:K_i<n_i} \ln\left(\int \prod_{j:g_{ij}\neq 0} f_\gamma(g_{ij}|s_{ij}, y_{ij}, g_{i,j-1}) f_\theta(y_i^{\text{mis}}|y_i^{\text{obs}}, x_i) dy_i^{\text{mis}}\right), \qquad (16)
\end{aligned}$$

where $g_i = (g_{i1}, ..., g_{in_i})$ is a vector of discrete variables for the missingness status of subject i; $f_\theta(y_i^{\text{obs}}, y_i^{\text{mis}}|x_i)$ is the density function of the outcome model defined above; $f_\gamma\left(g_{ij}|s_{ij}, y_{ij}, g_{i,j-1}\right)$ is the probability mass function of the missing-data model defined in Eqn (11), and if the general transitional model as specified in Eqn (8) is used for modeling G_{ij}, $f_\gamma\left(g_{ij}|s_{ij}, y_{ij}, g_{i,j-1}\right)$ is then replaced with $f_\gamma\left(g_{ij}|s_{ij}, y_{ij}, g_{i(j)}\right)$. We intend the integral sign to refer to summation with discrete outcomes.

It is readily seen that the components of y_i^{mis} after dropout do not enter the integral in Eqn (16), because these outcomes are deterministically missing. Thus, the dimensionality of the integration for the ith unit is $d_i = \sum_j I(g_{ij} = 1) + I(\text{any of } g_{ij} \text{ is } 2)$. Henceforth, the notation y_i^{mis} includes only the intermittent missing outcomes and the outcome at the time of dropout. With nonignorable missingness, the integral with respect to y_i^{mis} does not have a closed form, and we require a numerical method for its evaluation. The computational workload for such integration increases exponentially with the number of intermittent missing outcomes, rendering the evaluation of L difficult with even moderate intermittent missingness.

To derive $\nabla^2 L_{\theta,\gamma_1}$, we note that $\nabla^2 L_{\theta,\gamma_1} = (\nabla^2 L_{\theta,\gamma_{10}}, \nabla^2 L_{\theta,\gamma_{20}}, \nabla^2 L_{\theta,\gamma_{11}})$, where

$$
\begin{aligned}
\nabla^2 L_{\theta,\gamma_1^{10}} &= \sum_{i:K_i<n_i} \frac{\partial^2}{\partial\theta\partial\gamma_1^{10}} \ln \left(\int \prod_{j:g_{ij}\neq 0} f_\gamma(g_{ij}|s_{ij}, y_{ij}, g_{i,j-1}) f_\theta(y_i^{\text{mis}}|y_i^{\text{obs}}, x_i) dy_i^{\text{mis}} \right)\Bigg|_{\gamma_1=0} \\
&= \sum_{i:K_i<ni} \frac{\partial}{\partial\theta} \left(\frac{\int \frac{\partial \prod_{j:g_{ij}\neq 0} f_\gamma(g_{ij}|s_{ij},y_{ij},g_{i,j-1})}{\partial\gamma_1^{10}} f_\theta(y_i^{\text{mis}}|y_i^{\text{obs}}, x_i) dy_i^{\text{mis}}}{\int \prod_{j:g_{ij}\neq 0} f_\gamma(g_{ij}|s_{ij}, y_{ij}, g_{i,j-1}) f_\theta(y_i^{\text{mis}}|y_i^{\text{obs}}, x_i) dy_i^{\text{mis}}} \Bigg|_{\gamma_1=0} \right) \\
&= \sum_{i:K_i<ni} \frac{\partial}{\partial\theta} \left(\int \sum_{j:g_{ij}\neq 0} \frac{I(g_{i,j-1}=0) \frac{\partial P_{ij}^{g_{ij},g_{i,j-1}}}{\partial \phi_{ij}^{1,g_{i,j-1}}} \phi_{ij}^{1,g_{i,j-1}}}{P_{ij}^{g_{ij},g_{i,j-1}}} y_{ij} f_\theta(y_i^{\text{mis}}|y_i^{\text{obs}}, x_i) dy_i^{\text{mis}} \Bigg|_{\gamma_1=0} \right) \\
&= \sum_{i:K_i<n_i} \frac{\partial}{\partial\theta} \mathrm{E}((Y_i^{\text{mis}})^T A_i^{10} | y_i^{\text{obs}}, x_i) \Bigg|_{\gamma_1=0} \\
&= \sum_{i:K_i<n_i} \frac{\partial \mathrm{E}((Y_i^{\text{mis}})^T | y_i^{\text{obs}}, x_i)}{\partial\theta} \Bigg|_{\gamma_1=0} \cdot A_i^{10}. \qquad (17)
\end{aligned}
$$

If the lth component of y_i^{mis} corresponds to the jth element of y_i, the lth element of A_i^{10} is

$$
\begin{aligned}
A_{il}^{10} &= \frac{I(g_{i,j-1}=0) \frac{\partial P_{ij}^{g_{ij},g_{i,j-1}}}{\partial \phi_{ij}^{1,g_{i,j-1}}} \phi_{ij}^{1,g_{i,j-1}}}{P_{ij}^{g_{ij},g_{i,j-1}}} \Bigg|_{\hat\gamma_0(0),\gamma_1=0} \\
&= I(g_{i,j-1}=0) \left[I(g_{i,j}=1) - P_{ij}^{10} \right]_{\hat\gamma_0(0),\gamma_1=0}.
\end{aligned}
$$

$\nabla^2 L_{\theta,\gamma_1^{20}}$ and $\nabla^2 L_{\theta,\gamma_1^{11}}$ are derived similarly to Eqn (17) with A_i^{10} replaced by A_i^{20} and A_i^{11}, respectively, where

$$
\begin{aligned}
A_{il}^{20} &= I(g_{i,j-1}=0) \left[I(g_{i,j}=2) - P_{ij}^{20} \right]_{\hat\gamma_0(0),\gamma_1=0}, \\
A_{il}^{11} &= I(g_{i,j-1}=1) \left[P_{ij}^{01} \right]_{\hat\gamma_0(0),\gamma_1=0}.
\end{aligned}
$$

Appendix C: Derivation of $\frac{\partial \mathbf{E}((Y_i^{\mathbf{mis}})^T | y_i^{\mathbf{obs}}, x_i)}{\partial\theta}\Big|_{\gamma_1=0}$

We note that y_i^{mis} is a vector of length $d_i = \sum_j I(g_{ij}=1) + I(\text{any of } g_{ij} \text{ is } 2)$.

C.1 MMGM. Because

$$
\mathrm{E}((Y_i^{\text{mis}})^T | y_i^{\text{obs}}, x_i)\Big|_{\gamma_1=0} = \theta_1^T x_{i,M}^T + ((y_i^{\text{obs}})^T - \theta_1^T x_{i,O}^T) \Sigma_{i,OO}^{-1} \Sigma_{i,OM},
$$

by vector differentiation, we have

$$\begin{aligned}
\left.\frac{\partial \mathrm{E}((Y_i^{\mathrm{mis}})^T | y_i^{\mathrm{obs}}, x_i)}{\partial \theta_1}\right|_{\gamma_1=0} &= x_{i,M}^T - x_{i,O}^T \Sigma_{i,OO}^{-1} \Sigma_{i,OM} \\
\left.\frac{\partial \mathrm{E}((Y_i^{\mathrm{mis}})^T | y_i^{\mathrm{obs}}, x_i)}{\partial \theta_2}\right|_{\gamma_1=0} &= -((y_i^{\mathrm{obs}})^T - \theta_1^T x_{i,O}^T) \Sigma_{i,OO}^{-1} \frac{\partial \Sigma_{i,OO}}{\partial \theta_2} \Sigma_{i,OO}^{-1} \Sigma_{i,OM} \\
&\quad + ((y_i^{\mathrm{obs}})^T - \theta_1^T x_{i,O}^T) \Sigma_{i,OO}^{-1} \frac{\partial \Sigma_{i,OM}}{\partial \theta_2},
\end{aligned}$$

where $x_{i,O}$ and $x_{i,M}$ are the predictor matrices for Y_i^{obs} and Y_i^{mis}, respectively, and

$$\mathrm{Var}(Y_i^{\mathrm{obs}}, Y_i^{\mathrm{mis}} | X_i) = \begin{pmatrix} \Sigma_{i,OO} & \Sigma_{i,OM} \\ \Sigma_{i,MO} & \Sigma_{i,MM} \end{pmatrix}.$$

C.2 LMM. The linear mixed-effect model as specified in Eqn (7) can be re-expressed in a marginal multivariate normal model, where $\Sigma_i = \Lambda_i + Z_i V_b Z_i^T$. Then we can apply the above result for multivariate normal model, where

$$\frac{\partial \Sigma_i}{\partial \theta_2} = \frac{\partial \Lambda_i}{\partial \theta_2} + Z_i \frac{\partial V_b}{\partial \theta_2} Z_i^T$$