Measuring the Impact of Nonignorable Missingness Using the R Package isni

Hui Xie\textsuperscript{a,b}, Weihua Gao\textsuperscript{a}, Baodong Xing\textsuperscript{d}, Daniel F. Heitjan\textsuperscript{e,f}, Donald Hedeker\textsuperscript{g}, Chengbo Yuan\textsuperscript{a}

\textsuperscript{a} Division of Epidemiology & Biostatistics (M/C 923), School of Public Health, University of Illinois at Chicago, 1603 West Taylor Street, Room 984, Chicago, 60612-4336, IL
\textsuperscript{b} Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, V5S0G6, Canada
\textsuperscript{c} Arthritis Research Canada, Richmond, BC, V6X 2C7, Canada
\textsuperscript{d} AbbVie Inc., Chicago, 60632, IL
\textsuperscript{e} Department of Statistical Science, Southern Methodist University, Dallas, 75275-0332, TX
\textsuperscript{f} Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, 75390-9066, TX
\textsuperscript{g} Department of Public Health Sciences, The University of Chicago, Chicago, 60637, IL

A R T I C L E   I N F O

Article history:
Received 19 April 2018
Revised 18 June 2018
Accepted 22 June 2018

Keywords:
Analytical reliability
Data quality
Missing data
Missing not at random
Multivariate normal
Selection model

A B S T R A C T

Background and Objective: The popular assumption of ignorability simplifies analyses with incomplete data, but if it is not satisfied, results may be incorrect. Therefore it is necessary to assess the sensitivity of empirical findings to this assumption. We have created a user-friendly and freely available software program to conduct such analyses.

Method: One can evaluate the dependence of inferences on the assumption of ignorability 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 parameter estimates to the assumed degree of nonignorability in the neighborhood of an ignorable model. Computation of ISNI avoids the need to estimate a nonignorable model or to posit a specific magnitude of nonignorability. Our new R package, named isni, implements ISNI analysis for some common data structures and corresponding statistical models.

Result: The isni 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. It allows for arbitrary patterns of missingness caused by dropout and/or intermittent missingness. Examples illustrate its use and features.

Conclusions: The R package isni enables a systematic and efficient sensitivity analysis that informs evaluations of reliability and validity of empirical findings from incomplete data.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Ignorability is a key 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 need not model it in order to generate valid likelihood/Bayesian inferences [6,15]. 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 MAR does not hold, and therefore that the underlying missing data mechanism is nonignorable. For example, if patients drop out of a study because of excessive toxicity, their missingness may be related to their 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, standard 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, it is critical to have the means to conduct an analysis of sensitivity to departures from ignorability [4,9]. 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” [16]. Moreover, “[s]ensitivity analysis to missing data as-
sinations] is a relatively new area, and further research on the best methods is needed.” [9]

Estimating nonignorable models of missingness is conceptually and computationally challenging, which limits the types of sensitivity analyses that one can perform [23,26]. One approach is to compute the index of local sensitivity to nonignorability, or ISNI, first proposed by [17] in the context of the univariate generalized linear model. The ISNI method implements the local sensitivity approach that examines the effect on inferences of minor departures from ignorability [2,39]. The method has since been extended to a range of statistical models and coarsening types [51,12–27]. ISNI overcomes the computational obstacles to sensitivity analysis by requiring only readily-available MAR model computations, 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 common models — currently, generalized linear models for cross-sectional data, and marginal multivariate Gaussian models and linear mixed-effects models for longitudinal/clustered data. This article describes the software and its application.

We organize the article as follows: Section 2 reviews the ISNI method. Section 3 describes various aspects of using main functions in 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 details.

2. Review of the ISNI Method

Let \( Y \) be a vector of outcomes and \( G \) be a 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 \( Y = (Y_{\text{obs}}, Y_{\text{mis}}) \), where \( Y_{\text{obs}}, Y_{\text{mis}} \) 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 \( f_0(y) \), indexed by parameter \( \theta \), and the conditional probability mass function of \( G \) given \( Y = y \) is \( f_{0,y|\gamma_1}(g|y) \), indexed by parameters \( \gamma_0 \) and \( \gamma_1 \). We further assume that we can write the conditional probability function of \( G \) given \( Y = y \) as \( f_{0,y|\gamma_1}(g|y_{\text{obs}}, y_{\text{mis}}) \), such that the parameter \( \gamma_0 \) reflects the effect of fully observed data elements on the probability of missingness, and the parameter \( \gamma_1 \) reflects the effect of potentially unobserved outcomes on the probability of missingness. We denote the conditional distribution of the missingness indicator to be the missing data mechanism (MDM). Under these assumptions, the observed data \((y_{\text{obs}}, g)\) is missing at random (MAR) if, for every possible value of the parameters, \( f_{0,y|\gamma_1}(g|y_{\text{obs}}, y_{\text{mis}}) \) takes the same value for all \( y_{\text{mis}} \). 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 \( G \). An MMD is MAR if every possible data set generated under it is MAR.

Following common practice, we formulate the MDM to reduce to MAR when \( \gamma_1 = 0 \). When \( \gamma_1 \neq 0 \), the missingness probability depends on unobserved data, \( y_{\text{mis}} \), even after conditioning on the observed data. Because \( \gamma_1 \) 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

\[
L(\theta, \gamma_0, \gamma_1; y_{\text{obs}}, g) = \ln \int_{\Omega_{\gamma_{mis}}} f_0(y_{\text{obs}}, y_{\text{mis}}) f_{0,y|\gamma_1}(g|y_{\text{obs}}, y_{\text{mis}}) dy_{\text{mis}},
\]

where \( \Omega_{\gamma_{mis}} \) denotes the sample space of \( y_{\text{mis}} \), and \( f_0(y_{\text{obs}}, y_{\text{mis}}) \) is \( f_0(y_{\text{obs}}, y_{\text{mis}}) \) evaluated at the observed data \( y_{\text{obs}} \) and with the missing value \( y_{\text{mis}} \) substituted by a hypothesized postive value \( y_{\text{mis}}^* \) in \( \Omega_{\gamma_{mis}} \); the completeness indicator \( g \) dictates the identities of the observed and missing \( y \) values. Note that if we assume an MAR MDM (i.e., \( \gamma_1 = 0 \)), we have \( f_{0,y|\gamma_1}(g|y_{\text{obs}}) = f_{0|\gamma_1}(g|y_{\text{obs}}) \), because this conditional probability does not depend on values of \( y_{\text{mis}} \). Thus we can move this term out of the integration in Eqn (1), resulting in the simpler loglikelihood

\[
L(\theta, \gamma_0, \gamma_1; y_{\text{obs}}, g) = \ln \left( \int_{\Omega_{\gamma_{mis}}} f_0(y_{\text{obs}}, y_{\text{mis}}) dy_{\text{mis}} \right) f_{0,y|\gamma_1}(g|y_{\text{obs}}) \]

\[
= \ln f_0(y_{\text{obs}}) + \ln f_{0,y|\gamma_1}(g|y_{\text{obs}}).
\]

Under MAR and the additional assumption of parameter distinctness — i.e., that \( \theta \) and \( \gamma \) are independent (for Bayesian inference) or have parameter spaces that factor (for likelihood inference) — \( \ln f_0(y_{\text{obs}}) \) contains the totality of information on \( \theta \), and therefore it is unnecessary 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 \( y_{\text{mis}} \), one cannot identify or estimate such a nonignorable model without either additional data or untestable assumptions.

Practical analyses therefore typically assume MAR and base inferences on Eqn (2). To evaluate the robustness of such an inference, one can conduct a sensitivity analysis. One way to do this is to posit a range of values of \( \gamma_1 \) and determine the extent to which an estimate of \( \theta \), such as the MLE given \( y_{\text{obs}}, g, \gamma \), depends on the values of \( \gamma_1 \). The ISNI approach is to execute this analysis in a neighborhood of the MAR model by determining the rate of change of \( \hat{\theta}(y_{\text{obs}}, g) \) as a function of \( \gamma_1 \) at \( \gamma_1 = 0 \). That is, ISNI calculates the derivative of \( \hat{\theta}(y_{\text{obs}}, g) \) with respect to \( \gamma_1 \), evaluated locally at the ignorable model [17]. As we show in Appendix A, a general formula for ISNI is

\[
\text{ISNI} = \frac{\partial \hat{\theta}(y_{\text{obs}}, g)}{\partial \gamma_1} \bigg|_{\gamma_1=0} = -\nabla^2 L_{\theta,0,\gamma_1}^{-1} \nabla L_{\theta,0,\gamma_1} \bigg|_{\gamma_1=0},
\]

and \( L(\theta, \gamma_0, \gamma_1) \) denotes the loglikelihood of Eqn (1). The computation of ISNI thus involves two parts: First one computes \( \nabla^2 L_{\theta,0,\gamma_1}^{-1} \), which is the inverse of the second-order matrix \( \Omega_{\gamma_{mis}} \) under the MARD outcome model; this is readily available from standard statistical software. Second, one computes \( \nabla^2 L_{\theta,0,\gamma_1} \), which measures the nonorthogonality of \( \theta \) and \( \gamma_1 \). 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) [13] that assumes that scalar \( y \), given predictors \( x_i \), \( i = 1, \ldots, N \), are independent draws from the exponential family

\[
f_0(y|x_i) = \exp \left\{ \frac{y_i \Psi_i(\beta, x_i) - b(\Psi_i(\beta, x_i))}{\alpha_i(\tau)} + c(y_i, \tau) \right\},
\]

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

\[
P(G_i = 1 | s_i, y_i) = \frac{1}{1 + \exp\left[-(\gamma_0^i s_i + \gamma_1^i y_i)\right]},
\]

where \( G_i = 0(1) \) if the \( i \)th observation is observed (missing), and \( s_i \) includes a set of observed predictors. The parameters \( \gamma_0 \) and \( \gamma_1 \) associate the probability of missingness with the set of fully observed missingness predictors in \( s_i \) and the partially observed variable \( y_i \), respectively. 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 [25].

Following [17], under independence over units \( i \) we have

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

\[
= \sum_{i: g_i = 0} \frac{\partial \ln f_0(y_i^{obs}|x_i)}{\partial \theta} \frac{\partial y_i^{obs}}{\partial \gamma} \bigg|_{\gamma_0 = 0},
\]

\[
\nabla^2 L_{\theta, \gamma_1} = \frac{\partial^2 L(\theta, y_0, y_1)}{\partial \theta \partial \gamma^T} \bigg|_{\gamma_1 = 0},
\]

\[
= \sum_{i: g_i = 1} (1 - h_i) \frac{\partial E(Y_i^{obs}|x_i)}{\partial \theta} \bigg|_{\gamma_1 = 0}.
\]

ISNI for \( \theta = (\beta, \tau) \) simplifies to

\[
\text{ISNI} = -\left[ \sum_{i: g_i = 0} (1 - g_i) \left( y_i^{\beta, 0} - \frac{\partial \hat{h}_i^T}{\partial \hat{r}_{ij}} \right) \right]^{-1} \sum_{i: g_i = 0} g_i (1 - h_i) \frac{\partial \hat{h}_i^T}{\partial \hat{r}_{ij}},
\]

where \( h_i = h(\hat{y}_i^T s_i) \) is the predicted probability of being under MAR, and \( \hat{h}(0) \) and \( \hat{r}(0) \) are MLEs under MAR. This formula is the same as that presented in [17] except that we reverse the signs to reflect the reversed role of the indicator \( g \).

2.2. ISNIs for longitudinal/clumped data

Several authors have described generalizations of ISNI to longitudinal/clumped data [11, 20, 21, 23, 25, 26]. We describe below the ISNI method for the setting of longitudinal data with non-monotone missingness; adaptation to other types of clustered data is straightforward.

2.2.1. Models for the notional complete data

Let \( Y_j = (Y_{ij1}, \ldots, Y_{ijn}) \) denote the notional complete data for subject \( i \), where \( Y_{ij} \) is the outcome at measurement occasion \( j, i = 1, \ldots, N, j = 1, \ldots, n_i \). We assume that the density function of \( Y_j \) is \( f_0(y_j|x_j) \), where \( \theta \) is the vector of the parameters of interest with length \( p_0 \), and \( x_j \) 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). The MMGM for a continuous outcome is

\[
Y_i|x_i \overset{ind}{\sim} \text{MVN}(\theta_1, \Sigma_i(\theta_2)),
\]

where \( \theta_1 \) and \( \theta_2 \) are parameters of the population mean and the variance-covariance, respectively. The matrix \( \Sigma_i(\theta_2) \) must be positive definite. Under ignorable missingness, one typically estimates this model by generalized least-squares, for example with the R function glm().

Linear mixed model (LMM). The LMM for a continuous outcome is [8]

\[
Y_i|x_i, z_i \overset{ind}{\sim} \text{MVN}(x_i^g \beta + z_i b_i, \Lambda_i), \quad b_i \overset{iid}{\sim} N(0, V_b).
\]

Here \( \beta \) 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 \( \Lambda_i \) are variance-covariance matrices for the random effects and residuals, respectively. \( \Lambda_i \) depends on \( i \) only in that its size is \( n_i \times n_i \). Marginally,

\[
Y_i|x_i, z_i \overset{ind}{\sim} \text{MVN}(x_i^g \beta, \Lambda_i + z_i V_b z_i^T).
\]

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

2.2.2. A MDM for non-monotone missing data

Longitudinal studies typically exhibit 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 transition model [23]. Let \( G_i = (G_{i1}, \ldots, G_{in}) \) denote the vector of missingness status variables for subject \( i \), where \( G_{ij} = 1, \ldots, n_i \) denotes the missingness status of subject \( i \) at occasion \( j \), and

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

[23] described an approach that writes the MDM as a product of transition probabilities:

\[
f_\tau(g_{i1}, \ldots, g_{in}, y_i, x_i) = \prod_{j=2}^{n_i} f_{\tau(j)}(g_{ij}|g_{i1}, \ldots, g_{i,j-1}, y_j, x_j).
\]

We typically assume that all units are observed at baseline; i.e., \( f(O|y_i, x_i) = 1 \). One can then model the remaining univariate conditional distributions separately. We note first that each conditional probability can include all past missingness status variables, thereby naturally incorporating information on how past missingness affects current missingness. Second, one needs to decide what variables in \( Y_j \) enter each conditional distribution. Let \( s_{ij} = (Y_{ij}^{obs}|x_j) \) be a matrix containing fully observed predictors for missingness up to visit \( j \) for subject \( i \), where \( Y_{ij}^{obs} \) includes all observed outcome measurements prior to visit \( j \) for subject \( i \). Because we condition on all past the missingness status variables within each past missingness pattern, \( Y_{ij}^{obs} \) 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 \( Y_{ij}^{obs} \) for those observations with that same past missingness pattern.

We further let a numerical variable \( u \) index missingness status with \( u = 0, 1, 2 \) representing \( O, I, D \), respectively. Our MDM then assumes the following conditional distribution of \( G_{ij} \):

\[
(G_j|S_j = s_j, Y_j = y_j, G_{ij} = g_{ij}) \sim \text{Multinomial}\left(1, \left[p_0^{s_0,1}, p_0^{s_0,2}, p_0^{s_0,3}\right]\right),
\]

where \( g_{ij} = (g_{ij,1}, \ldots, g_{ij,n}) \) denotes the past missingness pattern prior to visit \( j \), and the cell probabilities \( p_{ij}^{s_0,1}, p_{ij}^{s_0,2}, p_{ij}^{s_0,3} \) are
specified as
\[
p_{ij}^{\text{obs}(i)} = \frac{\phi_{ij}^{\text{obs}(i)}}{\sum_{u=0}^{2} \phi_{ij}^{u}}, \quad u = 0, 1, 2;
\]
with
\[
\phi_{ij}^{\text{obs}(i)}(s_{ij}, y_{ij}) = \exp\left(\gamma_{ij}^{\text{obs}(i)} s_{ij} + \gamma_{ij}^{\text{obs}(i)} y_{ij}\right).
\]
\[\text{(10)}\]

When \(\gamma_{ij}^{\text{obs}(i)} = 0\) for all values of \(u\) and \(g_{ij}\), the MDM does not depend on the potentially unobserved outcome and thus is MAR.

When \(\gamma_{ij}^{\text{obs}(i)} \neq 0\) for some \(u\) and \(g_{ij}\), 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., [181]) 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 [23].

Xie [23] applied the transitional MDM to a longitudinal dataset with non-monotone missingness arising from a design with \(n_{t} = n\); this enables conditioning on the entire past missingness vector \(g_{ij}\). In the setting with varying \(n_{t}\), conditioning on \(g_{ij}\) may be impossible. In this case, a convenient alternative is a first-order transitional model, where one assumes that, conditional on \((s_{ij}, y_{ij}, g_{ij-1})\), the missingness status \(G_{ij}\) at the current visit is independent of missingness status at all other prior visits; thus Eqn (9) reduces to

\[G_{ij}|S_{ij} = s_{ij}, Y_{ij} = y_{ij}, G_{ij-1} = v \sim \text{Multinomial}(1, [p_{ij}^{0v}, p_{ij}^{1v}, p_{ij}^{2v}]);\]
\[p_{ij}^{uv} = \frac{\phi_{ij}^{uv}}{\sum_{u=0}^{2} \phi_{ij}^{u}};\]
where \(u = 0, 1, 2; v = 0, 1,\) and
\[\phi_{ij}^{uv}(s_{ij}, y_{ij}) = \exp(\gamma_{ij}^{0uv} s_{ij} + \gamma_{ij}^{1uv} y_{ij}).\]
\[\text{(11)}\]

By the definition of dropout, \(G_{ij} = 2\) deterministically when \(G_{ij-1} = 2\) (the prior visit is a dropout); by the definition of intermittent missingness, \(\phi_{ij}^{00} = 0\); and because the response probabilities must sum to unity, for identification purposes \(\phi_{ij}^{00} = \phi_{ij}^{11} = 1\). 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 \((Y_{1}, ..., Y_{j-1})\)).

The conditional probability \(f_{Y}(g_{ij}|S_{ij}, Y_{ij}, G_{ij-1})\) then takes the form

\[f_{Y}(g_{ij}|S_{ij}, Y_{ij}, G_{ij-1}) = \begin{cases} 1 & \text{if } g_{i,j-1} = 0, g_{ij} = 0, \\ \frac{\exp(\gamma_{ij}^{00} s_{ij} + \gamma_{ij}^{10} y_{ij})}{1 + \exp(\gamma_{ij}^{00} s_{ij} + \gamma_{ij}^{10} y_{ij})} & \text{if } g_{i,j-1} = 0, g_{ij} \neq 0, \\ \frac{1}{1 + \exp(\gamma_{ij}^{00} s_{ij} + \gamma_{ij}^{10} y_{ij})} & \text{if } g_{i,j-1} = 1, g_{ij} = 0, \\ 0 & \text{if } g_{i,j-1} = 1, g_{ij} = 1, \\ \frac{\exp(\gamma_{ij}^{00} s_{ij} + \gamma_{ij}^{10} y_{ij})}{1 + \exp(\gamma_{ij}^{00} s_{ij} + \gamma_{ij}^{10} y_{ij})} & \text{if } g_{i,j-1} = 1, g_{ij} = 2, \\ 0 & \text{if } g_{i,j-1} = 2, g_{ij} \neq 2, \\ \frac{1}{1 + \exp(\gamma_{ij}^{00} s_{ij} + \gamma_{ij}^{10} y_{ij})} & \text{if } g_{i,j-1} = 2, g_{ij} = 2. \end{cases}\]

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

The nonorthogonality term \(V^{2}L_{\gamma_{1}}\) in Eqn (3) has been derived for longitudinal data with monotone [11] and non-monotone missingness patterns [23,26]; see Appendix B. For longitudinal models with the first-order transitional MDM and \(\gamma_{1} = (\gamma_{1}^{10}, \gamma_{1}^{20}, \gamma_{1}^{11})\), ISNI = \(\{\text{ISNI}_{1^{10}}, \text{ISNI}_{1^{10}}, \text{ISNI}_{1^{11}}\}\) and

\[
V^{2}L_{\gamma_{1}} = \begin{bmatrix} \frac{\partial^{2}L}{\partial \theta \partial \gamma_{1}^{10}} & \frac{\partial^{2}L}{\partial \theta \partial \gamma_{1}^{20}} & \frac{\partial^{2}L}{\partial \theta \partial \gamma_{1}^{11}} \\
\frac{\partial^{2}L}{\partial \theta \partial \gamma_{1}^{10}} & \frac{\partial^{2}L}{\partial \theta \partial \gamma_{1}^{20}} & \frac{\partial^{2}L}{\partial \theta \partial \gamma_{1}^{11}} \\
\frac{\partial^{2}L}{\partial \theta \partial \gamma_{1}^{10}} & \frac{\partial^{2}L}{\partial \theta \partial \gamma_{1}^{20}} & \frac{2\partial^{2}L}{\partial \theta \partial \gamma_{1}^{11}} \\
\end{bmatrix} \bigg|_{\hat{\theta}(0), \hat{\gamma}_{0}(0), \gamma_{1}=0}.
\]

where \(K_{i}\) is the length of \(y_{i}^{\text{obs}}\) and \(\hat{\theta}(0)\) and \(\hat{\gamma}_{0}(0)\) are MLEs of \(\theta\) and \(\gamma_{0}\) under the MAR assumption.

The term \(\partial^{2}E((Y_{i}^{\text{mis}})^{T}y_{i}^{\text{obs}})/\partial \theta\) is a \(p_{i}\times d_{i}\) matrix where \(d_{i}\) is the number of missing outcomes for subject \(i\). Under MAR in the complete-outcome models of Section 2.2.1, \((Y_{i}^{\text{mis}})^{T}y_{i}^{\text{obs}}\) is a Gaussian distribution involving \(\theta\) parameters only and \(\partial^{2}E((Y_{i}^{\text{mis}})^{T}y_{i}^{\text{obs}})/\partial \theta\) can be derived in a closed form involving matrix multiplication and differentiation as shown in Appendix C.

The term \(A_{i}^{10} = [A_{i}^{10}, A_{i}^{20}, A_{i}^{11}]\) is a \(d_{i} \times 3\) matrix. For the \(l\)th \((l = 1, \ldots, d_{i})\) component of \(y_{i}^{\text{mis}}\) that corresponds to the \(j\)th element of \(y_{i}\), one can show that [23,26]

\[
A_{i}^{10} = \frac{\partial^{2}E((Y_{i}^{\text{mis}})^{T}y_{i}^{\text{obs}})/\partial \theta}{\partial \theta}(\beta_{i}(0), \gamma_{0}=0),
\]
\[
A_{i}^{20} = \frac{\partial^{2}E((Y_{i}^{\text{mis}})^{T}y_{i}^{\text{obs}})/\partial \theta}{\partial \theta}(\beta_{i}(0), \gamma_{0}=0),
\]
\[
A_{i}^{11} = \frac{\partial^{2}E((Y_{i}^{\text{mis}})^{T}y_{i}^{\text{obs}})/\partial \theta}{\partial \theta}(\beta_{i}(0), \gamma_{1}=0).
\]

Here, \(p_{i}^{10}, p_{i}^{20}\) and \(p_{i}^{11}\) are the missingness status transitional probabilities from Eqn (11), calculated under MAR (i.e., with \(\gamma_{1} = 0\)). We note that because dropout and intermittent missingness are competing causes of missingness, \(A_{i}^{10}\) and \(A_{i}^{20}\) are of opposite sign when \(g_{i,j-1} = 0\).

When \(\gamma_{1}\) is a scalar, ISNI approximates the changes of the estimates when \(\gamma_{1}\) is perturbed from 0 to 1. In our model, the nonignorability parameter \(\gamma_{1}\) is the vector \((\gamma_{1}^{10}, \gamma_{1}^{20}, \gamma_{1}^{11})\). Thus, ISNI is a vector of three elements where each approximates the changes in estimates when only the corresponding element in \(\gamma_{1}\) 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 \( \frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_1} \bigg|_{\gamma_1=0} \gamma_1 \). When one is willing to assume that intermittent missingness and dropout have roughly the same nonignorability mechanism, we can fix \( \gamma_1^1, \gamma_1^{20}, \) and \( \gamma_1^{11} \) at the same largest perturbation value. In this scenario, \( \gamma_1 \) becomes a scalar and \( A_i \) a vector where \( A_i = A_i^{01} + A_i^{20} + A_i^{11} \) and the components of \( A_i \) become \( \frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_1} \), the estimated probability of being observed for the missing observations under ignorability. For its simplicity and ease of interpretation, this is the default method in our package.

Alternatively, one could consider all perturbations of the elements of \( \gamma_1 \) that are within a hypercube of size 1 from the origin. 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 index of sensitivity:

\[
\text{MISNI} = \sum_{i=1}^{q} \left| \frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_i} \right|_{\delta(0), \delta(0), \gamma_1=0},
\]

where \( q \) is the length of \( \gamma_1 \); here, \( q = 3 \). MISNI has the interpretation of maximum sensitivity, \( \max_{|\gamma_i|=1, i=1...n} \left| \frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_1} \right| \), when each element of \( \gamma_1 \) is perturbed between \(-1\) and \(1\). Our package also produces MISNI. [24] 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 \( q \) is small, the dimension of the nonignorability parameter, is large. An alternative is to consider a hyperball of size that is independent of \( q \) [5]. 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 Eqn (10), the nonignorability parameter \( \gamma_1 = \bigcup_{i \geq j} \gamma_1^{10(i-j-1)}, \gamma_1^{20(i-j-1)}, \gamma_1^{11(i-j-1)} \), and the dimensionality of \( \gamma_1 \) depends on the number of unique past missing data patterns and can be large. For a more parsimonious sensitivity analysis, one option is to reduce the number of nonignorability parameters such that \( \gamma_1^{10(i-j-1)} = \gamma_1^{20(i-j-1)} = \gamma_1^{11(i-j-1)} = \gamma_1^{11(i-j-1)}, j \) reduces \( \gamma_1 \) to \( \gamma_1^{10}, \gamma_1^{20}, \gamma_1^{11}, \gamma_1 \), 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: \( A_1^{10} = A_1^{11} = 0 \) and \( \partial E(Y_i^{10}) \partial \theta \) reduces to a \( p \times 1 \) vector. The calculation of \( \chi_2^2 \) then reduces to that of [11] and [20].

2.3. Calibrating ISNI

With a logit link in the MDM, a scalar \( \gamma_1 \) 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 a 2.7-fold increase 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. We describe below two calibration approaches that facilitate interpretation by creating a scale-free index.

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

The second approach is to approximate the minimum standard-
variable name for the missingness indicator $g_i$ and the missingness predictor $s_i$ in Eqn (5). Details appear in Section 3.2.

family describes the error distribution to be used in the GLM for the outcome $y$ in Eqn (4). The options are gaussian (the default), poisson, binomial, gamma and inverse.gaussian. The current version implements ISNI computation with the canonical link.
data names 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 is an optional vector of “prior weights” to be used in the fitting process for the complete-data model and the MDM.
subset is an optional vector specifying a subset of observations to be used in the fitting process for the outcome model and the MDM.
start is a starting value for the parameters in the linear predictor of the outcome model.
offset is an optional vector specifying an a priori known component in the linear predictor of the outcome GLM. It can be NULL or a numeric vector of length equal to the number of observations.

Function isnimgrm() computes ISNI for the MMGM:

```r
isnimgrm=function(formula, data, cortype="CS", id, subset, weights,
predprobs, misni=FALSE)
```

**formula** is 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 $x_{ij}$ 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 (11). Details appear in Section 3.2.
data is as above.
cortype describes the within-subject correlation structure of $\Sigma_i$ in Eqn (7). The currently-available options are “CS” (the default), “AR1”, and “UN” for compound symmetry, autoregressive order 1, and unstructured, respectively.
id designates the level-2 clustering variable.
weights is an optional vector of frequency weights to be assigned to each id. When absent, each id is weighted equally.
subset is as above.
predprobs is 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.
misni is FALSE if using the default approach to computing ISNI with a scalar nonignorability parameter, and TRUE when computing ISNI with multiple nonignorability parameters.

Function isnimlm() computes ISNI for the LMM:

```r
isnimlm=function(formula, data, random, id, weights, subset, predprobs, misni=FALSE)
```

**formula** is 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 $x_{ij}$ in Eqn (8). 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 (11). Details appear in Section 3.2.
data is as above.
random is a one-sided formula that specifies the random-effects part of the complete-data model LMM using the variable names for $z_i$ in Eqn (8). The program uses a general positive-definite variance-covariance matrix $V_g$ for the random effects.
id designates the level-2 clustering variable.
subset is as above.
weights is an optional vector of frequency weights to be assigned to each id. When absent, each id is weighted equally.
predprobs 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.
misni is FALSE if using the default approach to computing ISNI with a scalar nonignorability parameter, and TRUE when computing ISNI with multiple nonignorability parameters.

### 3.2. Model specification

One specifies the variables in the complete-data model and MDM in the formula argument. The user must supply at least a single equation for the complete-data model in the form response ~ Xterms where response is the (numeric or factor) vector for the outcome, and Xterms is a series of terms, separated by + operators, that specify the linear predictor. For the longitudinal setting, both isnimgrm and isnimlm by default use the utility function definermissingstatus provided in the package to generate the missingness status variables $g_{ij}$, and then use Xterms as the predictors $s_{ij}$ for fitting a first-order transitional MDM in Eqn (11). With longitudinal data, it is necessary to sort observations by time within id.

The single-equation model specification uses the $x_i$ defined for the complete-data model as $s_i$ in the MDM. To use different sets of predictors, one specifies a two-equation formula. For this purpose, the formula argument in the isni functions uses the R package Formula [28]. For isnimlm, one can specify the model with response | is.na(response) ~ Xterms | Stems, which designates the complete-data model as response ~ Xterms and the MDM under MAR as is.na(response) ~ Stems. For both isnimgrm and isnimlm, one can write response | miss + missprior ~ Xterms | Stems. This specifies the outcome model as response ~ Xterms, and gives the MDM of Eqn (11) as miss ~ missprior + Stems, where miss and missprior are the variable names in data denoting the missingness status at the current visit and prior visits, respectively, and Stems is the MDM predictor $s_{ij}$.

For isnimlm, response ~ Xterms identifies the fixed-effect part of the linear mixed-effects model for the outcome. One specifies the random-effect part of the model as a one-sided formula via the argument random.

### 3.3. Formatting data for an ISNI analysis

The argument data names the input data frame. The user must observe two rules when preparing it 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 the MDM. For independent data, the missingness status variable is simply an indicator for missingness and is generated as is.na(response) inside isnimlm if not provided in 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 in via the formula argument to estimate a first-order transitional MDM. isni provides a utility function
definemissingstatus() to generate the missingness status variables for users. isnimgrm and isnimmm call definemissingstatus() to generate missingness status variables if the user does not supply them. Again, it is necessary to sort beforehand by time within id.

Independent data. We array the data in a rectangular matrix, with rows representing units. Fields include the main outcome of interest with 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. Here we have multiple level-1 observations (e.g., repeated measures) within a level-2 unit. The name of the level-2 variable appears in the argument id. The level-1 observations for the unit occupy a set of rows identified by a common value of id. Missing values are denoted NA.

Observations with missing outcomes contribute to the sensitivity analysis through $\nabla^2 L_{\theta,y}$. Thus, the master dataset must include places for all planned observations, present or missing. This differs from a standard ignorable analysis, where often one can simply omit the missing observations. An exception occurs in longitudinal data with dropout, because when a subject leaves the study, the probability of observing outcomes is 0 for all subsequent times. Thus one can omit all visits after the dropout visit 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 prior visit. The order is irrelevant. Both models include an intercept by default.

If users supply the missingness status variable in the master dataset, it should be 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).

3.4. Handling missing values in covariates

Currently, isni assumes that predictors are fully observed and handles missing values in the supplied x and s as follows: Observations with missing values in x are excluded, and the program notes in the R console that it has dropped observations due to missing values in the outcome covariates. If in the remaining data there are missing values in s, to avoid further reducing the sample size the program detects and removes any s variables that contain missing values. In the future we plan to implement an alternative ISNI method that can incorporate observations with missing values in covariates; see the Discussion below.

4. Examples

Next we describe some applications of the isni package.

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

Raab and Donelly [14] 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?”. Many declined to answer, leading to substantial missing data. We consider a simplified data set consisting of the response variable, with the student’s sex and faculty as predictors.

```r
# 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 med
225 yes male other
5812 <NA> female other
5115 <NA> female other
769 yes male other
2366 <NA> male other
2622 <NA> male other
```

This code loads isni and the data frame sos, and displays 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?” (no [reference], yes); gender is the student’s sex (male [reference], female); faculty is the student’s faculty (medical/dental/veterinary, other [reference]).

We estimated an ignorable model predicting response from sex, faculty and their interaction:

```r
> 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 0.7560 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; the other factors are not statistically significant.
The columns “MAR Est.” and “Std. Err” denote the logistic model estimates and their standard errors under MAR; “ISNI” and “c” denote the ISNI and c statistics. The ISNIs are equal in absolute values to those reported in [17], but with opposite signs because our package models the probability that an observation is missing rather than the probability that it is observed. Recall that ISNI denotes the approximate change in the MLEs when γij in the selection model is changed from 0 to 1. Under our nonignorable selection model, the assumption γij = 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” are 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, γ1 = 1, the MLE for the estimate should change from −0.73 to −0.90. If γ1 = −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. Raab and Donnelly also found that neither the gender nor the gender:faculty interaction term should be sensitive, as our findings confirm.

In the above we do not explicitly specify an MDM. One can generate an identical analysis with the two-equation model formula sexact | is.na(sexact) ~ gender+faculty | gender+faculty, which uses the operator | to separately specify variables used in the complete-data model and the MDM:

> summary(sos.sin)

Call:
  isignlm(formula=swmodel, family=binomial, data=sos)

MLM Est. Std. Err ISNI c
(Intercept) 0.293898 0.599624 0.56584 0.7034
genderfemale 0.00581 0.027804 0.00233 0.0255

4.2. Illustration: An ISNI analysis of simulated data

Here we illustrate ISNI analysis using simulated data with known MDMs. Specifically, we generated one complete data set with a sample size of 10,000 under the logistic model \( \logit(Pr(Y = 1|x)) = \beta_0 + \beta_1 x \) with \( \beta_0 = 1, \beta_1 = -0.7 \). The covariate x was generated as Bernoulli with probability 0.5. The MDM was the logistic regression \( \logit(Pr(G = 1|Y)) = \gamma_0 + \gamma_1 Y + \gamma_2 Y \gamma_1 Y \). We simulated one missingness pattern under each of the four parameter settings \( \gamma = (\gamma_0, \gamma_1, \gamma_2) \). We applied the isignlm() to each simulated dataset; results appear in Table 1.

The columns “Est” and “SE” denote the outcome logistic model estimate for \( \beta_0 \) or \( \beta_1 \) (as listed under “0”) and its standard error under MAR; “ISNI” and “c” denote the missing proportion, ISNI, and c statistic, respectively. One can use ISNI/γi to approximate the change in the estimate under the posited value of γi. For example, using the true value \( \gamma_1 = -1 \) from Dataset C, the estimate of \( \beta_0 \) changes to Est + ISNI × \( \gamma_1 \) = 1.651 + 0.607 × (−1) = 1.044, which is close to the true value of 1.0; the estimate of \( \beta_1 \) changes to −0.782 + (−0.81) × (−1) = −0.701, which is close to the true value of −0.7. Within each dataset, the estimates of \( \beta_0 \) and \( \beta_1 \) can have quite different ISNI and c values, suggesting that the missing proportion alone is insufficient to capture the individual sensitivity of different model parameters. The four datasets are sorted by missing proportion from low to high. We observe that for both \( \beta_0 \) and \( \beta_1 \), the absolute ISNI value increases and c value decreases as the fraction of data missing increases.

4.3. ISI analysis of an MMGM for longitudinal data

Ma et al. [11] analyzed data from a randomized trial comparing fluoxetine with a placebo in the treatment of prostate cancer. A sub-study collected QoL outcomes at baseline and 1, 3, and 6 months after randomization. We focus here on the emotional functioning (EF) scale outcome. Below is a sample of the data set qoef.
Table 1
ISNI Analysis Using Simulated Data From a Logistic Regression: \( \text{logit}(\Pr(Y = 1)) = \beta_0 + \beta_1 x \) with \( \beta_0 = 1 \) and \( \beta_1 = -0.7 \).

<table>
<thead>
<tr>
<th>Data</th>
<th>( y_{00} )</th>
<th>( y_{01} )</th>
<th>( y_{11} )</th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>SE</th>
<th>ISNI</th>
<th>( c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-0.2</td>
<td>-0.5</td>
<td>1</td>
<td>1.089</td>
<td>-0.754</td>
<td>0.092</td>
<td>0.074</td>
<td>1.090</td>
</tr>
<tr>
<td>B</td>
<td>-0.2</td>
<td>0</td>
<td>0.9</td>
<td>0.986</td>
<td>-0.740</td>
<td>0.034</td>
<td>0.117</td>
<td>0.287</td>
</tr>
<tr>
<td>C</td>
<td>1.2</td>
<td>-0.5</td>
<td>0.567</td>
<td>1.651</td>
<td>-0.782</td>
<td>0.140</td>
<td>0.067</td>
<td>0.230</td>
</tr>
<tr>
<td>D</td>
<td>1.2</td>
<td>0</td>
<td>0.722</td>
<td>1.028</td>
<td>-0.756</td>
<td>0.071</td>
<td>0.773</td>
<td>0.092</td>
</tr>
</tbody>
</table>

Table 2
Summary Statistics in the Prostate Cancer QoL Dataset.

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

Table 3
Missing Patterns in the Prostate Cancer QoL Dataset.

<table>
<thead>
<tr>
<th>EFO</th>
<th>EF1</th>
<th>EF2</th>
<th>EF3</th>
<th>Placebo (n = 352)</th>
<th>Flutamide (n = 363)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>239 67.8</td>
<td>258 71.1</td>
</tr>
<tr>
<td>P</td>
<td>P</td>
<td>P</td>
<td>A</td>
<td>38   10.7</td>
<td>37   10.2</td>
</tr>
<tr>
<td>P</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>19   5.4</td>
<td>14   3.8</td>
</tr>
<tr>
<td>P</td>
<td>A</td>
<td>A</td>
<td>P</td>
<td>20   5.6</td>
<td>22   6.1</td>
</tr>
<tr>
<td>P</td>
<td>A</td>
<td>P</td>
<td>P</td>
<td>13   3.7</td>
<td>13   3.6</td>
</tr>
<tr>
<td>P</td>
<td>A</td>
<td>P</td>
<td>A</td>
<td>11   3.1</td>
<td>15   4.1</td>
</tr>
<tr>
<td>P</td>
<td>A</td>
<td>A</td>
<td>P</td>
<td>7    2.0</td>
<td>1    0.3</td>
</tr>
<tr>
<td>P</td>
<td>A</td>
<td>A</td>
<td>P</td>
<td>5    1.4</td>
<td>3    0.8</td>
</tr>
</tbody>
</table>

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

Gaussian model, using function isnimgm(). We take the predictor vector to be

\[ X_i = (\text{Intercept}, \text{perf}, \text{sever}, \text{T1.0(pb)}, \text{T3.0(pb)}, \text{T6.0(pb)}). \]

The two predictors "perf" and "sever" are baseline covariates. The predictor T0(d) is an indicator for the baseline observation in arm a. The predictors Tr0(d) 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. One can specify the complete-data model as \( y_i = \text{perf} + \text{sever} + \text{as.factor}(\text{time}) + \text{group} + \text{as.factor}(\text{time})\cdot\text{group} \). To evaluate the robustness of the MAR analysis to nonignorable missingness, we assume the first-order transitional model of Eqn (11), where the missingness status variables at the current visit and at the previous visit and at the current time point are in qolef, respectively; the missingness predictor \( s_i \) 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:

> qolef().

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=internally missing)
- **gp** — missingness status in the prior visit (as above, plus 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); as in [11], we transform to the square-root scale. Table 2 presents the mean and SD of transformed EF and 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 of the subjects missing by the final visit. Our analysis excludes a small number of patients (≈ 3%) whose EF data were missing at baseline. Table 3 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.

[23] 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...
isnimgm() estimates the MAR model using gls() in R package nlme. The ignorable analysis suggests that, after adjustment for baseline performance and severity, placebo gives statistically significantly better EF at all three follow-up visits (estimates for as.factor(time)1:group, as.factor(time)3:group, and as.factor(time)6:group are significant and negative). "ISNI" values quantify the potential change from the MAR estimates with $\gamma_1 = 1$. 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 nonignorability. The treatment contrasts at follow-up – as.factor(time)1:group, as.factor(time)3:group, and as.factor(time)6:group – are insensitive, with $c < 1$.

One can designate the model formula without identifying the missingness variable. In this case, users should first sort data by time within id:

> summary(qolef.isni)

Call:
  isnimgm(formula = model, data = qolef, id = id)

<table>
<thead>
<tr>
<th></th>
<th>MAR Est.</th>
<th>Std. Err</th>
<th>ISNI</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>8.428061</td>
<td>0.102070</td>
<td>-0.302022</td>
<td>4.5088</td>
</tr>
<tr>
<td>perf</td>
<td>-0.239914</td>
<td>0.216802</td>
<td>0.976118</td>
<td>2.9650</td>
</tr>
<tr>
<td>severe</td>
<td>-0.154242</td>
<td>0.097230</td>
<td>0.346654</td>
<td>2.7543</td>
</tr>
<tr>
<td>as.factor(time)1</td>
<td>0.465742</td>
<td>0.064515</td>
<td>0.115753</td>
<td>0.7440</td>
</tr>
<tr>
<td>as.factor(time)3</td>
<td>0.488756</td>
<td>0.082559</td>
<td>0.139521</td>
<td>0.7899</td>
</tr>
<tr>
<td>as.factor(time)6</td>
<td>0.385020</td>
<td>0.094246</td>
<td>0.206403</td>
<td>0.6095</td>
</tr>
<tr>
<td>group</td>
<td>-0.012408</td>
<td>0.103233</td>
<td>-0.001340</td>
<td>0.9458</td>
</tr>
<tr>
<td>as.factor(time)1:group</td>
<td>-0.227693</td>
<td>0.090418</td>
<td>-0.075279</td>
<td>16.0341</td>
</tr>
<tr>
<td>as.factor(time)3:group</td>
<td>-0.215777</td>
<td>0.115774</td>
<td>-0.019549</td>
<td>7.8889</td>
</tr>
<tr>
<td>as.factor(time)6:group</td>
<td>-0.245304</td>
<td>0.131450</td>
<td>-0.009923</td>
<td>136.5847</td>
</tr>
<tr>
<td>sigma</td>
<td>1.273051</td>
<td>0.003007</td>
<td>0.215437</td>
<td>0.7526</td>
</tr>
<tr>
<td>rho</td>
<td>0.941949</td>
<td>0.014669</td>
<td>-0.009595</td>
<td>2.0976</td>
</tr>
</tbody>
</table>

By default, isnimgm() uses the compound symmetry correlation structure in the outcome model of Eqn (7). 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.

> qolef.isni.ar = isnimgm(model, id=id, cortype = 'AR1', data=qolef)

> summary(qolef.isni.ar)

Call:
  isnimgm(formula = model, data = qolef, cortype = 'AR1', id = id)

<table>
<thead>
<tr>
<th></th>
<th>MAR Est.</th>
<th>Std. Err</th>
<th>ISNI</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>8.434487</td>
<td>0.105008</td>
<td>-0.205590</td>
<td>5.3000</td>
</tr>
<tr>
<td>perf</td>
<td>-0.301055</td>
<td>0.224305</td>
<td>0.091956</td>
<td>3.2763</td>
</tr>
<tr>
<td>severe</td>
<td>-0.150384</td>
<td>0.102388</td>
<td>0.291823</td>
<td>4.6828</td>
</tr>
<tr>
<td>as.factor(time)1</td>
<td>0.456740</td>
<td>0.074576</td>
<td>0.135182</td>
<td>0.7369</td>
</tr>
<tr>
<td>as.factor(time)3</td>
<td>0.474532</td>
<td>0.077807</td>
<td>0.129672</td>
<td>0.8010</td>
</tr>
<tr>
<td>as.factor(time)6</td>
<td>0.365058</td>
<td>0.081839</td>
<td>0.172886</td>
<td>0.6319</td>
</tr>
<tr>
<td>group</td>
<td>-0.001957</td>
<td>0.101797</td>
<td>-0.000923</td>
<td>136.5847</td>
</tr>
<tr>
<td>as.factor(time)1:group</td>
<td>-0.226479</td>
<td>0.104531</td>
<td>-0.008104</td>
<td>16.8090</td>
</tr>
<tr>
<td>as.factor(time)3:group</td>
<td>-0.203433</td>
<td>0.108956</td>
<td>-0.013274</td>
<td>9.5456</td>
</tr>
<tr>
<td>as.factor(time)6:group</td>
<td>-0.213838</td>
<td>0.114211</td>
<td>-0.024300</td>
<td>6.2743</td>
</tr>
<tr>
<td>sigma</td>
<td>1.355358</td>
<td>0.025921</td>
<td>0.006980</td>
<td>5.1661</td>
</tr>
<tr>
<td>cor(1,2)</td>
<td>0.508482</td>
<td>0.024711</td>
<td>-0.006580</td>
<td>5.0225</td>
</tr>
<tr>
<td>cor(1,3)</td>
<td>0.472153</td>
<td>0.026089</td>
<td>0.016597</td>
<td>21.1662</td>
</tr>
<tr>
<td>cor(1,4)</td>
<td>0.445833</td>
<td>0.028275</td>
<td>0.006860</td>
<td>5.0178</td>
</tr>
<tr>
<td>cor(2,3)</td>
<td>0.707501</td>
<td>0.018507</td>
<td>0.007925</td>
<td>13.1171</td>
</tr>
<tr>
<td>cor(2,4)</td>
<td>0.595332</td>
<td>0.027904</td>
<td>0.007357</td>
<td>4.0171</td>
</tr>
<tr>
<td>rho</td>
<td>0.734721</td>
<td>0.016464</td>
<td>-0.004397</td>
<td>4.9979</td>
</tr>
</tbody>
</table>

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

To posit an MDM other than that of Eqn (11), one can supply predicted probabilities of being observed under the desired model through the optional argument predprob. The R code below uses function tmdm() to obtain the predicted probabilities of being observed for all observations and passes them in as the vector predprobs. isnimgm() then refrains from fitting the default first-order MDM and instead uses predprobs to evaluate ISNI:
The QoL dataset had both intermittent missingness (≈ 10%) and dropout (≈ 20%). In some applications, the missingness can be of only one type or the other, and thus the MDM reduces to a special case of the more general multimonial transitional model 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 contains only complete observations and those with intermittent missingness.

> # Run ISNI analysis on the subset that excludes intermittent missingness.
> qolef.isni.isnimgm(formula = models, id = id, data = qolef, subset = gl("T"))
> summary(qolef.isni)

<table>
<thead>
<tr>
<th></th>
<th>MAR Est.</th>
<th>Std. Err</th>
<th>ISNI c</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>perf</td>
<td>-0.270986</td>
<td>0.219348</td>
<td>0.5255</td>
</tr>
<tr>
<td>severe</td>
<td>-0.146629</td>
<td>0.100085</td>
<td>0.6866</td>
</tr>
<tr>
<td>as.factor(time)1</td>
<td>0.452168</td>
<td>0.070566</td>
<td>0.3383</td>
</tr>
<tr>
<td>as.factor(time)3</td>
<td>0.489599</td>
<td>0.071693</td>
<td>0.1702</td>
</tr>
<tr>
<td>group</td>
<td>-0.202108</td>
<td>0.099713</td>
<td>0.5534</td>
</tr>
<tr>
<td>sigma</td>
<td>1.328627</td>
<td>0.025275</td>
<td>0.2869</td>
</tr>
<tr>
<td>rho</td>
<td>0.548157</td>
<td>0.019662</td>
<td>0.3662</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>MAR Est.</th>
<th>Std. Err</th>
<th>ISNI c</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>perf</td>
<td>-0.270986</td>
<td>0.219348</td>
<td>0.5255</td>
</tr>
<tr>
<td>severe</td>
<td>-0.146629</td>
<td>0.100085</td>
<td>0.6866</td>
</tr>
<tr>
<td>as.factor(time)1</td>
<td>0.452168</td>
<td>0.070566</td>
<td>0.3383</td>
</tr>
<tr>
<td>as.factor(time)3</td>
<td>0.489599</td>
<td>0.071693</td>
<td>0.1702</td>
</tr>
<tr>
<td>group</td>
<td>-0.202108</td>
<td>0.099713</td>
<td>0.5534</td>
</tr>
<tr>
<td>sigma</td>
<td>1.328627</td>
<td>0.025275</td>
<td>0.2869</td>
</tr>
<tr>
<td>rho</td>
<td>0.548157</td>
<td>0.019662</td>
<td>0.3662</td>
</tr>
</tbody>
</table>

ISNIs from both summary(qolef.isni) and summary(qolef.isni) are generally smaller than those from the original dataset as shown in summary(qolef.isni). This is expected, because sensitivity increases with the proportion of missingness in a dataset.

Finally, we illustrate the computation of MISNI. As compared with summary(qolef.isni), which considers a scalar $Y_{1}$, MISNI shows a slight increase in sensitivity:

> qolef.isni.isnimgm(formula = models, id = id, data = qolef, misni = T)
> summary(qolef.isni)

<table>
<thead>
<tr>
<th></th>
<th>MAR Est.</th>
<th>Std. Err</th>
<th>ISNI c</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>perf</td>
<td>-0.270986</td>
<td>0.219348</td>
<td>0.5255</td>
</tr>
<tr>
<td>severe</td>
<td>-0.146629</td>
<td>0.100085</td>
<td>0.6866</td>
</tr>
<tr>
<td>as.factor(time)1</td>
<td>0.452168</td>
<td>0.070566</td>
<td>0.3383</td>
</tr>
<tr>
<td>as.factor(time)3</td>
<td>0.489599</td>
<td>0.071693</td>
<td>0.1702</td>
</tr>
<tr>
<td>group</td>
<td>-0.202108</td>
<td>0.099713</td>
<td>0.5534</td>
</tr>
<tr>
<td>sigma</td>
<td>1.328627</td>
<td>0.025275</td>
<td>0.2869</td>
</tr>
<tr>
<td>rho</td>
<td>0.548157</td>
<td>0.019662</td>
<td>0.3662</td>
</tr>
</tbody>
</table>

In the first-order transitional MDM, ISNI= (ISNI$_{10}$, ISNI$_{20}$, ISNI$_{11}$) (Eqn (13)) is a sensitivity vector whose components approximate the change in the estimate when the corresponding element in $Y_{1}$ is perturbed from 0 to 1. The above ISNI analysis provides parsimonious measures that combine these three elements to quantify sensitivity to both nonignorable dropout and intermittent missingness. The program outputs the sensitivity vector as a component called “isnivec” for the object “isnimgm” (and also the object “isnimm” described in the next subsection), which allows one to assess sensitivity by missingness types as follows:

> qolef.isni.isnivec

| ISNI ID  | ISNI DO | ISNI II | (Intercept) | -0.0003187537 | -0.0265677652 | 9.8706016-05 | perf | -0.0010973869 | 0.014130923 | -0.00359532-03 | severe | 0.0010998801 | 0.025008814 | 5.665167-05 | as.factor(time)1 | 0.0666836584 | 0.0713312349 | 3.290869-03 | as.factor(time)3 | 0.0349565621 | 0.082477934 | 1.0430096-02 | as.factor(time)6 | -0.0002040494 | 0.1822219019 | -8.609470-06 | group | -0.0015571978 | -0.009281642 | -6.248959-05 | as.factor(group)1 | -0.0145586569 | 0.0013932094 | -1.280454-03 | as.factor(group)3 | -0.0015131409 | -0.031899080 | -5.706744-03 | as.factor(group)6 | -0.0003079072 | -0.0282693641 | 6.400424-04 | sigma | 0.0007979717 | 0.0081877743 | -8.421882-05 | rho | 0.0010640228 | -0.008586857 | -1.182775-04 |

The column names “ISNI_1O”, “ISNI_DO” and “ISNI_II” represent sensitivity associated with nonignorability parameter $Y_{10}$, $Y_{20}$, $Y_{11}$ respectively. Thus they indicate sensitivity caused by missingness configuration “IO”, “DO” and “II”, respectively with the first and second letter inside each quote denoting the missingness types (O (observed), I (intermittently missing), and D (dropout)) in the current and prior visits, respectively. One can also assess the total sensitivity to intermittent missingness by combining “ISNI_1O” and “ISNI_II”.

4.4. ISNI analysis of an LMM for longitudinal data

Finally, we illustrate the LMM analysis using the 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 and ISNI results.

> data(qolef)
> models = l g + gp + perf + severe + as.factor(time) + group + as.factor(time) + gp + as.factor(time) + gp + as.factor(time) + gp + as.factor(time) + gp + severe
> isnm = MISNI(models, random = 1, id = id, data = qolef)
> summary(results)

<table>
<thead>
<tr>
<th></th>
<th>MAR Est.</th>
<th>Std. Err</th>
<th>ISNI c</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>perf</td>
<td>-0.270986</td>
<td>0.219348</td>
<td>0.5255</td>
</tr>
<tr>
<td>severe</td>
<td>-0.146629</td>
<td>0.100085</td>
<td>0.6866</td>
</tr>
<tr>
<td>as.factor(time)1</td>
<td>0.452168</td>
<td>0.070566</td>
<td>0.3383</td>
</tr>
<tr>
<td>as.factor(time)3</td>
<td>0.489599</td>
<td>0.071693</td>
<td>0.1702</td>
</tr>
<tr>
<td>group</td>
<td>-0.202108</td>
<td>0.099713</td>
<td>0.5534</td>
</tr>
<tr>
<td>sigma</td>
<td>1.328627</td>
<td>0.025275</td>
<td>0.2869</td>
</tr>
<tr>
<td>rho</td>
<td>0.548157</td>
<td>0.019662</td>
<td>0.3662</td>
</tr>
</tbody>
</table>

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 \( c > 1 \), suggesting that sensitivity to nonignorability is not of concern for this estimate. These results echo our findings under the MMGM.

```r
# Random intercept and slope model
> model1 = y ~ Ig + gp ~ time*group + perf + sever
> summary(lm(formula = model1, random = ~ 1 + time, id=id, data=qolef))
```

<table>
<thead>
<tr>
<th>MAR Est.</th>
<th>Std. Err</th>
<th>ISNI</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>8.633868</td>
<td>0.099126</td>
<td>0.0134140</td>
</tr>
<tr>
<td>time</td>
<td>0.067976</td>
<td>0.013094</td>
<td>0.0294745</td>
</tr>
<tr>
<td>group</td>
<td>-0.103960</td>
<td>0.092752</td>
<td>-0.003128</td>
</tr>
<tr>
<td>perf</td>
<td>-0.263945</td>
<td>0.221257</td>
<td>0.1001359</td>
</tr>
<tr>
<td>sever</td>
<td>-0.152619</td>
<td>0.099816</td>
<td>0.0326422</td>
</tr>
<tr>
<td>time:group</td>
<td>-0.032053</td>
<td>0.018209</td>
<td>-0.0064265</td>
</tr>
<tr>
<td>signav1</td>
<td>1.043033</td>
<td>0.039745</td>
<td>-0.0124504</td>
</tr>
<tr>
<td>signav2</td>
<td>0.118714</td>
<td>0.014123</td>
<td>0.0025849</td>
</tr>
<tr>
<td>rho12</td>
<td>-0.327000</td>
<td>0.080491</td>
<td>-0.0346455</td>
</tr>
<tr>
<td>sigmain</td>
<td>0.855890</td>
<td>0.017576</td>
<td>0.0045477</td>
</tr>
</tbody>
</table>

### 4.5. Hardware and software specifications

**ISNI** is written in R, and can be used under Windows, Linux and Macintosh. As a requirement to pass the R CRAN submission tests, for no individual example in the package does the ISNI computation time exceed 5 seconds.

### 4.6. Mode of availability

**ISNI** is freely available at the R CRAN repository (https://CRAN.R-project.org/package=ISNI) and can be installed as an add-on within the R console.

### 5. Discussion

There is a pressing need for statistical software for analysis of sensitivity to departures from the assumption of ignorable missingness. Our new R package **ISNI** implements such computations in three common settings: The GLM for univariate data and the MMGM and LMM for longitudinal/clustered data. Our next planned version will cover the generalized linear mixed model [20]. Other potential extensions are to ISNI analysis under the general coarse-data model [6,10,29–31] and to Bayesian versions of ISNI [21,31].

We emphasize that ISNI is a measure of local sensitivity, in that it quantifies the extent to which MLEs should change to reflect small departures from MAR. It is not intended to detect or estimate nonignorability; rather, it alerts the user to situations where, should nonignorability exist, estimates under an MAR model are unreliable.

ISNI is an increasing function of the fraction of missing observations, the predicted probabilities of being observed for the missing observations, and the influence of the missing observations on estimates. The most readily apparent component of sensitivity is the fraction of missingness; indeed, in the estimation of a univariate normal mean, ISNI is proportional to the fraction of missing observations [17]. The positioning of missing information in the data set also affects the value of ISNI. For example, holding the number of missing observations fixed, ISNI for a slope is large if the missing data are concentrated among high-influence points, but small if they are concentrated among low-influence points; see [17]. Note also that ISNI \( i = 0 \). and thus \( c = \infty \), whenever there are no missing observations, regardless of ignorability of the underlying true MDM.

In practice, the most effective measure one can take to reduce sensitivity is to avoid missing observations. Failing that, the next best thing is to prospectively gather robust information on the causes of missingness, which can then inform judgments on the plausible degree of nonignorability in the data.

As with regression influence diagnostics [1], ISNI is concerned only with the data at hand — defined here as \( \{y^{obs}, g\} \) — and not with other data sets that might have been gathered but were not. From this perspective, the sampling distribution of ISNI is irrelevant. Again, consider the case of a univariate normal mean, where ISNI is proportional to the fraction of missing observations. Clearly, it matters not that this proportion may have been higher or lower; for the analysis at hand, it only matters what the proportion is.

When ISNI identifies an estimate as sensitive to nonignorability, the investigator typically has only two choices: Flag the MAR analysis as unreliable, or conduct a speculative nonignorable analysis. When further data collection is possible, one may have the additional option of taking a refreshment sample [7], which can shed light on the MDM and inform a more robust nonignorable modeling exercise. In such a situation, ISNI can be useful in identifying targets for additional data collection. For example, an ISNI analysis of US young adult smoking data [26] revealed that the impact of nonignorable missingness was largest in black males, suggesting the need for more persistent efforts to acquire data in that group.

One cannot always distinguish intermittent missing data from true dropout on the basis of the missingness pattern alone. For example, a missing observation at only the final visit could represent dropout (abandoning the study) or intermittent missing data (skipping a single visit). Accurate analysis thus depends not just on specifying a suitably flexible MDM, but on correctly gauging the reasons for missed visits. Ideally, one would obtain such data directly from the subjects; if this is not possible, one can compute the range of ISNIs for possible combinations of missingness types [23].

Our use of a logistic model for \( f_Y(g|y) \) raises the concern that ISNI itself may be sensitive to untestable assumptions about the MDM. Although investigations to date have suggested that results are, as a practical matter, insensitive to the link function and the form of the linear predictor [25–27], robustness is not infinite. Therefore it may be desirable to extend the package to incorporate generalizations of the basic MDM [25,27]. We caution, however, that this is not simply a matter of model fit; the interpretation of the logistic \( \gamma_1 \) as a log odds ratio at every level of \( y \) greatly simplifies the conceptualization of a sensitivity analysis.

Another limitation of our package is that it allows missing data in the outcome only, and therefore it uses an ad hoc approach to repairing the data for analysis when there is missingness in the predictors. The method of [5], which addresses this problem to some extent, has not yet been implemented in a distribution package. With many observational data sets, a variable may be an “outcome” in one study and a “covariate” in another, depending on the context and objectives. Thus it is desirable to have methods that can assess sensitivity in a principled way regardless of the configuration of missing items. A general solution is to construct a comprehensive model for all the random variables, and to conduct an ISNI analysis for the parameters of the conditional distribution of the “outcomes” given the “predictors”.

There are various other approaches to the analysis of sensitivity to nonignorability. [19] proposed a model that generates individual-level influence diagnostics. Recently, [12] proposed an approach that uses imputations from pattern-mixture models to assess how strong departures from MAR must be in order to affect primary conclusions. Although these approaches differ from ours in form and concept, they could be valuable components of a comprehensive sensitivity analysis package.
Appendix A. Derivation of ISNI

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

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

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

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

Thus for any $\gamma_1$, we have

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

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

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

Appendix B. Derivation of $V^2L_{\theta, \gamma_1}$

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 $Y_i^{\text{obs}}$ refers to the components of $Y_i$ that are observed, and $Y_i^{\text{mis}}$ refers to the components of $Y_i$ that are missing. Let $K_i$ be the length of $Y_i^{\text{obs}}$. If subject $i$ completed all the intended visits of the study, then $K_i = n_i$, and $Y_i^{\text{mis}}$ 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 2.2.2 and 2.2.2. Then

$$L(\theta, \gamma_0, \gamma_1) = \sum_{i=1}^{N} L_i(\theta, \gamma_0, \gamma_1 ; \gamma_i^{\text{obs}}, \gamma_i^{\text{mis}}, \xi_i),$$

where $\gamma_i = (\gamma_{i1}, \ldots, \gamma_{in})$ is a vector of discrete variables for the missingness status of subject $i$: $f_0(y_i^{\text{obs}}, y_i^{\text{mis}} | \gamma_i)$ is the density function of the outcome model defined above; $f_y(\gamma_i | s_i, \xi_i, \gamma_i^{\text{mis}}, \gamma_i^{\text{obs}})$ is the probability mass function of the missing-data model defined in Eqn (12), and if the general transitional model as specified in Eqn (9) is used for modeling $\gamma_i$, $f_y(\gamma_i | s_i, \xi_i, \gamma_i^{\text{mis}}, \gamma_i^{\text{obs}})$ is then replaced with $f_y(\gamma_i | s_i, \xi_i, \gamma_i^{\text{mis}}, \gamma_i^{\text{obs}})$. We intend the integral sign to refer to summation with discrete outcomes.

The components of $y_i^{\text{mis}}$ after dropout do not enter the integral in Eqn (17), because these outcomes are deterministically missing. Thus, the dimensionality of the integration for the ith unit is $d_i = \sum_j l(\xi_{ij} = 1) + l(\gamma_{ij} = 2)$. Henceforth, the notation $y_i^{\text{mis}}$ includes only the intermittent missing outcomes and the outcome at the time of dropout. With nonignorable missingness, the integral with respect to $y_i^{\text{mis}}$ 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 $V^2L_{\theta, \gamma_1}$, we note that $V^2L_{\theta, \gamma_1} = (V^2L_{\theta, \gamma^{01}}, V^2L_{\theta, \gamma^{02}}, V^2L_{\theta, \gamma^{11}})$, where

$$V^2L_{\theta, \gamma^{ij}} = \sum_{i,k=1}^{\text{mis}} \sum_{\text{mis}} \frac{\partial^2}{\partial \theta^{ij} \partial \theta^{ij}} \ln \left( \frac{\int \prod_{j=1}^{n_i} f_y(y_i^{\text{obs}}, x_i^{\text{mis}}, y_i^{\text{mis}} | \gamma_i) \prod_{j=1}^{n_i} f_y(s_i, \xi_i, y_i^{\text{mis}}, \gamma_i | y_i^{\text{mis}}) \prod_{j=1}^{n_i} f_y(\gamma_i | s_i, \xi_i, y_i^{\text{mis}}, y_i^{\text{obs}}) dy_i^{\text{mis}}}{\int \prod_{j=1}^{n_i} f_y(y_i^{\text{obs}}, x_i^{\text{mis}}, y_i^{\text{mis}} | \gamma_i) \prod_{j=1}^{n_i} f_y(s_i, \xi_i, y_i^{\text{mis}}, \gamma_i | y_i^{\text{mis}}) \prod_{j=1}^{n_i} f_y(\gamma_i | s_i, \xi_i, y_i^{\text{mis}}, y_i^{\text{obs}}) dy_i^{\text{mis}}} \right) \left|_{\gamma^{ij} = 0} \right. \left|_{\gamma^{ij} = 0} \right.$$
If the $l$th component of $y_{i}^{\text{mis}}$ corresponds to the $j$th element of $y_i$, the $l$th element of $A_i^{10}$ is

$$A_{ij}^{10} = \frac{I(g_{i,l-1} = 0, p_{i,l}, j, j_{0}, 0, j_{1}, 0, 0, 0)}{p_{i,l}},$$

$$= I(g_{i,l-1} = 0, p_{i,l}, j, j_{0}, 0, j_{1}, 0, 0, 0),$$

where $\nabla^{2}L_{0}^{(10)}$ and $\nabla^{2}L_{0}^{(11)}$ are derived similarly to Eqn (18) with $A_i^{10}$ replaced by $A_i^{10}$ and $A_i^{11}$, respectively, where

$$A_{ij}^{20} = I(g_{i,l-1} = 0, p_{i,l}, j, j_{0}, 0, j_{1}, 0, 0, 0),$$

$$A_{ij}^{11} = I(g_{i,l-1} = 0, p_{i,l}, j, j_{0}, 0, j_{1}, 0, 0, 0).$$

**Appendix C. Derivation of $\frac{\partial E(Y_{i}^{\text{mis}}|Y_i^{\text{obs}}, x_i)}{\partial \theta}$ | $\gamma_i = 0$**

We note that $y_i^{\text{mis}}$ is a vector of length $d_i = \sum_j I(g_{ij} = 1) + I(\text{any of } j_k = 2)$. MMGM. Because $E((Y_i^{\text{mis}})^{T}|y_i^{\text{obs}}, x_i)|\gamma_i = 0 = \theta_i^T X_{iM} + (J_i^{\text{obs}})^{T} - \theta_i^T X_{iO}^{T} \Sigma_{1,00} \Sigma_{1,0M}$, by vector differentiation, we have

$$\frac{\partial E((Y_i^{\text{mis}})^{T}|y_i^{\text{obs}}, x_i)}{\partial \theta_i} | \gamma_i = 0 = X_{iM} - X_{iO}^{T} \Sigma_{1,00} \Sigma_{1,0M}^{-1} \Sigma_{1,OM},$$

$$\frac{\partial E((Y_i^{\text{mis}})^{T}|y_i^{\text{obs}}, x_i)}{\partial \theta_i} | \gamma_i = 0 = -(J_i^{\text{obs}})^{T} - \theta_i^T X_{iO}^{T} \Sigma_{1,00} \Sigma_{1,OM}^{-1} \Sigma_{1,OM},$$

$$+ (J_i^{\text{obs}})^{T} - \theta_i^T X_{iO}^{T} \Sigma_{1,00} \Sigma_{1,OM}^{-1} \frac{\partial \Sigma_{1,OM}}{\partial \theta_i} \Sigma_{1,OM}^{-1} \Sigma_{1,OM},$$

where $X_{iO}$ and $X_{iM}$ are the predictor matrices for $y_i^{\text{obs}}$ and $y_i^{\text{mis}}$, respectively, and

$$\text{Var}(Y_i^{\text{mis}}, y_i^{\text{obs}} | x_i) = \left( \begin{array}{cc} \Sigma_{1,00} & \Sigma_{1,OM} \\ \Sigma_{1,MO} & \Sigma_{1,MM} \end{array} \right).$$

**C.2 LMM.** One can rewrite the model of Eqn (8) as a marginal multivariate normal model with $\Sigma_i = \Lambda_i + Z_i \Sigma_{iZ} Z_i^{T}$.

Then we apply the above result to obtain

$$\frac{\partial \Sigma_i}{\partial \theta_i} = \frac{\partial \Lambda_i}{\partial \theta_i} + Z_i \frac{\partial \Sigma_{iZ}}{\partial \theta_i} Z_i^{T}.$$