Published in final edited form as:

Stat Med. 2016 December 30; 35(30): 5579–5602. doi:10.1002/sim.7078.

A Scalable Approach to Measuring the Impact of Nonignorable Nonresponse with an EMA Application

Weihua Gao^{*}, Donald Hedeker[†], Robin Mermelstein[‡], and Hui Xie[§]

^{*}Division of Epidemiology and Biostatistics, School of Public Health, University of Illinois at Chicago

[†]Department of Public Health Sciences, The University of Chicago

[‡]Department of Psychology and School of Public Health, University of Illinois at Chicago

[§]Division of Epidemiology and Biostatistics, School of Public Health, University of Illinois at Chicago and Faculty of Health Sciences, Simon Fraser University

Abstract

There is often a need to assess the dependence of standard analyses on the strong untestable assumption of *ignorable* missingness. To tackle this problem, past research developed simple sensitivity index measures assuming a linear impact of nonignorability and missingness in outcomes only. These restrictions limit their applicability for studies with missingness in both outcome and covariates. Nonignorable missingness in this setting poses significant new analytic challenges and calls for more general and flexible methods that are also computationally tractable even for large datasets. In this paper, we relax the restrictions of extant linear sensitivity index methods and develop nonlinear sensitivity indices that maintain computational simplicity and perform equally well when the impact of nonignorability is locally linear. On the other hand, they can substantially improve the effectiveness of local sensitivity analysis when regression outcomes and covariates are subject to concurrent missingness. In this situation, the local linear sensitivity analysis fails to detect the impact of nonignorability while the proposed nonlinear sensitivity measures can. Because the new sensitivity indices avoid fitting complicated nonignorable models, they are computationally tractable (i.e., scalable) for use in large datasets. We develop general formula for nonlinear sensitivity index measures, and evaluate the new measures in simulated data and a real dataset collected using the Ecological Momentary Assessment (EMA) method.

Keywords

Electronic Data Capture; Missing Data; Nonignorability; Nonlinear Sensitivity Index; Selection Bias

1 Introduction

Evidence-based population studies rely heavily on relevant and proper data. However, despite best efforts to collect representative samples to infer about the underlying populations, the available data are often plagued by various selection issues such that a subset of ideal data is systematically excluded with respect to particular attributes. For instance, frequently we can only obtain or use a subset of a random sample because at least one element of the ideal data is unobserved for some study units either due to uncontrollable behaviors of study units or unanticipated sampling requirements. Such selection issues may render the resulting data unrepresentative of the underlying populations, which threatens the validity of standard analyses designed for random samples by introducing potentially significant selection bias into these regular analyses.

Statistical methods dealing with selection issues in such situations almost always require making some important but unverifiable assumptions about the missing data process. One frequently invoked assumption is *ignorability* [1]. For likelihood-based and Bayesian inference, ignorability requires that data be missing at random (MAR) and that parameters governing the complete data generation process and those governing the missing data mechanism are distinct. When ignorability does not hold, standard analyses ignoring this selective data feature can lead to incorrect statistical inferences. Because the ignorability condition is typically unverifiable directly, existing research has shown that missing data analysis should not depend solely on the conclusion based on this single unverifiable assumption about the missing data process; instead it is an important component of analysis to assess whether the conclusions from the standard analysis hold over a range of plausible missing data mechanisms [2–4]. This view about missing data analysis has already had a profound impact on policy making. As detailed in the recent book by the NRC Panel on the Handling of Missing Data (2010), “Sensitivity 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.”

Recently there has been growing interest in measuring local sensitivity to nonignorable selection, where one examines the sensitivity of inferences to small departures from the ignorable models [5–7]. Troxel et al. [8] proposed an index of local sensitivity to nonignorability (ISNI) to evaluate the potential effect of nonignorability when the outcome follows generalized linear models. The method has been further extended to a causal model for noncompliance [9], longitudinal data [10–15], nonparametric and semiparametric models [16, 17], and coarsened data [18].

Although these approaches to local sensitivity analysis possess important merits and are adequate in a range of situations, they all assume that (1) missingness occurs only in the response outcome while covariates are fully observed and (2) the impact of nonignorability is locally linear. A common problem in many studies is that the outcome and covariates are subject to simultaneous missingness. This issue can arise in various contexts such as: missingness in both the outcome and subject-specific time-varying covariates to be collected in a visit missed by a subject in longitudinal clinical trials or observational studies; subject nonresponses to prompts for answering electronic questionnaires sent to mobile data

collection devices held by study participants (e.g., in EMA data); subject not wearing wireless sensors for measuring physical activities, health status and location data. The problem of missingness in both the outcome and covariates raises new statistical and computational challenges. In particular, the impact of nonignorable missingness is often non-monotonic with concurrent missingness in response and covariates, which cannot be captured by the first-order local linear index developed in the earlier papers. The main contribution of this article is to introduce a novel and computationally tractable nonlinear sensitivity index that (1) relaxes the assumption of local linear impact of nonignorability, and (2) allows for missingness in both responses and covariates. These innovations overcome key limitations of the existing linear sensitivity index measures and can substantially increase the applicability and effectiveness of the local sensitivity methods.

At the same time the new index maintains computational simplicity and the index algorithm is scalable to modern rich data, such as EMA data. These emerging types of data often have more complex missing data patterns and reasons, intensive measurements or a large amount of data. In these settings, the likelihood of nonignorable selection models for missing data analysis can involve a large number of high-dimensional integrals. The brute force approach that directly estimates a range of joint selection models with this complicated likelihood is computationally prohibitive. Even when it is possible, such analyses involve heavy computational workload, limiting the number, types, and sizes of analyses that can be performed. These issues can significantly hinder the use of principled statistical methods for quantifying the effects of nonignorable missingness, and call for new principled methods that are also computationally tractable for such rich data (i.e., scalability). Because our index methods avoid fitting any complicated nonignorable selection models, they provide principled sensitivity analysis methods that address the issue of scalability. Our experiences show that they are capable of reducing hours or days of computational time to just a few seconds.

There exist alternative approaches to nonignorable selection issues to which our approach can be related. One alternative approach is to directly estimate a joint model that includes a component describing selection mechanisms. A challenge for this approach is the model identifiability issue. Because typically observed data provide little information for distinguishing between different selection mechanisms, numerous studies have shown that the direct-estimation approach is often unidentifiable or unreliable [19–22]. Instrumental variables (IV) can help identify the model. However, clean and strong instrumental variables are often hard to find. Alternatively one can use complementary data, such as refreshment samples [23], to identify models. However, useful complementary data are often unavailable or expensive to collect. Our approach can be used as a screening method to tell when we need to search for alternative analysis strategies or collect additional data and perform arduous work of nonignorable modeling. Two other related approaches are global sensitivity analyses [24–26] and bound analyses [15, 27]. These two approaches have the merit of examining a broader range of inferences under alternative selection mechanisms. However, the resulting analyses can be too wide to be useful in practice because they can include unrealistically extreme scenarios of nonignorability [18]. Furthermore, the computation can become prohibitive because of the need to repeatedly fit alternative nonignorable models. In some situations, it can be advantageous to combine different methods. Xie and Qian [15]

combine a local sensitivity analysis with bound analysis to account for an unknown mixture of dropouts and intermittent missingness in longitudinal data.

Our application considers the nonresponse issue in EMA studies. EMA, as a real-time data capturing method, has become increasingly important in health studies [28]. By sending prompts to mobile devices held by study participants and asking them to provide answers to various survey questions in real-time, EMA studies can provide more accurate data. An ideal EMA study collects data for all planned measurements. However, like most studies involving human subjects, missing data are ubiquitous and unavoidable. For example, when studying the question “Are moods just prior to smoking different than moods during random background times”, there can be a moderate amount of nonresponses to the random prompts on study participants’ handheld computers. One can never be certain that these nonresponses are random. It may be suspected that subjects do not respond to the random prompts when their moods are worse, and thus the observed moods from answered random prompts are a biased representation of random background moods. We apply our new nonlinear index method to quantify the potential bias from such missing data in EMA studies.

In terms of the organization of this paper, after a brief review in Section 2, we develop general formula for nonlinear sensitivity index in Section 3. We discuss index calibration in Section 4 and then apply the methods to simulated data in Section 5 and a real EMA dataset in Section 6.

2 A Review of the ISNI methodology

Previous development in the ISNI methodology considers the case of missing outcomes only. Let Y be a random variable with density function $f_{\theta_Y}(y|x)$ on the sample space Ω_Y , where θ_Y is the parameter of interest and x contains data on a set of fully observed covariates. Let G be the completeness indicator for Y , where $G = 1$ if Y is observed and $G = 0$ if otherwise. The variable G follows the probability density function $f_{\gamma_0, \gamma_1}(g|s, y)$, where the parameter γ_0 associates the probability of missingness with a set of fully observed missingness predictors s , and the nonignorability parameter γ_1 associates the probability of missingness with the potentially unobserved outcome Y . Typically the missing data model is specified in a way such that it corresponds to MAR when $\gamma_1 = 0$. Viewing missing data as a special case of coarse data [29, 30], we let Y^* denote the observed coarsened data on Y such that $Y^* = \{Y\}$ if $G = 1$ and $Y^* = \Omega_Y$ if $G = 0$. Let L be the correct log-likelihood for $(\theta_Y, \gamma_0, \gamma_1)$ given the observed data consisting of N i.i.d copies of (Y^*, G) under the above general nonignorable selection model. Then for $i = 1, \dots, N$, we have

$$L(\theta_Y, \gamma_0, \gamma_1) = \sum_{i: g_i=1} \left[\ln f_{\theta_Y}(y_i|x_i) + \ln f_{\gamma_0, \gamma_1}(g_i|y_i, s_i) \right] + \sum_{i: g_i=0} \ln \int_{\Omega_{Y_i}} f_{\theta_Y}(y_i^{mis}|x_i) f_{\gamma_0, \gamma_1}(g_i|y_i^{mis}, s_i) dy_i^{mis}, \quad (1)$$

where $f_{\theta_y}(y_i^{mis}|x_i)$ and $f_{\gamma_0, \gamma_1}(g_i|y_i^{mis}, s_i)$ are $f_{\theta_y}(y_i|x_i)$ and $f_{\gamma_0, \gamma_1}(g_i|y_i, s_i)$, respectively, evaluated at y_i^{mis} .

Given the aforementioned identification issue that is well recognized in the literature, an alternative and more prudent approach is to perceive the above nonignorable model as provisional and to perform a sensitivity analysis over a plausible range of values for γ_1 . For

any value of γ_1 , we must obtain $(\hat{\theta}_y(\gamma_1), \hat{\gamma}_0(\gamma_1))$ by maximizing L over (θ_y, γ_0) , and compare inferences for the range of γ_1 values. However, computational cost can be very high because of the need to repeatedly evaluate the integrals with respect to missing outcomes. A feasible approach that overcomes these difficulties uses the first-order Taylor-

series approximation as $\hat{\theta}_y(\gamma_1) \approx \hat{\theta}_y(0) + \frac{\partial \hat{\theta}_y(\gamma_1)}{\partial \gamma_1}|_{\gamma_1=0} \times \gamma_1$ [5–8].

Both $\hat{\theta}_y(0)$ and $\frac{\partial \hat{\theta}_y(\gamma_1)}{\partial \gamma_1}|_{\gamma_1=0}$ above are easy to obtain: the MAR estimate $\hat{\theta}_y(0)$ is obtained

using standard software; $\frac{\partial \hat{\theta}_y(\gamma_1)}{\partial \gamma_1}|_{\gamma_1=0}$ measures the changing rate of $\hat{\theta}_y(\gamma_1)$ with respect to γ_1 at MAR, and is referred to as the *index of local sensitivity to nonignorability* (ISNI, [8]). The ISNI computation requires neither evaluating the aforementioned complicated integrals nor fitting any joint selection models, which permits simple and fast evaluation of $\hat{\theta}_y(\gamma_1)$. ISNI formulas have been developed for a range of problems with missing outcomes [9–18].

3 Developing Nonlinear Sensitivity Index Method

In this section we develop nonlinear sensitivity indices for selective missingness in both the outcome and covariates. In regression analysis, covariates are variables that are possibly predictive of the regression outcome. Covariates are included in the analysis because their relationship with the outcome may be of direct interest or they may serve as confounding, mediating or interaction variables. The development of ISNI so far assumes that only the outcome Y is subject to missingness while the covariates in X are fully observed. Thus, the current ISNI formulas are not computable with missingness in both the outcome and covariates because they require known covariate values. More importantly, as will be shown below, the impact of nonignorability takes a highly nonlinear form around the MAR model and consequently requires the development of new nonlinear sensitivity indices.

3.1 Selection Models

In order to permit simultaneous missingness in response Y and some covariates in X in evaluating the impact of nonrandom nonresponse, we expand the selection model as described in Section 2 to include a model for missing covariates. We consider here models for cross-sectional data although our index approach is general and can be extended to more complex models. We rewrite the covariates in the response model as (X, W) , where W denotes fully-observed covariates and X denotes the covariates subject to concurrent missingness with Y . Let $f_{\theta_X}(X|W)$ denote the probability model for X given the fully

observed covariates W . If X consists of a single covariate, a generalized linear model can be used for $f_{\theta_X}(X|W)$ where

$$f_{\theta_X}(x_i|w_i) = \exp \left\{ \frac{x_i \Psi_i(\phi, w_i) - b(\Psi_i(\phi, w_i))}{a(\tau)} + c(x_i, \tau) \right\},$$

Ψ_i is the canonical parameter as a function of the parameter ϕ ; functions $b(\cdot)$ and $c(\cdot, \cdot)$ determine a particular distribution in the exponential family; $a(\tau) = \tau/v$, with the dispersion parameter τ and a known weight v . If X consists of multiple covariates $(X_1, X_2, \dots, X_P)$, we can consider using a product conditional model where

$$f_{\theta_X}(X|W) = f_{\theta_{X_1}}(X_1|W) f_{\theta_{X_2}}(X_2|W, X_1) \cdots f_{\theta_{X_P}}(X_P|W, X_1, \dots, X_{P-1}).$$

In the product conditional model, each of $f_{\theta_{X_j}}(X_j|W, \tilde{X}_j)$, $j=1, \dots, P$ can be modeled using the above generalized linear model, where $\tilde{X}_j = (X_1, \dots, X_{j-1})$ when $j > 1$ and $\tilde{X}_j$ is null when $j = 1$. We further let $X^* = (X_1^*, X_2^*, \dots, X_P^*)$ denote the observed data on X , where $X_j^* = \{X_j\}$ if $G = 1$, $X_j^* = \Omega_{X_j}$ if $G = 0$ and Ω_{X_j} is the sample space of X_j , $j = 1, \dots, P$. When all the missing variables are continuous, a popular modeling approach is to model the missing covariates and outcome, sometimes after some transformation, using a multivariate normal model (MVN). We note that the MVN is equivalent to a product of univariate conditional normal models. As a result, the MVN is a special case of our product conditional modeling approach as described above. Our conditional modeling approach is more general and flexible because it can flexibly model non-normal continuous distributions such as Gamma, Inverse Gaussian etc. Chen, Xie and Qian [31] provide a more detailed discussion and comparison of product conditional models and the MVN approach.

To investigate the potential impact of nonignorable nonresponse, we need to posit a missing data model describing the response behavior. As shown extensively in literature, the joint selection model is typically unidentifiable because observed data provide little information for distinguishing different selection mechanisms. Even when the model is weakly identified, the estimation of selection models can be highly sensitive to the strong model assumptions, such as those made in the main outcome model. Thus, a sensible approach is to view the missing data model as provisional and use it as a device to perform sensitivity analysis. As such, it is important for the missingness mechanism model to be plausible and the parameters in the model to be interpretable. With this in mind, we assume the following model for describing the response behavior for subject i :

$$P(G_i=1|s_i, y_i, x_i) = h\left(\gamma_0^T s_i + \gamma_{1x}^T x_i + \gamma_{1y} y_i\right),$$

where $G_i = 1(0)$ if the i th subject responded (did not respond), $h(\cdot)$ is the inverse link function, s_i includes a set of observed predictors for response, $\gamma_1 = (\gamma_{1x}, \gamma_{1y})$ is a vector of nonignorability parameters that associate the probability of missingness with the outcome

and the covariates that were concurrently missing when there was no response. When the nonignorability parameter γ_1 departs from the point of zero, the nonresponse probability depends on the potentially unobserved outcome and covariate values. Consequently the observed outcomes become a selective subset of the outcomes to be collected, which can cause selection bias in the MAR estimates of the outcome model parameters. The general nonresponse model includes the following two special cases.

Special case 1: Y-dependent-only Nonignorability, where $\gamma_{1x} = 0$, and $P(G_i = 1) =$

$$h\left(\gamma_0^T s_i + \gamma_{1y} y_i\right).$$

Special case 2: X-dependent-only Nonignorability, where $\gamma_{1y} = 0$, and $P(G_i = 1) =$

$$h\left(\gamma_0^T s_i + \gamma_{1x}^T x_i\right).$$

3.2 Generalized nonlinear ISNIs

Under the above joint selection model, we have the log-likelihood as

$$L(\theta_y, \theta_x, \gamma_0, \gamma_1) = \sum_{i:g_i=1} \ln f_{\theta_y}(y_i|x_i, w_i) + \sum_{i:g_i=1} \ln f_{\theta_x}(x_i|w_i) + \sum_{i:g_i=1} \ln f_{\gamma}(g_i|s_i, y_i, x_i) + \sum_{i:g_i=0} \ln \left(\int_{\Omega_{Y_i}} \int_{\Omega_{X_{1i}}} \cdots \int_{\Omega_{X_{Pi}}} f_{\gamma}(g_i|s_i, y_i^{mis}, x_i^{mis}) f_{\theta_y}(y_i^{mis}|x_i^{mis}, w_i) f_{\theta_x}(x_i^{mis}|w_i) dy_i^{mis}, d. \right)$$

(2)

where $i = 1, \dots, N$; $f_{\theta_y}(y_i|x_i, w_i)$, $f_{\theta_x}(x_i|w_i)$ and $f_{\gamma}(g_i|s_i, y_i, x_i)$ denote the density function for the ideal outcome, ideal covariates and response behaviors, respectively, and

$f_{\theta_y}(y_i^{mis}|x_i^{mis}, w_i)$, $f_{\theta_x}(x_i^{mis}|w_i)$ and $f_{\gamma}(g_i|s_i, y_i^{mis}, x_i^{mis})$ are $f_{\theta_y}(y_i|x_i, w_i)$, $f_{\theta_x}(x_i|w_i)$

and $f_{\gamma}(g_i|s_i, y_i, x_i)$ respectively, evaluated at (x_i^{mis}, y_i^{mis}) . As compared with the likelihood in Eqn (1) which involves missingness only in the outcome, the nonignorable missing data problem with simultaneous missingness in both the outcome and covariates becomes considerably more difficult to handle because of the multidimensional integrals shown in Eqn (2). It is a computational challenge and time-consuming to directly estimate a range of joint selection models which requires evaluating the above complicated nonignorable likelihood many times. The computational cost can even become prohibitive for large datasets that have a large number of high-dimensional integrals. These difficult issues hinder the use of principled methods in many studies to quantify the impact of nonrandom missingness, and call for tractable methods that are easily scalable to those large data.

Besides the aforementioned computational challenge, the impact of nonignorability takes a highly nonlinear shape around the MAR model and thus the current linear ISNI approach loses its effectiveness. Thus, there is a need to generalize ISNI to make it applicable in a wider range of applications. Let $\hat{\theta}(\gamma_1) = (\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1), \hat{\gamma}_0(\gamma_1))$ be the MLEs given γ_1 ,

and $\hat{\theta}^k(\gamma_1)$ be the k th element of $\hat{\theta}(\gamma_1)$. We propose to employ a second-order Taylor-series expansion of the MLEs of model parameters as a function of γ_1 .

$$\hat{\theta}^k(\gamma_1) \stackrel{\text{Taylor Expansion}}{\approx} \hat{\theta}^k(0) + \gamma_1^T \overbrace{\frac{\partial \hat{\theta}^k(\gamma_1)}{\partial \gamma_1} \Big|_{\gamma_1=0}}^{\text{ISNIL}} + \frac{1}{2} \gamma_1^T \overbrace{\frac{\partial^2 \hat{\theta}^k(\gamma_1)}{\partial \gamma_1 \gamma_1^T} \Big|_{\gamma_1=0}}^{\text{ISNIQ}} \gamma_1. \quad (3)$$

3.3 Step 1: Extending Linear ISNI

In this and the following subsections, we take two steps to develop the nonlinear sensitivity indices. In this subsection, we first extend the linear ISNI to account for missing covariates, and then demonstrate empirically and theoretically that the extended linear ISNI (ISNIL in Eqn (3)) is inadequate when the outcome and covariates are subject to simultaneous missingness. This motivates the need for developing nonlinear sensitivity indices in Step 2 as described in the following subsection. As shown in Appendix A, the ISNIL for $\hat{\theta}_y$, the parameter estimates of primary interest, is

$$\begin{aligned} \frac{\partial \hat{\theta}_y(\gamma_1)}{\partial \gamma_1^T} \Big|_{\gamma_1=0} &= - \left[\overbrace{\nabla^2 L_{\theta_y, \theta_y}^{-1} \nabla^2 L_{\theta_y, \gamma_{1y}}}^{\text{ISNIL}_y(\hat{\theta}_y)}, \overbrace{\nabla^2 L_{\theta_y, \theta_y}^{-1} \nabla^2 L_{\theta_y, \gamma_{1x}}}^{\text{ISNIL}_x(\hat{\theta}_y)} \right], \\ \text{where } \nabla^2 L_{\theta_y, \theta_y} &= \sum_{i:g_i=1} \frac{\partial}{\partial \theta_y} \ln \frac{f_{\theta_y}(y_i^{obs} | x_i^{obs}, w_i)}{\partial \theta_y^T} \Big|_{\gamma_1=0}, \\ \nabla^2 L_{\theta_y, \gamma_{1y}} &= - \sum_{i:g_i=0} h_i \cdot E_{x_i^{mis} | w_i} \left(\frac{\partial E(y_i^{mis} | x_i^{mis}, w_i)}{\partial \theta_y} \right) \Big|_{\gamma_1=0}, \\ \nabla^2 L_{\theta_y, \gamma_{1x}} &= - \sum_{i:g_i=0} h_i \cdot \frac{\partial E(x_i^{mis} | w_i)}{\partial \theta_y} \Big|_{\gamma_1=0} = 0, \end{aligned}$$

and for arguments a, b , $\nabla^2 L_{a,b} = \frac{\partial^2 L(\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)}{\partial a \partial b} \Big|_{\hat{\theta}_y(0), \hat{\theta}_x(0), \hat{\gamma}_0(0), 0}$,

$h_i = \text{Prob}(G_i=1) = \frac{e^{\hat{\gamma}_0^T(0) s_i}}{1 + e^{\hat{\gamma}_0^T(0) s_i}}$ denoting the probability of being observed under the MAR model. The superscript *obs* (*mis*) indicates that the corresponding data element is observed (or missing).

As shown above, the first-order sensitivity consists of two components: ISNIL_y and ISNIL_x , which correspond to the sensitivity attributable to outcome-dependent nonignorability (i.e., with respect to γ_{1y}) and that attributable to covariates-dependent nonignorability (i.e., with respect to γ_{1x}), respectively. Both sensitivity components involve the product of two parts, with the first part being the inverse Fisher information matrix of the MAR model, and the second part evaluating the orthogonality of θ_y and nonignorability parameters (γ_{1y} and γ_{1x} , respectively). Note that in the case of unit nonresponses when the outcome and covariates are missing simultaneously, the MAR analysis coincides with the complete-case analysis. It then follows from [32] that in the case of unit nonresponse there is no selection bias on the

MAR estimates of the outcome models if the selection only depends on missing covariates and selection bias in the MAR estimates of the outcome model only occurs when the nonresponse depends on the missing outcome values. This is manifested by the results that $\text{ISNIL}_x(\hat{\theta}_y) = 0$ as the expected values of missing covariates given the observed covariate values are independent of θ_y under MAR.

3.4 An Example: Simple Linear Regression

To illustrate the extended linear ISNI and its limitations, we consider an example where the ideal outcome model is $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$, $i = 1, \dots, n$, the covariate $x_i \sim N(\mu_x, \sigma_x^2)$ and the response model $\text{logit}(P(G_i = 1)) = \gamma_0 + \gamma_{1x} x_i + \gamma_{1y} Y_i$. Using the general formula for ISNIL above, we can show the first-order linear ISNI for the parameter estimates of primary interest as follows

$$[\text{ISNIL}_x(\hat{\beta}_0), \text{ISNIL}_y(\hat{\beta}_0)] = \left[0, -\hat{\sigma}^2(0) \frac{n_m}{n}\right], \quad [\text{ISNIL}_x(\hat{\beta}_1), \text{ISNIL}_y(\hat{\beta}_1)] = [0, 0],$$

where n_m is the number of missing observations and $\hat{\sigma}^2(0)$ is the MAR estimate of σ^2 . The analysis shows that the MAR estimate $\hat{\beta}_1$ for the parameter associated with the missing covariate needs no first-order adjustment because both $\text{ISNIL}_y(\hat{\beta}_1)$ and $\text{ISNIL}_x(\hat{\beta}_1)$ are zeros. This seems to contradict with the intuition that the parameter estimates should be affected by the γ_1 value when the probability of missingness depends on the outcome.

To understand why the first-order sensitivity index is inadequate here, we consider an analysis of simulated data under the special case of outcome-dependent nonignorability, i.e., by setting $\gamma_{1x} = 0$. We simulated the ideal outcome Y from the above simple linear regression model with $\beta_0 = 0$, $\beta_1 = 1$, $\sigma^2 = 1$ and $i = 1, \dots, 1000$, and the covariate

$x_i \sim N(\mu_x, \sigma_x^2)$, where $\mu_x = 0$, $\sigma_x = 1$. The response behavior follows a logistic regression

model: $\text{logit}(\mu_i^G) = \gamma_0 + \gamma_{1y} y_i$, where $\gamma_0 = 1$, $\gamma_1 = 1$. Figure 1 (A) plots one fifth random subsample from the simulated data that has the proportion of missingness being 33.8%.

Figure 2 shows the analysis using the simulated dataset. The solid curve shows the exact sensitivity curve of $\hat{\beta}_1(\gamma_1)$, which was computed by maximizing the log-likelihood in Eqn (2) for the joint nonignorable selection models for a range of values for γ_1 . As shown in Fig 2, the parameter change around the MAR model takes a quadratic U-shaped nonlinear form. At the MAR model its first-order derivative is zero, just as shown from the computation of ISNIL above.

To further understand why the parameter estimate can take this non-monotonic U-shaped form when X and Y are subject to concurrent missingness, it is helpful to consider the following heuristic example. Fig 3 presents a comparative analysis of two simulated data sets. We first generated one complete data set from the simple linear regression model described above. We then generated missing data using the above logistic response behavior model where $\gamma_1 = 1$ for the left plot and $\gamma_1 = -1$ for the right plot in Fig 3. Each plot in

Figure 3 plots one fifth random subsample from the simulated data. The black dots and red pluses in the plots denote the observed data and missing data, respectively. The thick solid and dashed lines represent the fitted regression line using the complete data and observed data, respectively. When there is quadratic U-shaped impact of missing data on the MAR slope estimate as found above, we would expect the slopes of thick dashed line for $\gamma_1 = 1$ and $\gamma_1 = -1$ have the same directional change relative to the slope of the thick solid line. This is indeed the case: as shown in Fig 3 the slopes of the dashed lines fitted using the observed data for $\gamma_1 = 1$ and $\gamma_1 = -1$ are both smaller than the slope of the thick solid line. It is clear from Fig 3 that the missing data in the two plots of Figure 3 not only have very different outcome values but also very different covariate values, which can cause the slope estimate to have bias in the same direction, and consequently an U-shaped form of parameter changes around the MAR model. In contrast, for a dataset in which only Y is subject to missingness and X is always observed, the predicted position for missing data is forced to have the same X value with all posited γ_1 values. Therefore opposite nonignorability (i.e. when the posited γ_1 values have opposite signs) will typically give opposite directional change in estimates and yield approximately linear impact around MAR. Our aim here is to use this heuristic example to aid understanding the source of the observed nonlinear U-shaped impact of missing data and to motivate our development of nonlinear ISNI methods. In the following section, we will formally develop the nonlinear sensitivity indices when both X and Y are subject to missingness as well as not only for the outcome-dependent nonignorability but also for the more general outcome- and covariate-dependent nonignorability.

3.5 Step 2: Developing Measures of Higher-order Sensitivity

For notational simplicity, we rewrite $\gamma_1 = (\gamma_{1x}, \gamma_{1y}) = (\gamma_{11}, \gamma_{12}, \dots, \gamma_{1P}, \gamma_{1Y})$. We let γ_{1p} and γ_{1q} denote two generic elements in the γ_1 vector. As shown in Appendix A we have

$$\begin{aligned} \text{ISNIQ}_{pq}(\hat{\theta}_y) &= \frac{\partial^2 \hat{\theta}_y(\gamma_1)}{\partial \gamma_{1p} \gamma_{1q}} \Big|_{\gamma_1=0} = -\nabla^2 L_{\theta_y, \theta_y}^{-1} \nabla^3 L_{\theta_y, \theta_y, \gamma_{1q}} \text{ISNIL}_p(\hat{\theta}_y) - \nabla^2 L_{\theta_y, \theta_y}^{-1} \nabla^3 L_{\theta_y, \theta_y, \gamma_{1p}} \text{ISNIL}_q(\hat{\theta}_y) \\ &\quad - \nabla^2 L_{\theta_y, \theta_y}^{-1} \left(\sum_{j=1}^{n_{\theta_y}} \nabla^3 L_{\theta_y, \theta_y, \theta_y^j} \text{ISNIL}_q(\hat{\theta}_y^j) \right) \text{ISNIL}_p(\hat{\theta}_y) \\ &\quad - \nabla^2 L_{\theta_y, \theta_y}^{-1} \nabla^3 L_{\theta_y, \theta_x, \gamma_{1q}} \text{ISNIL}_p(\hat{\theta}_x) - \nabla^2 L_{\theta_y, \theta_y}^{-1} \nabla^3 L_{\theta_y, \theta_x, \gamma_{1p}} \text{ISNIL}_q(\hat{\theta}_x) \\ &\quad - \nabla^2 L_{\theta_y, \theta_y}^{-1} \nabla^3 L_{\theta_y, \gamma_0, \gamma_{1q}} \text{ISNIL}_p(\hat{\gamma}_0) - \nabla^2 L_{\theta_y, \theta_y}^{-1} \nabla^3 L_{\theta_y, \gamma_0, \gamma_{1p}} \text{ISNIL}_q(\hat{\gamma}_0) \\ &\quad - \nabla^2 L_{\theta_y, \theta_y}^{-1} \nabla^3 L_{\theta_y, \gamma_{1p}, \gamma_{1q}}, \end{aligned} \quad (4)$$

where for arguments a, b, c , $\nabla^3 L_{a,b,c} = \frac{\partial^3 L(\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)}{\partial a \partial b \partial c} \Big|_{\hat{\theta}_y(0), \hat{\theta}_x(0), \hat{\gamma}_0(0), 0}$

$\text{ISNIL}_p(u) = \frac{\partial u}{\partial \gamma_{1p}}$ and $\text{ISNIL}_q(u) = \frac{\partial u}{\partial \gamma_{1q}}$ for the argument u , and $\hat{\theta}_y^j$ denotes the j th element of $\hat{\theta}$. As shown, $\text{ISNIQ}(\hat{\theta}_y)$ is a sum of terms, where each term is generally a product of three components with the first component evaluating the inverse Fisher information matrix

for the ideal outcome model, the second component capturing the orthogonality among θ_y , γ_1 and the other model parameters, and the last component assessing the first-order sensitivity of the other model parameters with respect to γ_1 . All three components in each term are evaluated under the MAR model that simplifies computation considerably.

According to the above general result, we have

$$\begin{aligned} \text{ISNIQ}_{yy}(\hat{\theta}_y) &= \frac{\partial^2 \hat{\theta}_y(\gamma_1)}{\partial \gamma_{1y}^2} \Big|_{\gamma_1=0} \\ &= -2\nabla^2 L_{\theta_y, \theta_y}^{-1} \nabla^3 L_{\theta_y, \theta_y, \gamma_{1y}} \text{ISNIL}_y(\hat{\theta}_y) - \nabla^2 L_{\theta_y, \theta_y}^{-1} \left(\sum_{j=1}^{n_{\theta_y}} \nabla^3 L_{\theta_y, \theta_y, \theta_y^j} \text{ISNIL}_y(\hat{\theta}_y^j) \right) \text{ISNIL}_y(\hat{\theta}_y) \\ &\quad - 2\nabla^2 L_{\theta_y, \theta_y}^{-1} \nabla^3 L_{\theta_y, \theta_x, \gamma_{1y}} \text{ISNIL}_y(\hat{\theta}_x) - 2\nabla^2 L_{\theta_y, \theta_y}^{-1} \nabla^3 L_{\theta_y, \theta_x, \gamma_{1y}} \text{ISNIL}_y(\hat{\gamma}_0) - \nabla^2 L_{\theta_y, \theta_y}^{-1} \nabla^3 L_{\theta_y, \gamma_{1y}, \gamma_{1y}}, \end{aligned}$$

$$\begin{aligned} \text{ISNIQ}_{yq}(\hat{\theta}_y) &= \frac{\partial^2 \hat{\theta}_y(\gamma_1)}{\partial \gamma_{1y} \partial \gamma_{1q}} \Big|_{\gamma_1=0} = -\nabla^2 L_{\theta_y, \theta_y}^{-1} \nabla^3 L_{\theta_y, \theta_x, \gamma_{1y}} \text{ISNIL}_q(\hat{\theta}_x) - \nabla^2 L_{\theta_y, \theta_y}^{-1} \nabla^3 L_{\theta_y, \gamma_0, \gamma_{1y}} \text{ISNIL}_q(\hat{\gamma}_0) \\ &\quad - \nabla^2 L_{\theta_y, \theta_y}^{-1} \nabla^3 L_{\theta_y, \gamma_{1y}, \gamma_{1q}}, \quad \text{for } \gamma_{1q} \in (\gamma_{11}, \dots, \gamma_{1P}) \end{aligned}$$

$$\text{ISNIQ}_{pq}(\hat{\theta}_y) = \frac{\partial^2 \hat{\theta}_y(\gamma_1)}{\partial \gamma_{1p} \partial \gamma_{1q}} \Big|_{\gamma_1=0} = 0, \quad \text{for } \gamma_{1p} \in \gamma_{1x} = (\gamma_{11}, \dots, \gamma_{1P}) \quad \text{and} \quad \gamma_{1q} \in \gamma_{1x} = (\gamma_{11}, \dots, \gamma_{1P}).$$

As shown above, $\text{ISNIQ}_{pq}(\hat{\theta}_y) = 0$ when $\gamma_{1p}, \gamma_{1q} \in \gamma_{1x}$ for the same reason as that for the result of $\text{ISNIL}_x = 0$.

3.6 An Example: Simple Linear Regression: Continue

For the joint selection models considered Section 3.4, using the above general formula for ISNIQ and the terms derived in Appendix B, we have for $\gamma_1 = (\gamma_{1x}, \gamma_{1y}) = (\gamma_{11}, \gamma_{1y})$

$$\text{ISNIQ}(\hat{\beta}_0) = \frac{\partial^2 \hat{\beta}_0(\gamma_1)}{\partial \gamma_1 \partial \gamma_1^T} \Big|_{\gamma_1=0} = \begin{bmatrix} 0 & -\hat{\sigma}^2(0) \frac{n_o n_m}{n^2} \frac{n_o \hat{\sigma}_x^2(0)}{s_{xoxo}} \bar{x}_o \\ -\hat{\sigma}^2(0) \frac{n_o n_m}{n^2} \frac{n_o \hat{\sigma}_x^2(0)}{s_{xoxo}} \bar{x}_o & -2\hat{\beta}_1(0) \hat{\sigma}^2(0) \frac{n_o n_m}{n^2} \frac{n_o \hat{\sigma}_x^2(0)}{s_{xoxo}} \bar{x}_o \end{bmatrix}$$

$$\text{ISNIQ}(\hat{\beta}_1) = \frac{\partial^2 \hat{\beta}_1(\gamma_1)}{\partial \gamma_1 \partial \gamma_1^T} \Big|_{\gamma_1=0} = \begin{bmatrix} 0 & \hat{\sigma}^2(0) \frac{n_o n_m}{n^2} \frac{n_o \hat{\sigma}_x^2(0)}{s_{xoxo}} \\ \hat{\sigma}^2(0) \frac{n_o n_m}{n^2} \frac{n_o \hat{\sigma}_x^2(0)}{s_{xoxo}} & 2\hat{\beta}_1(0) \hat{\sigma}^2(0) \frac{n_o n_m}{n^2} \frac{n_o \hat{\sigma}_x^2(0)}{s_{xoxo}} \end{bmatrix}$$

where $S_{xoxo} = \sum_{i: g_i=1} (x_i - \bar{x}_o)^2$, and $\hat{\beta}_1(0)$, $\hat{\sigma}^2(0)$ and $\hat{\sigma}_x^2(0)$ are the MAR estimates of β_1 , σ^2 and σ_x^2 , and $n_o = n - n_m$. All these entries in the ISNIQ matrices are evaluated using the MAR estimates only. As shown above and in Section 3.4, when Y and X are subject to concurrent missingness, the MAR slope parameter $\hat{\beta}_1(0)$ has no first-order sensitivity (i.e.,

both $\text{ISNI}_{L_x}(\hat{\beta}_1(0))$ and $\text{ISNI}_{L_y}(\hat{\beta}_1(0))$ are zeros) whereas the ISNIQ matrix for $\hat{\beta}_1(0)$ has non-zero entries, indicating the importance of examining the higher order sensitivity.

4 Index Calibration

4.1 Y-dependent Nonignorability

The ISNI value depends on the scale of the outcome Y and the covariates in X when they are continuous because in this case the magnitude of $\gamma_1 = \mathbf{1}$ depends on the scale of Y and X . As a result, scale-free index calibration can facilitate the use and interpretation of the index. We first consider approaches to calibrating the sensitivity index in the simpler case of Y -dependent nonignorability where $\gamma_1 = (\gamma_{1_y})$, which effectively sets $\gamma_{1_x} = \mathbf{0}$. One approach is to evaluate changes in parameter estimates for a magnitude of nonignorability where one SD change in Y is associated with an odds ratio of $e^1 = 2.7$ in the probability of being

observed, i.e. when $\gamma_1 = \pm \frac{1}{\sigma_Y}$, which can be considered as one standardized magnitude of nonignorability.

Alternatively, we can compute the minimum magnitude of nonignorability that is needed for the selection bias to be equal to one standard error and evaluate whether such nonignorability is feasible. Specifically, note that for θ^j (the j th element of θ) we have

$\hat{\theta}^j(\gamma_1) - \hat{\theta}^j(0) \approx \gamma_1 \text{ISNI}_{L_y}(\hat{\theta}^j) + \frac{\gamma_1^2}{2} \text{ISNIQ}_{yy}(\hat{\theta}^j)$. From the above equation, we compute the smallest value of γ_1 , denoted as $\tilde{\gamma}_1$, that causes the right hand side of the above equation to be same as one standard error. To put the $\tilde{\gamma}_1$ value in the scale of standardized magnitude of nonignorability defined above, we define a sensitivity transformation c statistic as $c = |\tilde{\gamma}_1 * \sigma_Y|$. 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 observed. Note that when the c statistic is large, it means that only extreme nonignorability can make the selection bias to be as large as the sampling error, and consequently nonignorability is of less concern. When the c statistic is small, modest nonignorability can cause selection bias to be as large as the sampling error, and consequently nonignorability is of concern. Following [8], we suggest using $c = 1$ as a general cutoff value for important

sensitivity. Note that when the $\text{ISNIQ} = 0$, we have $c = \left| \frac{\text{S. E.} * \sigma_Y}{\text{ISNI}} \right|$ which reduces to the c statistics defined in [8] for the case of linear ISNI only.

For the case where Y has only a single natural scale, such as a binary variable, re-scaling is not required. One can directly evaluate changes in parameter estimates for a magnitude of nonignorability such that one unit change in Y is associated with an odds ratio of $e^1 = 2.7$ in the probability of being observed, i.e., when $\gamma_1 = \pm 1$. Alternatively, one can compute the sensitivity transformation c statistic as $c = |\tilde{\gamma}_1|$.

4.2 Y-and-X-dependent Nonignorability

We next consider the situation of Y -and- X -dependent nonignorability where $\gamma_1 = (\gamma_{11}, \dots, \gamma_{1P}, \gamma_{1y})$. For continuous outcome and covariates, one can hypothetically standardize these continuous variables to all have SD of one, and consider the missing data model

$Prob(G_i=1) = h\left(s_i^T \gamma_0 + (y_i^s, x_i^s)^T \gamma_1^*\right)$, where $(x_i^s, y_i^s) = \left(\frac{x_{1i}}{\sigma_{x_1}}, \dots, \frac{x_{Pi}}{\sigma_{x_P}}, \frac{y_i}{\sigma_y}\right)$. This rescales continuous variables in the unit of their SDs. If any of these variables has only a single natural scale, such as a binary variable, rescaling this variable by standardization is not required and the natural scale for this variable is used in calibration. Extending the idea for the above univariate nonignorability case to the multidimensional situation, we can examine the changes in the parameter estimates when we perturb γ_1^* from the point of zero to points on the hyperball of radius one, which is equivalent to a magnitude of nonignorability such that $\|\gamma_1^*\| = 1$, where $\|\cdot\|$ is the Euclidean distance of the perturbation from the point zero. It is readily seen that

$$\|\gamma_1^*\| = \sqrt{\gamma_{11}^2 \sigma_{x_1}^2 + \dots + \gamma_{1P}^2 \sigma_{x_P}^2 + \gamma_{1y}^2 \sigma_y^2}, \quad (5)$$

where in the right-hand side of the above equation the outcome and covariates are all on the original scales. We could consider a particular perturbation point in the unit hyperball with additional inputs given from subject experts. In the absence of such external inputs, we recommend considering the range (minimum and maximum) of changes in estimates achievable in the hyperball as

$$\text{range} \left(\overbrace{\gamma_1^T \frac{\partial \hat{\theta}^j(\gamma_1)}{\partial \gamma_1} \Big|_{\gamma_1=0}}^{\text{ISNIL}} + \frac{1}{2} \gamma_1^T \overbrace{\frac{\partial^2 \hat{\theta}^j(\gamma_1)}{\partial \gamma_1 \gamma_1^T} \Big|_{\gamma_1=0}}^{\text{ISNIQ}} \gamma_1 \right), \quad \text{subject to } \sqrt{\gamma_{11}^2 \sigma_{x_1}^2 + \dots + \gamma_{1P}^2 \sigma_{x_P}^2 + \gamma_{1y}^2 \sigma_y^2} = 1.$$

Note that by setting $\gamma_{1x} = (\gamma_{11}, \dots, \gamma_{1P}) = 0$ this evaluation reduces to the special case of the univariate Y -dependent nonignorability. Unlike the univariate case, the multidimensional case considers a broader range of configuration of nonignorability and reasons for nonignorable nonresponse with the cumulative magnitude of nonignorability roughly comparable to that in the univariate case. Except in the simpler univariate nonignorability case and some other simple cases where close-form solutions for obtaining the range are available, in general obtaining the range requires a numerical search procedure (e.g., using optimization procedures implementing the Lagrange multipliers method). Finally, we define the c value as the minimum value of $\|\gamma_1^*\| = 1$ for which the maximal absolute change in the parameter estimate $\hat{\theta}^j$ is the same as its standard error.

5 Illustration on Simulated Data

To illustrate the use of the nonlinear ISNI and demonstrate its superiority, we apply it to simulated data from the simple linear regression example described above. Specifically, we generated random ideal datasets from a simple linear regression model: $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ where $\beta_0 = 0$, $\beta_1 = 1$, $\sigma^2 = 1$ and $i = 1, \dots, 1000$, and the covariate $x_i \sim N(\mu_x, \sigma_x^2)$, where $\mu_x = 0$, $\sigma_x = 1$. The response behavior follows a logistic regression model:

$\text{logit}(\mu_i^G) = \gamma_0 + \gamma_{1y}y_i + \gamma_{1x}x_i$, where the binary indicator $G = 1$ (responded) and $G = 0$ (nonresponded with x and y both missing). The values of γ_0 and $\gamma_1 = (\gamma_{1x}, \gamma_{1y})$ are varied so that the resulting datasets have varying amounts of missingness and varying degrees and types of nonignorable missingness. As shown before, for the case of X -dependent-only nonignorable nonresponse where the response probability depends on X but independent of Y (i.e., $\gamma_{1y} = 0$), there is no biasing effect of nonresponse on the ideal outcome model parameters, and consequently there is no need to do the ISNI adjustment. Therefore we summarize the simulation results for the other two more complex and relevant cases of nonignorable nonresponse.

We first consider Y -dependent-only nonignorable missingness and generated twelve random samples with $\gamma_{1x} = 0$, $\gamma_0 = 1, 2$ or 5 , $\gamma_{1y} = \pm 1$ or ± 0.5 . Figure 2 show the analysis based on the simulated dataset with $\gamma_0 = 1$ and $\gamma_{1y} = 1$ with the proportion of missingness being 33.8%. The solid curve shows the exact sensitivity curve of $\hat{\beta}_1(\gamma_1)$, which was computed by maximizing the the log-likelihood of the joint nonignorable selection models in Eqn (2) for a range of values for γ_{1y} with γ_{1x} fixed at zero. This is computationally intensive and is not required in ISNI analyses. We compute two ISNI-based approximate sensitivity curves as

$$\hat{\beta}_1^{\text{ISNIL}}(\gamma_1) = \hat{\beta}_1(0) + \text{ISNIL}_{yy} * \gamma_{1y}, \quad \hat{\beta}_1^{\text{ISNIQ}}(\gamma_1) = \hat{\beta}_1(0) + \text{ISNIL}_{yy} * \gamma_{1y} + \frac{\text{ISNIQ}_{yy}}{2} \gamma_{1y}^2,$$

where $\text{ISNIL}_{yy} = 0$ and $\text{ISNIQ}_{yy} = 0.290$ for the dataset using the formula derived in Section 3.4 and 3.6, and the MAR estimate $\hat{\beta}_1(0)$ is 0.81 as shown in Table 1. Assuming Y -dependent nonignorability, the computation of ISNI-adjusted estimates requires only

knowing ISNIL_{yy} and ISNIQ_{yy} . The columns " $\frac{\Delta \hat{\beta}_1}{\Delta \gamma_{1y}}$ " and " $\frac{\Delta^2 \hat{\beta}_1}{\Delta \gamma_{1y}^2}$ " in Table 1 present the numerically evaluated first- and second- derivatives at the MAR model using the values obtained from the exact sensitivity curve. In the simulated dataset, the numerically evaluated derivatives are $5e-16$ (i.e. 5×10^{-16}) and 0.292, respectively, and are very close to the ISNIL and ISNIQ values, demonstrating that ISNIL and ISNIQ indeed calculate the first and second derivatives of the sensitivity curves at the ignorable model, but without the need to fit any nonignorable models.

Figure 2 demonstrates the superiority of the higher-order ISNI approximation (broken line) over the first-order ISNI approximation (dotted line). In Fig 2, the first-order ISNIL approximation, as the tangent line of the exact sensitivity curve at MAR, is flat, even though

the estimates on the exact sensitivity curve (the solid line) vary substantially around the MAR model. Clearly in this case it is necessary to compute the higher order ISNI to adequately capture the sensitivity of estimates to nonignorability. The conclusions from the plots (not shown here) of the other simulated data sets are the same in that they all show that the NISNI method significantly improves upon the first-order approximation and provides reasonably accurate and quick assessment of the impact of moderate nonignorability on the MAR analysis.

We next consider the outcome-and-covariate-dependent nonignorable nonresponse and generated twenty-four random datasets with $\gamma_0 = 1, 2$ or 5 , and (a)

$$\gamma_{1x} = \sqrt{0.5} \quad \text{or} \quad -\sqrt{0.5}, \gamma_{1y} = \sqrt{0.5} \quad \text{or} \quad -\sqrt{0.5} \quad \text{or (b)}$$

$$\gamma_{1x} = 0.5/\sqrt{2} \quad \text{or} \quad -0.5/\sqrt{2}, \gamma_{1y} = 0.5/\sqrt{2} \quad \text{or} \quad -0.5/\sqrt{2}. \quad \text{In (a) and (b), we have } \|\gamma_1\| = 1 \text{ and } 0.5, \text{ respectively, the same as that in the above } Y\text{-dependent-only case. Figure 1 (B)}$$

plots the simulated dataset with $\gamma_0 = 1, \gamma_{1y} = \gamma_{1x} = \sqrt{0.5}$ and proportion of missingness being 31.4%. In the scatterplot the observations with small values of Y and X tend to be unobserved, causing downward bias in the MAR slope estimate which is 0.82 as shown in Table 2. We then use the following equations to compute the ISNI-adjusted estimates.

$$\hat{\beta}_1^{\text{ISNIL}}(\gamma_1) = \hat{\beta}_1(0) + \text{ISNIL}^T * \gamma_1, \quad \hat{\beta}_1^{\text{ISNIQ}}(\gamma_1) = \hat{\beta}_1(0) + \text{ISNIL}^T * \gamma_1 + \gamma_1^T \frac{\text{ISNIQ}}{2} \gamma_1,$$

where as shown in Table 2, $\text{ISNIL} = \frac{\partial \hat{\beta}_1(\gamma_1)}{\partial \gamma_1} = (0, 0)$, and

$$\text{ISNIQ} = \frac{\partial^2 \hat{\beta}_1(\gamma_1)}{\partial \gamma_1 \partial \gamma_1^T} = \begin{bmatrix} 0 & 0.194 \\ 0.194 & 0.320 \end{bmatrix} \text{ and } \gamma_1 = (\gamma_{1x}, \gamma_{1y}). \text{ Figure 4 plots the two ISNI-}$$

based approximate sensitivity surfaces: the circles represent the values of $\hat{\beta}_1^{\text{ISNIQ}}$ for the grid of γ_1 values, and the values of $\hat{\beta}_1^{\text{ISNIL}}$ fall on the flat plane formed by the dotted line. The surface formed by the solid lines in Figure 4 plots the exact sensitivity surface, $\hat{\beta}_1(\gamma_1)$, which was computed by maximizing the the log-likelihood in Eqn (2) for the joint nonignorable selection models at a grid of values for $(\gamma_{1x}, \gamma_{1y})$. The computation of the exact sensitivity curve involves fitting complicated nonignorable models repeatedly that takes hours to compute. In contrast the computation of nonlinear ISNIs takes only a few seconds. It is clear that the surface formed by $\hat{\beta}_1^{\text{ISNIQ}}$ provides much better approximation to

the exact sensitivity surface of $\hat{\beta}_1(\gamma_1)$ than that using $\hat{\beta}_1^{\text{ISNIL}}$. The columns $\frac{\Delta \hat{\beta}_1}{\Delta \gamma_1}$ and $\frac{\Delta^2 \hat{\beta}_1}{\Delta \gamma_1^2}$ in Table 2 present the numerically evaluated first- and second- derivatives at the MAR model using the values obtained from the exact sensitivity surface. In the simulated dataset, the

$$\text{numerically evaluated derivatives are } (-5e-06, 7e-05) \text{ and } \begin{bmatrix} -2e-03 & 0.192 \\ 0.192 & 0.318 \end{bmatrix},$$

respectively, and are very close to the ISNIL and ISNIQ values, demonstrating again that ISNIL and ISNIQ indeed calculate the first and second derivatives of the sensitivity surface at the ignorable model, but without the need to fit any nonignorable models. Although both

linear components, $ISNIL_X$ and $ISNIL_Y$ are zeros, the reasons are different. In Figure 4, when $\gamma_{1Y} = 0$, the estimate remains unchanged implying X -dependent-only nonignorability does not impact the estimates in the outcome model. On the other hand, $ISNIL_Y = 0$ because the estimate varies around the MAR model in a highly nonlinear U-shaped form around the MAR model. Because $ISNIL$ are zeros, the first-order ISNI sensitivity surface becomes a flat plane at the MAR estimate, and cannot capture the sensitivity of estimates around MAR model. Similar to the result for the Y -dependent-only nonignorability, the generalized NISNI method significantly improves upon the linear ISNI method and provides adequate and fast evaluation of the local sensitivity of MAR estimates to nonignorability.

Table 1 and Table 2 summarize the simulation results for all simulated datasets. The variation in the response model parameters gives us a range of missing proportions. As shown in these tables, the larger the missing data proportion, the smaller the c statistic indicating larger sensitivity to nonignorable nonresponse when the amount of missing data increases. Furthermore, the c statistic tends to be smaller for Y -and- X -dependent nonignorability than for Y -dependent-only nonignorability. This is expected because Y -and- X -dependent nonignorability includes Y -dependently-only nonignorability as a special case and has the potential to identify larger sensitivity for a broader configurations of nonignorable nonresponse. Finally, within either type of nonignorable nonresponse, the value of NISNI or c statistic is not related to the γ_1 value because the sensitivity analysis should not be expected to inform the magnitude of nonignorability.

6 Application to an EMA study

The NIH has designated the topic of real-time data capture as an important and innovative research area [28], and EMA has become a vital data collection method in cancer research (e.g., [33, 34]). Data come from a longitudinal study of the natural history of smoking among adolescents (PO1CA098262, PI R. Mermelstein). The Electronic Diary component of the study involved sampling novice smokers who were in 9th and 10th grades; 461 adolescents (55.1% female) were recruited as part of the larger study (total N of 1263) and completed the baseline assessment for this study. Data collection occurred via hand-held palmtop computers. Each data collection wave included 7 consecutive days of monitoring. Each random prompt was date- and time-stamped and recorded whether the interview was completed, missed, delayed, or disbanded. The random interviews asked about mood, activity, location, companionship, presence of other smokers, and other behaviors. In addition to the random prompts, participants were trained to event record smoking episodes. The “smoke” interviews included the same questions as the random prompts, and in addition, asked about specific smoking-related items (e.g., how much smoked, how the cigarette was obtained, etc.). Thus, we will be able to compare the subjective and objective contexts surrounding the smoke and random times.

Although the smoking-mood link has been well studied among adults, in particular those trying to stop smoking, much less is known about how mood is associated with smoking among adolescents. Understanding more about the mood-smoking relationship in adolescents can help identify who is most at risk for smoking escalation and developing nicotine dependency, as well as providing insights into intervention development.

As compared with more traditional data collection methods, EMA provides an excellent window into the lives of subjects. In smoking-related studies, EMA captures subtle variations in mood as they occur, and can do so more accurately than other measurement modalities [34, 35]. With the use of random assessments that are independent from the occurrence of specific smoking events, EMA can provide useful comparison information about background mood, and allow us to address several critical questions about mood-smoking relationships among adolescents, including: Does smoking help to regulate mood and how? Are moods just prior to smoking different than moods during random background times? In addition, identifying potential mediating variables may also help in the prediction of smoking escalation among relatively novice smokers.

In this section we apply the NISNI method developed here to quantify the impact of nonignorable prompt nonresponses in EMA data analysis, and to more reliably address these questions concerning the mood-smoking relationship by accounting for the impact of potential nonignorable prompt nonresponse. In the application, we focus on the cross-sectional data from the first collection occasion of the EMA data. About 3% of the observations in the subsample are smoking events with the rest of the subsample coming from random prompts. About 20% of these random prompts were not answered. One study aim is to investigate the mood-smoking relationship in young adults, and to consider the following linear regression model for the ideal outcome:

$$Mood_i = (Intercept, SmkEvent, Social)_i^T \beta + \epsilon_i,$$

where the outcome *Mood* is the negative mood prior to the prompt signal or smoking event with higher values indicating worse mood, *SmkEvent* denotes a variable indicating whether the observation is from a random prompt (=0) or a smoking event (=1); *Social* denotes a measurement of social isolation prior to the prompt signal or smoking event. Both *Mood* and *Social* are the averages of several individual items and as a result take nearly continuous values. Specifically, *Mood* is the average of a subject's evaluation of the following five items: I felt sad, I felt stressed, I felt angry, I felt frustrated, and I felt irritable, where each of the five items is on a rating scale from 1 to 10. *Social* is the average of a subject's evaluation of the following three items: I felt ignored, I felt left out, and I felt lonely, where each item is on a rating scale from 1 to 10. Subjects are queried about their social isolation and negative mood status just prior to receiving the random prompt signal or prior to smoking. The purpose of the analysis is to use the data from both random prompts and smoking reports to examine the following question when studying mood-smoking relationship: Are moods just prior to smoking different than moods during random background times? For this purpose, the regression coefficient for *SmkEvent* compares the mood outcome prior to smoking with that at random background times. There are various reasons to incorporate additional variables such as *Social* as covariates in the above regression model: the *Social* variable can mediate the mood-smoking relationship or serve as a confounding contextual variable that simultaneously affects mood and smoking behavior but not in the mood-smoking causal pathway.

Missing data arise because of nonresponses to the random prompts, which leads to the mood outcome being missing, as well as the contextual factor *Social* associated with the missed random prompts. However, if mood affects adolescents' compliance with answering data collection prompts, then our conclusions about the mood-smoking relationship may be biased. For example, if high positive moods are associated with missing, then the overall random background mood appears less positive (worse), and it will be harder to find differences in mood level between smoking and random times leading one to conclude that smoking is not associated with negative moods. If, on the other hand, negative mood states are associated with missingness, then background level of mood looks better (more positive), and we may conclude that there is a stronger relationship between negative mood and smoking than really exists. Thus, evaluating and controlling for the potential bias from such missing data is extremely important in EMA studies. To investigate the potential impact of nonignorable nonresponse, we assume the following prompt nonresponse model: $\text{logit}P(G_i = 1) = s_i\gamma_0 + \gamma_{1_x} * \text{social}_i + \gamma_{1_y} \text{mood}_i$, where $s_i = (\text{intercept}, \text{SmkEvent})_i$. We assume a normal linear regression model for the ideal data on the covariate *Social*, where $\text{Social}_i = (\text{Intercept}, \text{SmkEvent})_i^T \delta + e_i$.

We first conducted the MAR analysis and then conduct both linear and nonlinear ISNI analysis assuming *mood*-dependently-only nonignorability by fixing $\gamma_{1_x} = 0$ in the above prompt nonresponse model. The MAR analysis results for the *mood* outcome regression model are summarized in Table 3 and show that smoking is indeed associated with higher negative mood just prior to smoking as compared with that at random background times, and this smoking-mood relationship is statistically significant after controlling for social isolation with a p-value of 0.02. The MAR analysis also shows the statistically significant association between the outcome *mood* and the covariate *Social*. These MAR estimates are potentially biased with nonignorable nonresponse to random prompts. Thus a commonly asked question is how reliable or trustworthy the standard MAR analysis is. The purpose of our index development (including index calibration) as a sensitivity analysis tool is to provide researchers with new scientific knowledge to directly investigate this question. One can use ISNI analysis to conveniently gauge the sensitivity of these MAR estimates to nonignorable nonresponse, which avoids fitting any complicated nonignorable models and only requires readily-available MAR estimates. We note two important limitations of the extant ISNI methods that prevent their use in this type of EMA data analysis to quantify the impact of nonignorable nonresponses. First, the extant ISNI methods require that all the covariates in a regression model be fully observed and are not computable with missing covariates, such as when the covariate *Social* is subject to missingness. Second, the extant ISNI methods can fail to detect the nonlinear U-shaped impact of nonignorability around the MAR model when the regression outcome and covariates (e.g., *mood* and *Social* in this EMA data analysis) are subject to concurrent missingness.

To overcome these limitations of the extant ISNI methods in examining the likely impact of nonignorable prompt nonresponse on the MAR estimates, we apply our new NISNI method proposed in this paper to the EMA data. The results on both the ISNIL and ISNIQ values, as well as the calibrated range of parameter changes and *c* statistics are reported in Table 3. The

column “Est $\left(\gamma_{1y} = \pm \frac{1}{\sigma_y}\right)$ ” presents the approximate range of the estimates when

$\gamma_{1y} = \pm \frac{1}{\sigma_y}$. This examines the range of MLE estimates for a moderate nonignorability such that one standard deviation change in Y (i.e., mood) is associated with an odds ratio of prompt response being $e^1 = 2.7$ or $e^{-1} = 0.37$. With this moderate nonignorability, the range of possible MLE coefficient estimates for *SmkEvent* is (0.862, 1.362) with the left endpoint of the sensitivity interval having a p-value of 0.07. The c statistics are 1.93 under both the ISNIL and ISNIQ, suggesting that the potential impact of nonignorable nonresponse on the estimate is small relative to its sampling error. The overall conclusion here is that for the MAR estimate of the *SmkEvent* parameter, strong non-ignorability is needed to have selection bias due to nonignorable nonresponse to be comparable to sampling error although the strength of statistical evidence for the mood-smoking relationship could be somewhat reduced in that the p-value could increase from 0.02 to 0.07 for moderate non-ignorability. One may consider ways to reduce sampling error of the mood-smoking relationship MAR estimate to increase the robustness of its testing result to nonignorable missingness. The c statistic for the *intercept* parameter in the *mood* outcome model using ISNIQ (=0.86) is noticeably smaller than that using ISNIL (=1.04), suggesting larger sensitivity detected when permitting the nonlinear impact of nonignorable nonresponse. The sensitivity of the *intercept* parameter in the *mood* outcome regression model is understandable as it measures the conditional mean of mood from the random prompts which is likely subject to sensitivity to selective missingness. The most significant difference between the linear and nonlinear ISNI analysis is for the parameter estimate of *Social*. The c statistic using ISNIQ is 0.96 while it is ∞ when using the ISNIL alone, indicating that the linear ISNI analysis is unable to detect any impact of nonignorability on the coefficient of the covariate *Social* whereas there are in fact significant changes in parameter estimates relative to sampling error. In this case, it is critical to use nonlinear indices to measure properly the impact of nonignorability.

We further conduct nonlinear ISNI analysis that permits *mood*-and-*Social*-dependent non-ignorability where the nonresponse probabilities are allowed to depend on both mood and social isolation, with results reported in Table 4. As compared with results assuming *mood*-dependently-only nonignorability, the c statistics are somewhat smaller. Despite numerical differences in the calibrated range of parameter changes and the c values, the sensitivity analysis results remain qualitatively unchanged with respect to the type of nonignorability (*mood*-and-*Social*-dependency vs. *mood*-dependency-only).

7 Discussion

Missing data are the rules rather than the exception in many studies. There is a strong need to evaluate the impact of nonignorable missingness on standard MAR analysis so that researchers can judge when such nonignorable missingness may be a concern and require attention. For example, systematic biases in nonresponses may affect conclusions about the effects of controlling variables or patterns of health behaviors, therefore demanding quantifications of such potential biases. Modern electronic data capturing methods, such as the EMA methods, can collect a large amount of detailed data and raise new challenges

when evaluating the impact of nonignorability in these contexts, including computational prohibitiveness and more complex effects of nonignorability. Developing simple-to-compute index methods to measure the sensitivity to nonignorability tailored for these new type of data is attractive. This research represents our first step to develop such tools for these emerging types of data, including those from EMA studies.

Prior research has focused on the linear sensitivity index for the situation of missing response outcomes. In this work, we have developed more general nonlinear sensitivity index measures that can handle concurrent missingness in response and covariates and more complex nonignorability mechanisms, and can capture a highly nonlinear impact of nonignorability around the MAR model. Our data analyses demonstrate that the existing linear sensitivity index can fail to detect the dependence of the estimates on the MAR assumption whereas the newly developed nonlinear index measures can, and the difference can lead to qualitatively different conclusions regarding the impact of nonignorability on the MAR estimates. At the same time, these new indices maintain computational simplicity by avoiding fitting any complicated nonignorable models, and thus are well suited for use in rich data environment nowadays.

In practice, our sensitivity indices can be useful in the following ways. First, the investigators can use the indices to incorporate sensitivity analysis results into the primary reporting. Tables 3 and 4 provide examples of how our sensitivity analysis results can be incorporated into the primary reporting and be used to evaluate the impact of moderate and plausible nonignorability on parameter estimates relative to sampling errors.

Second, our sensitivity indices can be used for informing more efficient efforts to collect additional data. The investigator may choose to collect additional data, such as refreshment samples [23], to better understand the missing data mechanism, and then explore the use of the information regarding the missing data mechanism provided by the additional data. These efforts would require additional resources and time in collecting data as well as constructing and fitting more complicated models. Our method can be useful for quickly screening for datasets in which the impact of nonignorability is important before deciding to invest a great deal of valuable resources to collect additional data and to perform arduous modeling.

Although this research is motivated by the nonresponse issue in EMA data, there are a much broader range of applications for the NISNI method developed here. First, the proposed method can easily generalize to a variety of relevant research areas, including the assessment of pain and symptoms as well as diet and exercise, and to other types of real-time data that become increasingly available because of technological advancement. Second, similar issues occur often in more traditional types of data. For example, when a subject misses one visit in a clinical trial study or in an observational cohort study using more traditional data collection methods, both the outcome and subject-specific time varying covariates to be collected in the missed visit will then be missing simultaneously, as noted in [15]. This creates the same type of missing data pattern as considered here for EMA data with prompt nonresponses. The NISNI method developed here can also be applied in these more traditional types of data to quantify the potential impact of nonignorability.

One limitation of this paper is that we restrict our analysis to the cross-sectional EMA data. Although the cross-sectional EMA data maintains the benefit of ensuring ecological validity of data by measuring the participants' current states in everyday environments and minimizing the recall bias as compared with traditional retrospective data, the analysis does not exploit its intensive longitudinal information within the entire EMA measurement period. It is natural to extend the methodology presented here to the longitudinal setting, which will involve the joint longitudinal modeling and analysis of the outcome and covariates that are subject to missingness. This involves substantially more modeling and analytical work, which we plan to pursue in the future.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Research reported in this work was supported by the National Cancer Institute [grants R01CA178061 and P01CA98262]. We thank an Associate Editor, referees and seminar participants at Cornell, Duke, Emory, Fred Hutchinson Cancer Research Center, Fudan University, Georgia State University, Simon Fraser University, SUNY at Buffalo, University of Alberta, University of British Columbia, University of Hong Kong, and the Joint Statistical Meeting for their constructive comments that led to improvements in the article.

References

1. Rubin DB. Inference and missing data (with discussion). *Biometrika*. 1976; 63:581–592.
2. Little RJA. Modeling the drop-out mechanism in longitudinal studies. *Journal of the American Statistical Association*. 1995; 90:1112–121.
3. Molenberghs, G.; Verbeke, G. *Models for discrete longitudinal data*. Springer; New York, NY: 2005.
4. Daniels, MJ.; Hogan, JW. *Missing data in longitudinal studies: strategies for Bayesian modeling and sensitivity analysis*. Chapman & Hall; Boca Raton: 2008.
5. Copas JB, Li HG. Inference for non-random samples (with discussion). *Journal of the Royal Statistical Society B*. 1997; 59:55–95.
6. Copas JB, Eguchi S. Local sensitivity approximations for selectivity bias. *Journal of the Royal Statistical Society B*. 2001; 63:871–895.
7. Verbeke G, Molenberghs G, Thijs H, Lesaffre E, Kenward MG. Sensitivity analysis for nonrandom dropout: a local influence approach. *Biometrics*. 2001; 57:7–14. [PubMed: 11252620]
8. Troxel AB, Ma G, Heitjan DF. An index of local sensitivity to nonignorability. *Statistica Sinica*. 2004; 14:1221–1237.
9. Xie H, Heitjan DF. Sensitivity analysis of causal inference in a clinical trial subject to crossover. *Clinical Trials*. 2004; 1:21–30. [PubMed: 16281459]
10. Ma G, Troxel AB, Heitjan DF. An index of local sensitivity to nonignorable dropout in longitudinal modeling. *Statistics in Medicine*. 2005; 24:2129–2150. [PubMed: 15909292]
11. Xie H. A local sensitivity analysis Approach to longitudinal non-gaussian data with non-ignorable dropout. *Statistics in Medicine*. 2008; 27:3155–3177. [PubMed: 17948917]
12. Xie H. Bayesian inference from incomplete longitudinal data: a simple method to quantify sensitivity to nonignorable dropout. *Statistics in Medicine*. 2009; 28:2725–2747. [PubMed: 19572257]
13. Xie H. Analyzing longitudinal clinical trial data with nonignorable missingness and unknown missingness reasons. *Computational Statistics & Data Analysis*. 2012; 56:1287–1300.
14. Xie H, Heitjan DF. Local sensitivity to nonignorability: dependence on the assumed dropout mechanism. *Statistics in Biopharmaceutical Research*. 2009; 29:243–257.

15. Xie H, Qian Y. Measuring the impact of nonignorability in panel data with non-monotone nonresponse. *Journal of Applied Econometrics*. 2012; 27:129–159.
16. Xie H. Adjusting for nonignorable missingness when estimating generalized additive models. *Biometrical Journal*. 2010; 52:186–200. [PubMed: 20422637]
17. Xie H, Qian Y, Qu L. A semiparametric approach for analyzing nonignorable missing data. *Statistica Sinica*. 2011; 21:1881–1899.
18. Zhang J, Heitjan DF. Impact of nonignorable coarsening on bayesian inference. *Biostatistics*. 2007; 8:722–743. [PubMed: 17215331]
19. Diggle JP, Kenward MG. Informative drop-out in longitudinal data analysis (with discussion). *Journal of the Royal Statistical Society C*. 1994; 43:49–73.
20. Troxel AB, Harrington DP, Lipsitz SR. Analysis of longitudinal measurements with non-ignorable non-monotone missing values. *Applied Statistics*. 1998; 47:425–438.
21. Kenward MG. Selection models for repeated measurements with non-random dropout: an illustration of sensitivity. *Statistics in Medicine*. 1998; 17:2723–2732. [PubMed: 9881418]
22. Molenberghs G, Beunckens C, Sotito C, Kenward M. Every missing not at random model has got a missing at random counterpart with equal fit. *Journal of the Royal Statistical Society B*. 2008; 70:371–388.
23. Hirano K, Ridder GW, Rubin DB. Combining panels with attrition and refreshment samples. *Econometrica*. 2001; 69:1645–1659.
24. Vach W, Blettner M. Logistic regression with incompletely observed categorical covariates — investigating the sensitivity against violation of the missing at random assumption. *Statistics in Medicine*. 1995; 14:1315–1329. [PubMed: 7569490]
25. Rotnitzky A, Robins JM, Scharfstein DO. Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association*. 1998; 93:1321–1339.
26. Scharfstein D, Rotnitzky A, Robins JM. Adjusting for nonignorable dropout using semiparametric models (with discussion). *Journal of the American Statistical Association*. 1999; 94:1096–1146.
27. Park Y, Tian L, Wei LJ. One- and two-sample nonparametric inference procedures in the presence of a mixture of independent and dependent censoring. *Biostatistics*. 2006; 7:252–267. [PubMed: 16282531]
28. Stone, AA.; Shiffman, S.; Atienza, A.; Nebeling, L. *The science of real-time data capture*. Oxford University Press; New York: 2007.
29. Heitjan DF, Rubin DR. Ignorability and coarse Data. *The Annals of Statistics*. 1991; 19:2244–2253.
30. Heitjan DF. Ignorability, sufficiency and ancillarity. *Journal of Royal Statistical Society B*. 1997; 59:375–381.
31. Chen HY, Xie H, Qian Y. Multiple Imputation for Missing Values through Conditional Semiparametric Odds Ratio Models. *Biometrics*. 2011; 67:799–809. [PubMed: 21210771]
32. Little RJA. Regression with missing X's: a review. *Journal of the American Statistical Association*. 1992; 87:1127–1137.
33. Shiffman S. Dynamic influences on smoking relapse process. *Journal of Personality*. 2005; 73:1715–1748. [PubMed: 16274451]
34. Mermelstein, R.; Hedeker, D.; Flay, B.; Shiffman, S. *Real-time data capture and adolescent cigarette smoking: moods and smoking (the science of real-time data capture: self-report in health research)*. Oxford University Press; New York: 2007. p. 117-135.
35. Whalen CK, Jamner LD, Henker B, Delfino RJ. Smoking and moods in adolescents with depressive and aggressive dispositions evidence from surveys and electronic diaries. *Health Psychology*. 2001; 20:99–111. [PubMed: 11315734]

Figure 1.
Scatterplots of Simulated Data with Nonignorable Missingness. +: Missing observations

Figure 2.

ISNI Approximation to $\hat{\beta}_1(\gamma_1)$ with Y -Dependent-only Nonignorable Nonresponse.

Figure 3.

A Comparative Analysis of Two Datasets Simulated under $\gamma_1 = 1$, and -1 , respectively. +: Missing observations. The thick solid line represents the regression line fitted using the complete data. The missing observations in the two plots not only have very different Y values but also very different X values, which causes the slopes of the MAR regression lines (the thick dashed lines in the two plots) fitted using observed data to have bias in the same direction. The dotted horizontal and vertical lines represent the average value of y and x using observed data, respectively.

Figure 4.

ISNI Approximation to $\hat{\beta}_1(\gamma_1)$ with Y -and X -Dependent Nonignorable Nonresponse. The solid surface is the exact sensitivity surface that plots $\hat{\beta}_1(\gamma_{1x}, \gamma_{1y})$ for a grid of values of $(\gamma_{1x}, \gamma_{1y})$, which involves fitting complicated nonignorable models repeatedly at these grid values that takes hours to compute. The points denote the estimates using the nonlinear ISNI (NISNI) method which avoids fitting any nonignorable model and takes only a few seconds to compute. The at plane formed by the dotted lines represents the estimates using the linear ISNI method.

Table 1

An Application of NISNI to Simulated Data For Simple Linear Regressions with Y -dependent-only Nonignorable Nonresponse.

γ_0	γ_{1y}	Prop_M	$\hat{\beta}_1(0)$ (S.E)	$\frac{\Delta \hat{\beta}_1}{\Delta \gamma_{1y}}$	$\frac{\Delta \hat{\beta}_1}{\Delta \gamma_{1y}^2}$	ISNIL	ISNIQ	c
1	1	33.8%	0.81 (0.04)	5e-16	0.292	0	0.290	0.59
	0.5	29.5%	0.94 (0.04)	-2e-15	0.398	0	0.394	0.60
	-1	31.0%	0.91 (0.04)	5e-16	0.328	0	0.325	0.60
	-0.5	29.8%	0.97 (0.04)	-1e-15	0.408	0	0.406	0.61
2	1	17.4%	0.88 (0.03)	-1e-4	0.220	0	0.219	0.70
	0.5	14.2%	0.98 (0.04)	-6e-16	0.240	0	0.236	0.76
	-1	18.1%	0.95 (0.04)	-8e-5	0.260	0	0.254	0.69
	-0.5	13.8%	0.94 (0.03)	4e-5	0.207	0	0.201	0.77
5	1	1.2%	0.96 (0.03)	-1e-15	0.025	0	0.020	2.54
	0.5	1.1%	0.99 (0.03)	-1e-16	0.020	0	0.018	2.67
	-1	1.7%	0.95 (0.03)	8e-5	0.038	0	0.029	2.08
	-0.5	0.4%	1.05 (0.03)	-3e-15	0.025	0	0.01	5.65

Table 2

An Application of NISNI to Simulated Data For Simple Linear Regressions with Y -and- X -dependent Nonignorable Nonresponse.

γ_0	γ_{1y}	γ_{1x}	Prop_M	$\hat{\beta}_1(0)(S.E.)$	$\frac{\Delta \hat{\beta}_1}{\Delta \gamma_1}$	$\frac{\Delta^2 \hat{\beta}_1}{\Delta \gamma_1 \Delta \gamma_1^T}$	ISNIL	ISNIQ	c
1	$\sqrt{0.5}$	$\sqrt{0.5}$	31.4%	0.82 (0.04)	(-5e-06, 7e-05)	(0.318, 0.192, -2e-03)	(0, 0)	(0.320, 0.194, 0)	0.52
	$0.5 / \sqrt{2}$	$0.5 / \sqrt{2}$	31.6%	0.90 (0.04)	(-6e-06, -5e-06)	(0.384, 0.211, -4e-04)	(0, 0)	(0.384, 0.212, 0)	0.50
	$-\sqrt{0.5}$	$-\sqrt{0.5}$	32.7%	0.89 (0.04)	(-5e-05, 4e-05)	(0.406, 0.224, -2e-03)	(0, 0)	(0.404, 0.225, 0)	0.49
	$-0.5\sqrt{2}$	$-0.5\sqrt{2}$	28.2%	0.93 (0.04)	(-2e-05, 5e-06)	(0.345, 0.185, -8e-04)	(0, 0)	(0.348, 0.185, 0)	0.51
5	$\sqrt{0.5}$	$-\sqrt{0.5}$	28.9%	0.98 (0.04)	(2e-05, -2e-05)	(0.364, 0.183, -4e-04)	(0, 0)	(0.364, 0.184, 0)	0.52
	$0.5 / \sqrt{2}$	$-0.5 / \sqrt{2}$	27.9%	1.02 (0.04)	(2e-05, -1e-05)	(0.399, 0.196, -2e-03)	(0, 0)	(0.401, 0.197, 0)	0.52
	$\sqrt{0.5}$	$-\sqrt{0.5}$	28.8%	0.99 (0.04)	(6e-06, -1e-05)	(0.389, 0.195, -1e-03)	(0, 0)	(0.389, 0.196, 0)	0.51
	$0.5 / \sqrt{2}$	$-0.5 / \sqrt{2}$	28.3%	1.05 (0.04)	(4e-16, -4e-06)	(0.405, 0.192, -5e-04)	(0, 0)	(0.405, 0.192, 0)	0.52
2	$\sqrt{0.5}$	$\sqrt{0.5}$	21.4%	0.95 (0.04)	(-8e-06, 1e-15)	(0.297, 0.158, 3e-04)	(0, 0)	(0.299, 0.157, 0)	0.56
	$0.5 / \sqrt{2}$	$0.5 / \sqrt{2}$	15.1%	0.94 (0.04)	(2e-15, 2e-15)	(0.243, 0.131, -2e-04)	(0, 0)	(0.242, 0.129, 0)	0.62
	$-0.5 / \sqrt{2}$	$-0.5 / \sqrt{2}$	18.3%	0.91 (0.04)	(2e-05, -1e-05)	(0.276, 0.153, -3e-03)	(0, 0)	(0.275, 0.152, 0)	0.57
	$\sqrt{0.5}$	$-\sqrt{0.5}$	16.4%	0.97 (0.04)	(-3e-06, -1e-05)	(0.233, 0.119, -2e-03)	(0, 0)	(0.233, 0.119, 0)	0.61
	$0.5 / \sqrt{2}$	$-0.5\sqrt{2}$	14.2%	0.99 (0.03)	(-9e-06, 5e-06)	(0.227, 0.112, -1e-04)	(0, 0)	(0.225, 0.113, 0)	0.64
	$0.5 / \sqrt{2}$	$-0.5 / \sqrt{2}$	12.8%	0.96 (0.04)	(-8e-06, -1e-05)	(0.211, 0.110, -4e-05)	(0, 0)	(0.211, 0.109, 0)	0.67
	$\sqrt{-0.5}$	$\sqrt{0.5}$	14.1%	1.02 (0.03)	(1e-05, 1e-05)	(0.243, 0.119, -2e-03)	(0, 0)	(0.244, 0.119, 0)	0.65
	$0.5 / \sqrt{2}$	$-0.5 / \sqrt{2}$	13.3%	0.96 (0.03)	(4e-06, -6e-06)	(0.210, 0.108, -1e-03)	(0, 0)	(0.209, 0.109, 0)	0.65
	$-\sqrt{0.5}$	$\sqrt{0.5}$	1.6%	0.98 (0.03)	(-6e-06, 5e-06)	(0.032, 0.016, 2e-03)	(0, 0)	(0.027, 0.014, 0)	2.1
	$-0.5 / \sqrt{2}$	$0.5 / \sqrt{2}$	1.2%	1.00 (0.03)	(1e-06, 6e-06)	(0.025, 0.011, 3e-04)	(0, 0)	(0.018, 0.009, 0)	2.52
	$0.5 / \sqrt{2}$	$0.5 / \sqrt{2}$	2.6%	0.97 (0.03)	(-4e-06, -3e-06)	(0.049, 0.025, -6e-04)	(0, 0)	(0.045, 0.023, 0)	1.64
	$-0.5 / \sqrt{0.5}$	$-0.5 / \sqrt{0.5}$	0.6%	1.03 (0.03)	(-5e-05, -1e-06)	(0.016, 0.007, -3e-03)	(0, 0)	(0.008, 0.004, 0)	3.79
	$\sqrt{0.5}$	$-\sqrt{0.5}$	0.5%	1.05 (0.03)	(-1e-05, 4e-06)	(0.011, 0.005, 3e-03)	(0, 0)	(0.006, 0.003, 0)	4.02
	$0.5 / \sqrt{2}$	$-0.5 / \sqrt{2}$	0.6%	0.98 (0.03)	(-6e-15, -6e-15)	(0.005, 0.004, -4e-03)	(0, 0)	(0.007, 0.004, 0)	3.98
	$-\sqrt{0.5}$	$\sqrt{0.5}$	0.7%	1.07 (0.03)	(-8e-06, e-05)	(0.020, 0.008, -1e-03)	(0, 0)	(0.011, 0.005, 0)	3.57
	$-0.5\sqrt{2}$	$0.5\sqrt{2}$	0.6%	1.02 (0.03)	(5e-15, 5e-15)	(0.004, 0.002, 2e-03)	(0, 0)	(0.007, 0.003, 0)	3.98

Table 3

NISNI Analysis of Smoking-Mood Relationship Using EMA Data Assuming Y -Dependent-only
Nonignorable Nonresponse

Parameter	MAR Est	SE	ISNIL _y	ISNIQ _{yy}	Linear Correction		Nonlinear Correction	
					$\text{Est}\left(\gamma_{1y} = \pm \frac{1}{\sigma_Y}\right)$	c	$\text{Est}\left(\gamma_{1y} = \pm \frac{1}{\sigma_Y}\right)$	c
Intercept	0.342	0.518	-1.324	-1.728	[-0.156, 0.840]	1.04	[-0.279, 0.718]	0.86
SmkEvent	1.112	0.481	0.666	0.003	[0.862, 1.362]	1.93	[0.862, 1.363]	1.93
Social	0.750	0.049	0	0.709	0.750	∞	[0.750, 0.810]	0.96

Table 4

NISNI Analysis of Smoking-Mood Relationship Using EMA Data Assuming Y -and-X-Dependent Nonignorable Nonresponse

Parameter	MAR	SE	ISNIL	ISNIQ	Linear Correction		Nonlinear Correction	
					$\text{Est} \left(\gamma_{1y} = \pm \frac{1}{\sigma_Y} \right)$	c	$\text{Est} \left(\ \gamma_1^* \ = 1 \right)$	c
Intercept	0.342	0.518	(−1.324, 0)	(−1.728, −1.018, 0)	[−0.156, 0.840]	1.04	[−0.289, 0.741]	0.85
SmkEvent	1.112	0.481	(0.666, 0)	(0.003, 0.001, 0)	[0.862, 1.362]	1.93	[0.862, 1.363]	1.93
Social	0.750	0.049	(0, 0)	(0.709, 0.474, 0)	0.750	OO	[0.738, 0.823]	0.88