Local Nonlinear Sensitivity to Nonigorable Selection

by

Weihua Gao
BS, JiLin Institute of Technology, Changchun, China, 1998
MS, University of Illinois at Chicago, Chicago, Illinois, 2004

Thesis submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Public Health Sciences
in the Graduate College of the
University of Illinois at Chicago, 2018

Chicago, Illinois

Defense Committee:
Hui Xie, PhD, Associate Professor of Biostatistics, UIC, Advisor and Chair
Michael Berbaum, PhD, IHRP Methodology Research Core Director, UIC
Hua Yun Chen, PhD, Professor of Biostatistics, UIC
Robin Mermelstein, PhD, Professor of Psychology and IHRP Director, UIC
Donald Hedeker, PhD, Professor of Biostatistics, U of Chicago

Dedicated to my parents, husband Baodong, and kids Gloria and Grace

ACKNOWLEDGMENT

First, I would like to express my sincere gratitude to my advisor Dr. Hui Xie for introducing me into the field of missing-data analysis, for his continuous support of my Ph.D study, and for his patience, motivation, and immense knowledge. His guidance helped me enormously. My research reported in this work was also supported by Dr. Xie's grant (statistical qualification of the impact of missing data in EMA studies (R01CA178061, H. Xie, PI)), funded by the National Cancer Institute of the National Institutes of Health.

I am grateful to my committee members: Dr. Michael Berbaum, Dr. Hua Yun Chen, Dr. Donald Hedeker, and Dr. Robin Mermelstein. Their thoughtful suggestions motivated great improvement on my dissertation. Particularly, I would like to thank Dr. Mermelstein for providing me with great datasets in Application section (NIH-funded longitudinal study of the natural history of smoking among adolescents (PO1CA098262, R. Mermelstein, PI)).

I also want to acknowledge the faculty and staff members of the Epidemiology and Biostatistics Division in School of Public Health for their great support during my studying at UIC.

Lastly, I am grateful to my parents, my husband, and my kids for their constant encouragement and emotional support in my life. I am also grateful to other family members and friends who have supported me along the way.

Contribution of Authors

Chapter 1 is a literature review that places my dissertation question in the context of the larger field and highlights the significance of my research question. Chapter 2 represents a series of my own unpublished work directed at answering the question of what is nonlinear sensitivity index for selection when only the outcome is subject to selective observation. Partial of Chapter 3 and 4 represent a published manuscript for which I was the primary author and major driver of the research. My advisor, Dr. Hui Xie contributed to the writing of the manuscript. Chapter 5 represents my synthesis of the research presented in this dissertation and my overarching conclusions. The future directions of this field and this research question are discussed.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

CHAPTER **PAGE**

LIST OF TABLES

LIST OF TABLES (Continued)

TABLE **PAGE**

LIST OF FIGURES

LIST OF ABBREVIATIONS

AT	As-Treated
EM	Expectation-Maximization
EMA	Ecological Momentary Assessment
GAM	Generalized Additive Model
GEE	Generalized Estimating Equation
IPCW	Inverse Probablity of Censoring Weighted
IQR	Inter-Quantile Range
ISNI	Index of Sensitivity to Nonignorablity
ISNIL	Index of Sensitivity to Nonignorablity in Linear form
ISNIQ	Index of Sensitivity to Nonignorablity in Quadratic form
MI	Multiple Imputation
MCAR	Missing Completely At Random
MAR	Missing At Random
MNAR	Missing Not At Random
ML	Maximum Likelihood

LIST OF ABBREVIATIONS (Continued)

MLE	Maximum Likelihood Estimate
MS	Multiple Sclerosis
NCI	National Cancer Institute
NISNI	Nonlinear Index of Sensitivity to Nonignorablity
SD	Standard Deviation
RS	Regression Spline

CHAPTER 1

INTRODUCTION

It is common for incomplete data to be encountered in many types of research, such as surveys, clinical trials, market data, and especially in longitudinal studies. In observational studies, missing data can arise because respondents refuse to answer a question owing to privacy issues, respondents taking the survey do not understand the question, or respondents do not have enough time to complete the questionnaire. In a randomized clinical trial, missing data can arise when lab measurements are very difficult to obtain. Patients drop out from the treatment due to side effects during the trial. Furthermore missing data can cause serious problems. Researchers may not have enough data to perform the analysis and may lose power to obtain statistically significant results. The results may be misleading and biased. Missing data really challenge researchers' ability to draw valid and conclusive inferences. Below are two illustrative examples that will be used later in this paper.

Example 1: Uncontrolled crossover in clinical trials. Randomized clinical trials are the gold standard in evaluating treatment effectiveness. However, patient noncompliance such as dropout, non-response, or self-switching treatment arms after randomization can ruin randomization, such that the resulting data are not random samples of the population of interest. This example is a randomized clinical trial in multiple sclerosis (1). In this subset of the data, 14 subjects were randomized to a placebo control and 11 to treatment with azathioprine and methylprednisolone. The study is interested in whether the test treatment reduces levels of

AD25, an assay of immune function defined as the antibody-dependent cellular cytotoxicity at an effector:target ratio of 25:1. The complication arises because there were 3 patients who were switched from treatment to placebo. There were no crossovers from placebo to treatment arm because subjects were not allowed to switch to active treatment. Because of the crossover, the outcome for those crossovers under the originally assigned treatment arms are missing. If these crossovers are systematically different from the remaining subjects, the observed data would be a biased representation of population of interest and consequently can introduce bias to the analysis.

Example 2: Non-responses in Ecological Momentary Assessment (EMA) studies. As a real-time data capturing method, it has become increasingly important in health studies. In particular, EMA has become a new and vital approach to understanding health-related behaviors, e.g., in smoking and cancer research, as detailed by (2). They wrote"The National Cancer Institute (NCI) has designated the topic of real-time data capture as an important and innovative research area." It also discussed "the state of the science of real-time data capture and its application to health and cancer research." 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. Like in other studies involving human subjects, missing data due to non-response is common in EMA studies.

1.1 Methods for Missing Data

Missing data inevitably occurs in the majority of studies despite efforts to minimize missingness through design and data collection stages. Researchers have to decide how to deal with it. There are several general methods to handle missing data. (3) summarized four methods to handle missing data. The first method is the simplest approach called complete case analysis, which is also known as listwise deletion. This method excludes missing cases and only analyzes the complete cases. In other words, the inference drawn is based on the subjects with the observed data. This method usually leads to biased estimates and loses power, especially when having a large amount of missing data. If the observed data is a random sample from the full data with a small portion of missingness, unbiased parameters might be achieved. But the analysis still has potential loss of precision.

The second method is weighting, extended from complete case analysis. This approach adapts sampling method in survey data and involves an estimate of the probability of completeness. Therefore the weight is inversely proportional to their probability of selection multiplied by the probability of completeness. The estimate of population mean is expressed as

$$\sum_{i=1}^{n}(\pi_i\hat{p}_i)^{-1}y_i/\sum_{i=1}^{n}(\pi_i\hat{p}_i)^{-1}$$

where y_i is observed data, π_i is the known probability of selection into the sample, and $\hat{p}_i$ is the probability of completeness. This method can reduce estimates' bias from the complete case analysis. But the computation of variance using this method is not straight-forward and is intensive.

The third method is imputation. Obviously imputation means filling in the missing data by a value generated using different imputation approaches. After conducting imputation, the "new" data will be analyzed as if they were not missing. Imputations can be single imputation and multiple imputation. The following imputation methods are single imputation methods. Hot deck imputation was used quite widely in the Census Bureau as far back as 60 years ago. Another earlier imputation approach is called mean imputation. The average of the completed data is filled in the missing data. A relatively more advanced imputation is regression imputation. Using regression model to predict the missing data based on the other observed variables. And then the missing data is filled by the predicted value. All these imputation approaches have a common characteristics in that the imputation increases the sample size and reduces the variance. The single imputation methods do not take into account of uncertainty created by missing data. To solve this problem, Multiple Imputation (MI) was developed by (4). Its procedure firstly generates a set of imputed datasets with choice of imputation mechanisms according to the missing data pattern. Then standard statistical methods are used to analyze each of the imputed datasets. Lastly, the estimates of parameters of interest are averaged across multiple imputed datasets except the standard error, which is calculated based on combination of the within variance of each imputed dataset and the variance between the imputed datasets. Thus MI incorporates missing-data uncertainty.

The last method Rubin described is a model-based procedure called likelihood-based procedure. This approach models the observed data, draws inferences based on likelihood, and estimates the parameters of interest by maximizing likelihood. The estimates of variance by this procedure will consider missingness in the data. To take into account the effect of missing data, one could generate a likelihood including a model of the missing data mechanism besides the model for observed data. Therefore this method has an advantage of flexibility. If models are correctly specified, the inferences drawn based on the model are more efficient compared to the methods mentioned above. (5) defined a full model including both the distributions of data and missing-data mechanism. The distribution of missing-data mechanism is actually the conditional distribution of the missing-data indicator (G) conditioning on the data (Y), $f_{\gamma}^{G|Y}(g)$, where $Y = (Y^{mis}, Y^{obs})$ and Y^{obs} denotes the observed data and Y^{mis} denotes the missing data, G is a completeness indicator (1 represents the datum was observed and 0 represents the datum was missing), and γ is a vector of unknown parameters in distribution of missing-data mechanism. Then the density of the full model is described as joint densities of the distribution of Y and the conditional distribution of G on Y. That is

$$f(Y^{obs}, Y^{mis}, G|\theta, \gamma) = f(Y^{obs}, Y^{mis}|\theta) f(G|Y^{obs}, Y^{mis}, \gamma)$$

Where θ is a vector of the parameters of interest. To obtain complete likelihood (L_C), Y^{miss} should be integrated out from the joint density of (Y,G) as follows

$$f(Y^{obs}, G|\theta, \gamma) = \int f(Y^{obs}, Y^{mis}|\theta) f(G|Y^{obs}, Y^{mis}, \gamma) dY^{mis}$$

Thus the complete likelihood of (θ, γ) is any function of θ and γ proportional to the above equation, which is

$$L_C(\theta, \gamma|Y^{obs}, G) \sim f(Y^{obs}, G|\theta, \gamma)$$

The model based on the complete likelihood is called a selection model (6). The selection model includes two parts: a complete-data model ($f(Y^{obs}|\theta)$ and a missing-data mechanism model ($f(G|Y^{obs}, \gamma)$). We will focus on the selection model in this paper.

1.2 Missing Data Mechanism

A data analyst often faces challenges in modeling the missing-data mechanism, as usually little information is provided for the missing-data. Knowing types of the missing data mechanisms is very important and necessary to help the statistician determine whether the analysis can ignore the missing-data mechanism. (5) has clearly described three missing data mechanisms. The first missing data mechanism is called Missing Completely At Random (MCAR) if missingness does not depend on the data values, missing or observed. Formally, MCAR holds when $f_\gamma^{G|Y}(g) = f_\gamma^{G}(g)$ for all $Y = (Y^{obs}, Y^{mis})$ and γ. For example, in a clinical trial, data on birthweight is missing because the process of transfering these data to a central server is interrupted, which does not relate to any characteristic of the child. The second missing data mechanism is called Missing At Random (MAR) if missingness only depends on the observed data values. Formally, MAR holds when $f_\gamma^{G|Y}(g) = f_\gamma^{G|Y^{obs}}(g)$ for all Y^{mis} and γ. For example, data on birthweight are missing due to observed characteristics, such as child gender, mother's smoking status and so on. The last missing data mechanism is called Missing Not At Random (MNAR) if missingness depends on the missing data values. Formally, MNAR hold when $f_\gamma^{G|Y}(g) = f_\gamma^{G|Y^{obs},Y^{mis}}(g)$ for all Y^{mis} and γ. For example, participants failed to complete the asthma symptom severity rating because of their severe asthma symptoms. Missingness with the MNAR mechaism is called nonignorable missingness.

1.3 Nonignorable Modeling

To avoiding estimating parameters in the missing-data mechanism model, many standard analyses on the available data need to invoke the strong and untestable assumption of ignorable selection in the sense of ignoring the missing observation and deriving a simple likelihood (L_I), which is $L_I(\theta|Y^{obs}) = f(Y^{obs}|\theta)$. It is intended as an approximation to the truth. Very often this occurs when the observed data is selective and thus unrepresentative of the underlying population. For instance, frequently we can only have or use a subset of a random sample because at least one element of ideal data is unobserved for some study units either due to sample design or uncontrollable behaviors of study units. In all these situations, the observed data can be considered as a selective subset of a random sample from the underlying population. In the previous section, three missing data mechanisms were defined. The sufficient condition of ignorailty for modeling the missing-data mechanism is the data with MCAR or MAR mechanism and the parameters of interest in the completed-data model and the parameters in the completeness indicator model ($f_Y^{G|Y}(g)$) are distinct. Therefore the statistician can draw a conclusion based on a subsample of completed cases with unbiased estimates. The inference can be addressed based on maximizing a likelihood of the observed-data model or Bayesian analysis. For the ignorability situation, the inferences of θ based on this L_I will be the same as the ones based on L_C because L_I is proportional to L_C. Thus the inferences drawn from L_I are valid. When ignorability is questionable, in other words, the missing data is MNAR, the inference of θ from L_C is different from L_I. Standard statistical procedures ignoring this selective feature likely lead to biased parameter estimates and invalid statistical inference. A particularly challenging case is when the selection mechanism depends on the unobserved data. That is, selection is nonignorable where the probability of data elements being missing depends on the unobserved components of data matrix conditioning on the observed data (7). Thus L_C should be used for drawing inferences of θ for these circumstances.

A variety of popular models, including selection models, shared parameter models, and pattern-mixture models were developed to investigate nonignorable missingness. (8) discussed the selection models and pattern-mixture models for handling drop-out mechanism in longitudinal data. The selection model combines a model for the ideal complete data with a model that models the drop-out mechanism with a set of binary missing data indicators. Even though the selection models allow the statistician to directly use L_C to make inference, there exist well-known difficulties in using L_C to make inference. If the parameter γ_1 in the drop-out mechanism model associated with unobserved data are unidentifiable with flat likelihood, the joint estimation of θ and γ_1 is impossible, in the sense that unique ML estimates are not available, or weakly identified. To increase the model identifiability often requires the existence of instrumental variables or complementary data. However, clean and strong instrumental variables are often hard to identify or useful complementary data containing such instruments are often expensive to collect. Even when the model is identifiable, the inference can be highly sensitive to untestable distributional assumptions in (9), (10), and (11). The pattern-mixture models extend the selection models by stratifying the missing data into different patterns and then constructing a corresponding complete-data model within strata. In the pattern-mixture model framework, subjects in the same strata share the same pattern of missing data. The complete data model is estimated for each pattern and then the pattern-specific estimates are averaged into an overall result. The pattern-mixture model faces the same problems as the selection models because some parameters in the model cannot be estimated from data owing to missing (unobserved) data within the pattern. But one can impose some assumptions regarding the inestimable parameters, which are referred to "identifying restrictions". The shared parameter models in (12), (13), and (14) usually fit longitudinal data with nonignorable dropout. The shared paramter model shares a vector of random effects (b_i) between the complete data model and the missing data mechanism model. In the model, the data Y_i and the missing data mechanism indicator G_i are assumed to be conditionally independent of each other, given b_i. This allows us to essentially eliminate the density for the Y^{mis}. Maximum likelihood can be computationally difficult. Based on previous work, (15) proposed a stochastic EM algorithm to obtain the parameters estimates in the shared parameter model. These approaches require strict and untestable assumptions.

1.4 Sensitivity Analysis

1.4.1 Global Sensitivity Analysis

As discussed above, it is hard to verify that missingness and the unobserved values are unrelated. In general it is impossible to judge the validity of the assumption that missing data can be adequately predicted from the observed data alone. Consequently it is crucial to assess the reliability and validity of empirical findings in the presence of potentially selectively observed data. Existing research has shown that missing data analysis should not depend solely on a 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 in (8), (3), and (16). The recent book by the Panel on the Handling of Missing Data in Clinical Trials formed by the Division of Behavioral and Social Sciences and Education of National Research Council on behalf of the U.S. Food and Drug Administration (17) points out that "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" and "The treatment of missing data in clinical trials, being a crucial issue, should have a higher priority for sponsors of statistical research, such as the National Institutes of Health and the National Science Foundation". Although the guidelines above are primarily for regulating the analysis and reporting of clinical trial data, they have important implications for handling missing data issues in general and for the need to develop missing data methods suitable in broader settings. Evaluating the sensitivity of standard analysis assuming ignorable selection typically involves augmenting the outcome model with a selection model that describes how the ideal data are selected for observation. One can then examine the change in statistical inferences at models with different magnitude of nonignorablity parameters. If the inference varies substantially over the plausible range of nonignorability parameters, the inference based solely on the ignorability assumption (MLE obtained from $\mathrm{L_I}$) is considered questionable. Otherwise, the analysis based on $\mathrm{L_I}$ assuming MAR mechanism is considered trustworthy.

(18) and (19) present sensitivity analysis in the context of selection models for semi-parametric inference. Their first paper presented a method based on an augmented inverse

probability of censoring weighted (IPCW) estimator for making inferences about parameters β of a model for the conditional mean of longitudinal outcomes (Y_i) with nonignorable missingness under the condition that missing data probability can be modeled and the parameter is Consistent and Asymptotically Normal. Having $Y_i = (Y_{i1}, ..., Y_{iT})^T$ is a (Tx1) vector of the outcomes measured at time t and X_i is a (T X P) matrix of covariates for the ith subject and $\beta = (\beta_1, ..., \beta_P)$ is a (1xP) vector of parameters, where i = 1, ..., N, t = 1, ...T, and p = 1, ..., P, they proposed the following estimating equation for estimating β with repeated outcomes and nonignorable nonresponse:

$$\sum_{i=1}^{N} \left[\frac{I(R_i = 1)}{\hat{\pi}(1)} d(X_i; \beta) Y_i - g(X_i; \beta) + A_i \right] = 0,$$

where

$$A_i = \sum_{r \neq 1} \left\{ I(R_i = r) - \frac{I(R_i = 1)}{\hat{\pi}_i(1)} \hat{\pi}_i(r) \right\} \phi_r(W_{(r)i})$$

In the above equation, R_i is the vector of response patterns for subject i, $R_1 = 1$ indicates fully observed across all timepoints for ith subject; "$d(X_i; \beta)$ is a (P X T) matrix of fixed functions of X_i and β"(18); $g(X_i; \beta)$ is the mean vector of Y_i where g(.) is a smooth function of β; $\hat{\pi}(1)$ is the estimated "probability of observing the complete outcome vector Y"; $\hat{\pi}(r)$ is the estimated probability of response pattern $R_i = r$; $\phi_r(W_{(r)i})$ is an arbitrary, investigator-chosen vector function of the observed data at the time this subject dropped out. "Subjects with incomplete observations contribute to the estimation of β via the term A_i"(18) noted. Since their estimator is viewed as an extension of generalized estimating equation (GEE) estimators for nonignorable missing outcomes. It has the following advantages: (1) not requiring "full specification of a parametric likelihood"(18); (2) not requiring numerical integration; (3) robust consistency of estimate to model misspecification even in the presence of nonignorable missingness. Due to the common problem of unidentifiability in selection models, the authors performed sensitivity analysis regarding the nonresponse model parameter τ, which is equivalent to γ_1 in our setting. By varying τ within a sort of acceptable values, they evaluate the inference by the change of parameter β. The disadvantage of this approach is the difficulty in handling complex missing data patterns.

Their first paper (18) assumed a logistic regression model for the nonresponse probability, but they assumed a semi-parametric model for the nonresponse probability in their second paper (19). They modeled the nonresponse probability by modeling time to dropout using a Cox proportional hazards model. They proposed two models for the conditional hazard of dropouts. The first model assumed a hazard function of time to drop out Q

$$\lambda_Q(t|\bar{V}(T), Y) = \lambda_0(t|\bar{V}(t))exp(\alpha_0 Y)$$

In the equation above, $\bar{V}(T)$ is the history of all other variables that would be recorded through time t in the absence of dropout; α_0 is a scalar parameter. When $\alpha_0 \neq 0$, then drop-out is nonignorable; and $\lambda_0(t|\bar{V}(t))$ is an unrestricted positive function of t giving the process $\bar{V}(t)$. Due to the curse of dimensionality in moderate sample sizes, they proposed a second model

$$\lambda_Q(t|\bar{V}(T), Y) = \lambda_0(t)exp[\gamma_0^{'} W(t)]exp(\alpha_0 Y)$$

In this model, they imposed the restriction on $\lambda_0(t|\bar{V}(t))$ to $\lambda_0(t)\exp(\gamma_0^{'}W(t))$, where $\lambda_0(t)$ is an unspecified baseline hazard function; $W(t) = w(t, V(t)), w(.)$ is a known function that maps $(t, V(t))$ to R^q; γ_0 is an unknown parameter vector. In this setting, the parameters are jointly identified. Then they performed sensitivity analysis regarding the nonresponse model parameter α_0, which is equivalent of γ_1 in our setting. By varying α_0 within a sort of acceptable values, they evaluate the inference by the change of the parameter β.

(20) proposed a multivariate regression pattern-mixture model for analyzing longitudinal data with continuous variables Y(1) and Y(2) when values of Y(2) are nonignorable missing. Having a sample of N independent observations, $y = (y_1, \cdots, y_p)^T$, and covariates $x = (x_1, \cdots, x_q)^T$, where x is fully observed, $Y(1) = (Y_1, \cdots, Y_{p1})^T$ are observed for all cases; $Y(2) = (Y_{p1+1}, \cdots, Y_p)^T$ are observed for n_0 cases and missing for $n_1 = n - n_0$ cases. They also defined missingness indicator variable m with $m = 0$ if Y(2) is observed and $m = 1$ if Y(2) is missing. The proposed model is

$$(m|x) \sim_{ind} \mathrm{Bernoulli}(p(x|\pi)), \text{ where } \log(p(x|\pi)/(1 - p(x|\pi))) = \pi^T x$$

$$(y|x, m = k) \sim_{ind} N_p(B^{(k)}x, \Sigma^{(k)})$$

There exists two patterns, k = 0, 1. The parameters of the model are the (q x 1) vector π and $\phi(k) = (B(k), (k))$. The parameters $\phi^{(1)}$ are clearly underidentified, Little assumed a missingness mechanism such that

$$\Pr(m = 1|y(1), y(2), x) = f(Cy(1) + \Lambda y(2), x)$$

Where f is an arbitrary unspecified function, C and Λ are known matrices and Λ is full rank matrix. When Λ is unknown, they performed sensitivity analysis for a range of plausible choices of Λ.

(21) presented the sensitivity analysis under nonignorable dropout in reparameterizing the pattern-mixture model. They also consider a multivariate regression pattern-mixture model. Based on the (20) methods, Daniels and Hogan rewrote the pattern-mixture model in terms of between-pattern location and scale changes. In this way, sensitivity analyses were reduced to a series of complete-data problems. Assuming dropout is monotone, the missingness pattern indicator is $u_i \in 1, \cdots, K$, where K denotes completeness and $\pi^{(k)} = pr(u_i = k)$. They introduced two unidentified components δ and C, such that

$$\delta^{(k)} = \mu^{(k)} - \mu^{(k+1)} \text{ and } C^{(k)} = (\Sigma^{(k)})^{(1/2)}(\Sigma^{(k+1)})^{-1/2}$$

For $k = 1, \cdots, K - 1$. Therefore their sensitivity analyses of posterior inferences about the difference in treatment arm means were conducted upon the varied combinations of $\delta^{(k)}, C^{(k)}$ under differences between unobserved data within a pattern and patterns with more complete data, such as starting with MAR and then moving to nonignorable mechanisms. The normal-mixture pattern-mixture model has an attractive feature that the marginal distribution of the observed data is fixed or nearly fixed when evaluating unidentified δ and C in a wide range of nonignorable missing data mechanisms

1.4.2 Local Sensitivity Analysis

The above sensitivity analyses are called *global* sensitivity analysis. There is an increasing interest in *local* sensitivity approximation of selection bias due to nonignorability. The local sensitivity analysis is derivative-based, varied one at a time by a small amount around a fixed point, and calculates the effect of individual perturbations on the output. Usually the fixed point is at ignorability. If the results of analysis of sensitivity to ignorability has little influence on the parameter estimates of the inferences of interest, one can directly report the results obtained by L_I under the assumption of ignorablity, instead of using L_C, which might have unidentifiability and untestable issues.

(22) proposed local sensitivity in the neighborhood of the MAR model. In their approach, the complete data Y and a latent variable Z each follows a normal linear model and their error terms are correlated with correlation coefficient ρ. The relationship of Y and Z is that Y is observed if $Z > 0$. Then they introduced a nonignorability parameter θ which is a convenient reparameterization of ρ in this way

$$\theta = \frac{\rho}{(1-\rho^2)^{1/2}}$$

Based on the derived likelihood for nonignorable selection model, the statistic T of the interested inferences is derived

$$T(\theta) = T(0) + A\eta$$

Here A is the sensitivity multiplier calculated from data; $\eta(\theta)$ is the single log-odds parameter related to θ. Then they examined sensitivity of various inferences for small changes of θ around the ignorable model ($\theta = 0$).

(23) extended the approach to surround the MAR model in the neighborhood of Kullback-Leibler divergence (N). For some small ϵ_0, the neighborhood is defined as

$$N = \{g_{YZ} : D(g_{YZ}, f_{YZ}) < \frac{1}{2}\epsilon_0^2\}$$

where g_{YZ} is the actual joint distribution of Y and Z, f_{YZ} is the joint distribution of Y and Z under the assumption of ignorability, and $D(f_1, f_2)$ is the Kullback-Leibler divergence

$$D(f_1, f_2) = \int \log\left(\frac{f_1}{f_2}\right) f_1 d_y d_z = E_{f_1}\left\{\log\left(\frac{f_1}{f_2}\right)\right\}$$

Then they compare the inference for models in the neighborhood N with that for a ignorability model. They proved that the sensitivity multiplier A is the upper bound for all models in the neighborhood N constrained by the single parameter $\eta(\theta)$. Here $\eta(\theta)$ is a sensitivity parameter and $\eta^2(\theta)$ is defined as the variance of the log odds of selection models with respect to the outcome.

(24) proposed sensitivity analysis based on a local influence approach (25) when the subjects drop out nonrandomly in the longitudinal data under a normal distribution. Their proposed method is different from other local sensitivity analysis in the way that they assign a perturbation around the MAR model to the nonignorability parameter for each subject (γ_{1i}) within the linear predictor of the selection model and measure sensitivity on likelihood displacement instead of estimates of parameter of interest, which is defined as:

$$LD(\gamma_{1i}) = 2[\ln L_C(\hat{\beta}(0), \hat{\gamma}_0(0); 0) - \ln L_C(\hat{\beta}(\gamma_{1i}), \hat{\gamma}_0(\gamma_{1i}); \gamma_{1i})]$$

Here $L_C(\hat{\beta}(0), \hat{\gamma}_0(0); 0)$ corresponds to an MAR dropout model ($\gamma_{1i} = (0, \cdots, 0)'$) and $L_C(\hat{\beta}(\gamma_{1i}), \hat{\gamma}_0(\gamma_{1i}); \gamma_{1i})$ corresponds to an NMAR dropout model with small perturbation around MAR. With their method, they can identify subjects with high influence on the nonrandom dropout model.

1.4.3 An index Approach to Local Sensitivity to Nonignorability

(6) extended the approach of (22) and proposed an Index of Sensitivity to Nonignorablity (ISNI) evaluated at the MLE of the parameter estimates under the assumption of ignorability. They derived a general expression for ISNI based on Taylor-series approximation to the nonignorable likelihood with a general parametric complete data model jointly with a general parametric selection model. In their selection model, there is a single nonignorability parameter, named γ_1, which is linking the outcome y_i to the completeness indicator that takes the value 1 for observed and 0 for missing. Unlike η in (22), the nonignorability parameter γ_1

is the regression coefficient in the missing-data model. The ISNI is the first derivative of the parameter estimates (θ) with respect to γ_1 in the following equation:

$$\mathrm{ISNI} = -\left(\nabla^2 L_{\theta,\theta}\right)^{-1} \nabla^2 L_{\theta,\gamma_1},$$

Where θ is a vector of parameters of interest and ∇ is derivative. The ISNI is actually the product of the estimated variance-covariance matrix of θ under MAR and the orthogonality of θ and γ_1. Compared with other sensitivity analyses mentioned above, ISNI analysis provides a closed-form formula for the sensitivity. The main advantage of this approach is that computing ISNI does not need to estimate a nonignorable model. Hence the ISNI could be used as preliminary screening for sensitivity to nonignorability. If the ratio of ISNI to the standard error of a parameter of interest is greater than 1, the model is potentially nonignorable. It is necessary to obtain more information to model the missing-data mechanism and use L_C to draw inferences.

After the (6) work on the ISNI, there has been a tremendous amount of work done in developing ISNI to use in the other types of data. (26) applied the approach of (6) to the problem of measuring sensitivity of the As-treated (AT) to nonignorable crossover in randomized trials and extended their ISNI formula to allow the nonignorable missing-data mechanism to vary across groups of subjects. Instead of one nonignorability parameter (γ_1) in (6), they proposed two nonignorability parameters: γ_{1c} if subject i is randomized to control and γ_{1t} if subject i is randomized to test. The probability of accepting the randomization arm depends on the outcomes in the following way:

$$P(G_i = a|R, Y_{ci}, Y_{ti}) = \begin{cases} h(\gamma_{0c} + \gamma_{1c} y_{ci}), & \text{if } \gamma_i = c \\ h(\gamma_{0t} + \gamma_{1t} y_{ti}), & \text{if } \gamma_i = t \end{cases}$$

Where $G_i = a$ indicates subject i accepts randomized therapy; otherwise subject i switches from randomized therapy. Using this extended ISNI approach, one can evaluate whether the AT estimate of the difference between treatment and control is sensitive to nonignorable crossover.

(27) applied ISNI to survival data with nonignorable censoring. They assumed a scaled beta model for the censoring process to take into account of the dependence of censoring time on survival time in addition to assuming a parametric model for the survival time. Then they assessed the sensitivity of traditional model-based analyses to nonignorability using ISNI.

(28) applied ISNI to measure the sensitivity of MAR estimates to small departures from ignorability for multivariate normal outcomes. (29) extended the work of (28) to apply the ISNI methodology to handle the generalized linear mixed model for longitudinal data subject to nonignorable dropouts and to measure the sensitivity of inferences in the neighborhood of MAR. (28) used a multivariate normal model for the outcomes of interest and a regression model with logit or probit link for missingness probabilities. They allowed missingness probabilities

to depend on the current values of predictors (x_{ij}), the previous outcome observation ($y_{i,j-1}$), and the current outcome (y_{ij}) in the following way:

$$P(G_{ij} = 1|X_{ij}, Y_i, G_{i,j-1} = 1) = h(x_{ij}\gamma_{01} + y_{i,j-1}\gamma_{02} + y_{ij}\gamma_1)$$

Where $G_{ij} = 1$ indicates the subject i completes the observation at j timepoint, $j = 2, \cdots, m_i$ and h is a link function. The computation involves estimating a mixed-effect model and a selection model for the drop-out, together with some simple arithmetic calculations. (29) had the same setup as (28) for the drop-out model. For the model of the outcomes of interest, (29) modeled the distribution of the outcomes in the form of the exponential family. In the derivation the ISNI process, (29) used the score function and the information matrix to derive the terms in the general formula of ISNI (6). Due to the lack of closed forms for calculating the conditional expectations of the score function and the information matrix, they used the method of Gaussian quadrature to approximate these conditional expectations. (28) and (29) assumed monotonic missingness, thus their derived ISNIs do not extend to intermittent missingness data.

The above ISNI applications have been applied to a parametric form for the mean structure in the outcome model. (30) relaxed the parametric assumption in the outcome model and investigated the usage of ISNI methodology to evaluate the potential effects of nonignorable missing data mechanism on the generalized additive models (GAM) estimates in the neighborhood of the MAR model. The GAM allows nonparametric functional forms for all or some of the predictors in outcomes models. This is useful when the relationship of the outcome and a covariate is not well understood. The backfitting algorithm is often used to estimate GAMs in complete data, but it has not been adapted to handle the nonignorable missing data. The alternative Regression Spline (RS) method can be used to handle the nonignorable missing data, as it is a widely used method for performing nonparametric curve estimation. Thus (30) adopted the GAM + RS approach to model the complete outcome in the nonignorable selection model. That is

$$g(\mu_i^Y) = \eta_i \approx X_i^T \beta + \sum_{q=1}^{Q} W_{iq}^T \phi_q$$

where μ_i^Y is the mean of the outcome, $X_i^T\beta$ is a parametric component, $\sum_{q=1}^{Q} W_{iq}^T \phi_q$ is a non-parametric component, $W_{iq}^T = [1, T_q, \cdots, T_q^r, (T_q - \tau_{q1})_+^r, \cdots, (T_q - \tau_{qK_q})_+^r]$, and ϕ_q is the corresponding vector of coefficient parameters. They modeled the missingness mechanism with a logistic regression model. With these specifications, they applied ISNI to evaluate the effect of nonignorable missing on the estimation of GAMs.

The application of ISNI we reviewed above have all used logit or probit link functions to model the missingness mechanism. (31) proposed an extended ISNI with a generalized logistic model for the missingness mechanism. The generalized logistic model relaxes the assumption of a known link function by embedding the logistic model within "two families of power transformations to model symmetric and asymmetric departures from the logistic model in binary regression" (32). The logistic model is a special case in the families. The common link functions included in the families are logit, complementary log-log (cloglog), linear, and approximated probit and arcsine. (31) proposed the extended ISNI: multiplication of ISNI by an F factor.

The F factor is "the average over the dropouts of their individual nonignorability factors" F_i

For a family of symmetric link functions

$$F_i = \begin{cases} 1 - \frac{1}{4}\lambda^2 \hat{\eta}_{0i}^2, & \text{if } |\frac{1}{2}\lambda\hat{\eta}_{0i}| < 1 \\ 0, & otherwise \end{cases}$$

For a family of asymmetric link functions

$$F_i = \begin{cases} h(\hat{\eta}_{0i})(1 + \lambda e^{\hat{\eta}_{0i}})/e^{\hat{\eta}_{0i}}, & \text{if } \lambda e^{\hat{\eta}_{0i}} > -1 \\ 0, & otherwise \end{cases}$$

Where λ is a scalar. When $\lambda = 1$ it indicates a logistic model for both symmetric and asymmetric families. When $\lambda = 0$, it indicates a linear model for symmetric family and a cloglog model for asymmetric family. $\hat{\eta}_{0i} = \hat{\gamma}_{0i}^T S_i$, S is the vector of fully observed dropout predictors. "The extended ISNI analysis allows the degree of nonignorability to vary among subjects" by allowing "individual nonignorability parameters to vary only subject to the one-parameter extension of the logistic model"(31). (31) conducted sensitivity analysis over a plausible range of λ, which should include "commonly used link functions as well as the MLE of λ assuming MAR"(31), to test dependence of the link function to nonignorability.

(33) further investigated relaxing a missing data model from parametric functional forms to a semi-parametric approach: a generalized additive missing data model in addition to relaxing

the linear relationship with non-linear relationship of missingness predictors and the missingness indicator (30).

The above reviewed papers for ISNI methods assumed monotone pattern of missingness. But non-monotone pattern of missigness (intermittent missingness) often plagues panel data. (34) and (35) extended the ISNI method to nonignorable non-monotone missingness in panel data or longitudinal clinical trials for various types of outcome models, such as the marginal multivariate Gaussian model, generalized linear mixed model, and panel tobit model. Now the missingness indicator G_{ij} at time j for subject i in the missing-data model denoted as

$$G_{ij} = \begin{cases} 0, & \text{if subject i is observed at time j} \\ 1, & \text{if subject i is intermittent missing at visit j} \\ 2, & \text{if subject i drops out at visit j} \end{cases}$$

They employed a multinomial logit model for non-monotone missingness indicator G_{ij} given the missingness status at the previous visits, which depended on the covariates, the prior observed outcomes and the potentially unobserved current outcome using a Markov model of order q:

$$\begin{aligned} P_{ij}^{uv_1\cdots v_q} &= \frac{\phi_{ij}^{uv_1\cdots v_q}}{\sum\limits_{U=0}^{2} \phi_{ij}^{Uv_1\cdots v_q}} \\ \phi_{ij}^{uv_1\cdots v_q} &= exp(\gamma_0^{uv_1\cdots v_q} s_{ij} + \gamma_1^{uv_1\cdots v_q} y_{ij}) \end{aligned}$$

Where u is an index of the status of missingness $0, 1, 2$, $\nu_1 \cdots \nu_q$ are index of status happen in history of missingness through $G_{i,j-1}, \cdots, G_{i,j-q}$ with 0 as observed and 1 as unobserved, and S_{ij} is a vector of fully observed predictors for missinginess including the previous and current values of predictors and the previous observed outcome.

Due to the model's constraints: 1) $G_{ij} = 2$ implied $G_{it} = 2$ for $t > j$; 2) $\phi_{ij}^{2g_{i(j)}} = 0$ when $G_{ij-1} = 1$; 3) $\phi_{ij}^{0g_{i(j)}} = 1$ for $g_{ij-1} = 0$ or 1 as the response probabilities must sum to 1, the nonignorable parameter γ_1 expanded to a vector of unconstrained parameters $(\gamma_{10}, \gamma_{20}, \gamma_{11})$. The first digit represented current missing status (1=intermittent missing, 2=dropout) and the second digit represented previous missing status (0=observed, 1=unobserved). Three elements of γ_1 perturbed from 0 to 1 formed a hypercube in space. Under this circumstance, (34) and (35) considered the maximum sensitivity in this hypercube for the ISNI. The computation of the ISNI for nonignorable non-monotone missingness did not have closed form and needed numerical evaluations with respecting to the missing data.

In this thesis, our approach builds on earlier work developing ISNI. We extend the local sensitivity analysis of likelihood-based estimates with missing data to allow for nonlinear impact of nonignorable selection mechanism. We develop formule and algorithms for computing the nonlinear impact of nonignorability. The resulting nonlinear local sensitivity index can provide substantial improvement in some important situations, such as when estimating the error variances or when both the covariates and outcome are subject to missingness. This approach extends the linear local sensitivity index (6). At the same time it maintains computational feasibility as it avoids fitting any complicated nonignorable selection models. Our approach can be used as a screening method to tell whether we need to collect additional data and perform arduous work of nonignorable modeling.

The rest of the thesis is organized as follows: In Chapter 2 we develop general nonlinear sensitivity index for selection when only the outcome is subject to selective observation. Then we illustrate it in specific situations: univariate normal data and the simple linear regression model. Last we conduct simulation studies for these two special situations and compared the selection bias between our proposed nonlinear sensitivity index and linear sensitivity index (6). In Chapter 3, with the same process as in Chapter 2, we develop a general nonlinear sensitivity index for nonignorable selection when both the outcome and covariates are subject to selective observation and illustrated it in specific situations: simple linear regression with both the continuous outcome and covariates and simple linear regression with the continuous outcome and the binary covariates. Then we conclude this chapter with simulation studies for these two special situations. Chapter 4 applies our methodology to two real examples mentioned in the beginning of Chapter 1 (Introduction). We conclude with a discussion in Chapter 5.

CHAPTER 2

NONLINEAR SENSITIVITY INDEX FOR MISSINGNESS IN THE OUTCOME ONLY

2.1 Selection Model and Linear Local Sensitivity Analysis

We first describe the joint selection model that consists of the following two components: First is a outcome model. Let the outcome Y have a density function $f_\theta(y|x)$ for independent subjects $i = 1, \cdots, N$ from the population of interest, where X is a vector of fully-observed predictors of Y. For illustrative purposes, here we consider models for cross-sectional ideal outcomes, though our methods are general and can be extended to more complex models. The outcome Y_i is independently drawn from the exponential family that includes normal, binomial, Poisson and Gamma distributions as special cases.

Second is missing data model. As indicated in the Introduction, selection models are commonly used to capture the nonignorable missing data. Let the missing indicator G have a density function of $h_\gamma(s, y)$ with the value 1 for subjects who are observed and 0 for subjects who are missing or unobserved, where s is a set of fully-observed predictors for the missing-data mechanism model. We model the probability of Y being observed as follows:

$$\mathrm{Prob}(G_i = 1|y_i, s_i) = h(\gamma_0^T s_i + \gamma_1 y_i)$$

Where $h(\cdot)$ is the specified monotonic link function, e.g., the logit or probit.

We assume Y_i is independent of S_i conditioning on X_i, which means S_i is not used for predicting Y_i, but only used to predict the probability of missingness. Let θ be a p x 1 vector of unknown parameters with parameter space ω_θ being an open subset of R_p, and let $\gamma = (\gamma_0, \gamma_1)$ be a q x 1 vector of pre-specified nonignorability parameters with parameter space ω_γ being an open subset of R_q.

The log likelihood of the joint model is

$$L(\theta, \gamma_0, \gamma_1) = \sum_{i:g_i=1} \ln f_\theta(y_i^{obs}|x_i) + \sum_{i:g_i=1} \ln h(\gamma_0 s_i + \gamma_1 y_i^{obs}) + \sum_{i:g_i=0} \ln \int f_\theta(y_i^{mis}|x_i)[1 - h(\gamma_0 s_i + \gamma_1 y_i^{mis})] dy_i^{mis} \tag{2.1}$$

Observed data provide little information for distinguishing between different missingness mechanisms, and the model is usually unidentifiable; even when the model is identified, the inference obtained from the model is highly sensitive to the strong model assumptions that are untestable with the observed data in (36), (8), (18), and (19). Thus an alternative and more prudent approach is to perform a sensitivity analysis on the above nonignorable model over a plausible range of values for γ_1 in a neighborhood of the MAR model from (22), (23), (24), and (6). When γ_1 is 0, the missing-data mechanism is MAR, and the model is the ignorable model. For the sensitivity analysis, we must first obtain MLE$(\hat{\theta}(\gamma_1), \hat{\gamma_0}(\gamma_1))$ by maximizing L over (θ, γ_0) for any value of γ_1, and then compare inferences for the range of γ_1 values. Usually we are interested in estimating θ. We will evaluate how $\hat{\theta}(\gamma_1)$ departs from $\hat{\theta}(0)$ when γ_1 changes from zero. By a Taylor-series approximation, we have

$$\hat{\theta}(\gamma_1) \approx \hat{\theta}(0) + \left.\frac{\partial\hat{\theta}(\gamma_1)}{\partial\gamma_1}\right|_{\gamma_1=0} \times \gamma_1$$

The MAR estimate $\hat{\theta}(0)$ can be obtained from standard software. $\left.\frac{\partial\hat{\theta}(\gamma_1)}{\partial\gamma_1^T}\right|_{\gamma_1=0}$ measures the changing rate of $\hat{\theta}(\gamma_1)$ as a function of γ_1, which is referred to as the Index of Local Sensitivity to NonIgnorability (ISNI).

(6) derived the linear Index of Sensitivity to NonIgnorability (ISNI), which is the first derivative of $\hat{\theta}$ with respective to γ_1, evaluated at the MAR model, i.e. $\gamma_1 = 0$ as follows

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

The ISNI derived by (6) is the combination of the estimated variance-covariance matrix of $\hat{\theta}$ in an ignorable model and a measure of the orthogonality of θ and γ_1. The computation of this index is easy and intuitive; it does not require evaluation of complicated integrals or fitting any joint selection models. Thus, it permits simple and fast evaluation of $\theta(\widehat{\gamma_1})$. ISNI formulas have been developed for a range of problems with missing outcomes in (26), (28), (29), (37), (31), (30), (33), (35), (34), (27), (38), and (39).

2.2 Nonlinear Sensitivity Index Development

Although the linear sensitivity index measures are adequate for a range of important applications previously studied, there is a need to develop more accurate simple index measures that can more faithfully capture the sensitivity in general situations. For example, when estimating the error variances, or when both the covariates and outcome are subject to missingness, the selection bias can take a highly nonlinear form such that we must develop nonlinear sensitivity index measures. More generally, standard deviations can be important for quality control, measuring risk in finance, and consequently is an important estimate. The development of ISNI so far assumes that only the outcome Y is subject to missingness while the covariates X are fully observed. However, a common problem in EMA studies is the simultaneous missingness in both the outcome variable and important covariates that would be collected concurrently during those nonresponded prompts. The current ISNI formulas are not computable in this situation because they require known covariate values. Furthermore, the impact of nonignorability takes a highly nonlinear shape around the MAR model, and consequently, the current linear ISNI loses its effectiveness. Therefore, there is a need to generalize ISNI to make it applicable and effective to a much broader range of studies. Our main strategy is to employ a second-order Taylor-series expansion of the MLEs of model parameters as a function of γ_1 as follows

$$\hat{\theta}(\gamma_1) \approx \hat{\theta}(0) + \left.\frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_1^T}\right|_{\gamma_1=0} \times \gamma_1 + \left.\frac{\partial^2 \hat{\theta}(\gamma_1)}{\partial \gamma_1^2}\right|_{\gamma_1=0} \times \frac{\gamma_1^2}{2}$$

The following steps show the detailed derivation of Index of Sensitivity to NonIgnorability in Quadratic form (ISNIQ) Given a fixed γ_1, MLE estimates $\hat{\theta}(\gamma_1)$ and $\hat{\gamma_0}(\gamma_1)$ satisfy the following condition

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

Taking the first derivative of both sides with respect to γ_1 we have

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

Therefore for any value of γ_1, the first derivative of the $(\hat{\theta}(\gamma_1), \hat{\gamma_0}(\gamma_1))$ with respect to γ_1 is as follows

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

We call the first derivative of the $(\hat{\theta}(\gamma_1), \hat{\gamma_0}(\gamma_1))$ with respect to γ_1 as Index of Sensitivity to NonIgnorability in Linear form (ISNIL).

Next, we take the second derivative of the $(\hat{\theta}(\gamma_1), \hat{\gamma_0}(\gamma_1))$ with respect to γ_1. We then have

$$\begin{aligned}\frac{\partial}{\partial\gamma_1}\begin{pmatrix}\frac{\partial\hat{\theta}(\gamma_1)}{\partial\gamma_1}\\ \\ \frac{\partial\hat{\gamma_0}(\gamma_1)}{\partial\gamma_1}\end{pmatrix} =& A^{-1}\frac{\partial}{\partial\gamma_1}\left[\frac{\partial^2 L(\hat{\theta}(\gamma_1), \hat{\gamma_0}(\gamma_1), \gamma_1)}{\partial(\theta^T, \gamma_0^T)^T \partial(\theta^T, \gamma_0^T)}\right] A^{-1} \frac{\partial^2 L(\hat{\theta}(\gamma_1), \hat{\gamma_0}(\gamma_1), \gamma_1)}{\partial(\theta^T, \gamma_0^T)^T \partial\gamma_1} \\ & - A^{-1}\frac{\partial}{\partial\gamma_1}\left[\frac{\partial^2 L(\hat{\theta}(\gamma_1), \hat{\gamma_0}(\gamma_1), \gamma_1)}{\partial(\theta^T, \gamma_0^T)^T \partial\gamma_1}\right]\end{aligned}$$

Where $A^{-1} = \left[\frac{\partial^2 L(\hat{\theta}(\gamma_1),\hat{\gamma_0}(\gamma_1),\gamma_1)}{\partial(\theta^T,\gamma_0^T)^T\partial(\theta^T,\gamma_0^T)}\right]^{-1}$

$$\frac{\partial}{\partial\gamma_1}\left[\frac{\partial^2 L(\hat{\theta}(\gamma_1),\hat{\gamma_0}(\gamma_1),\gamma_1)}{\partial(\theta^T,\gamma_0^T)^T\partial(\theta^T,\gamma_0^T)}\right] = \frac{\partial^3 L(\hat{\theta}(\gamma_1),\hat{\gamma_0}(\gamma_1),\gamma_1)}{\partial(\theta^T,\gamma_0^T)^T\partial(\theta^T,\gamma_0^T)\partial\gamma_1} + \sum_{j=1}^{n_\theta}\frac{\partial^3 L(\hat{\theta}(\gamma_1),\hat{\gamma_0}(\gamma_1),\gamma_1)}{\partial(\theta^T,\gamma_0^T)^T\partial(\theta^T,\gamma_0^T)\partial\theta_j}\frac{\partial\hat{\theta_j}(\gamma_1)}{\partial\gamma_1}$$
$$+\sum_{k=1}^{n_{\gamma_0}}\frac{\partial^3 L(\hat{\theta}(\gamma_1),\hat{\gamma_0}(\gamma_1),\gamma_1)}{\partial(\theta^T,\gamma_0^T)^T\partial(\theta^T,\gamma_0^T)\partial\gamma_{0k}}\frac{\partial\hat{\gamma_{0k}}(\gamma_1)}{\partial\gamma_1}$$

And

$$\frac{\partial}{\partial\gamma_1}\left[\frac{\partial^2 L(\hat{\theta}(\gamma_1),\hat{\gamma_0}(\gamma_1),\gamma_1)}{\partial(\theta^T,\gamma_0^T)^T\partial\gamma_1}\right] = \frac{\partial^3 L(\hat{\theta}(\gamma_1),\hat{\gamma_0}(\gamma_1),\gamma_1)}{\partial(\theta^T,\gamma_0^T)^T\partial(\theta^T,\gamma_0^T)\partial\gamma_1}\frac{\partial(\hat{\theta}^T,\hat{\gamma_0}^T)^T}{\partial\gamma_1} + \frac{\partial^3 L(\hat{\theta}(\gamma_1),\hat{\gamma_0}(\gamma_1),\gamma_1)}{\partial(\theta^T,\gamma_0^T)^T\partial\gamma_1^2}$$

We call the second derivative of the $(\hat{\theta}(\gamma_1),\hat{\gamma_0}(\gamma_1))$ with respect to γ_1 an Index of Sensitivity to NonIgnorability in Quadratic form (ISNIQ).

Then we rewrite the above equation as follows

$$\mathrm{ISNIQ} = \begin{pmatrix} \nabla^2 L_{\theta,\theta} & \nabla^2 L_{\theta,\gamma_0} \\ \nabla^2 L_{\gamma_0,\theta} & \nabla^2 L_{\gamma_0,\gamma_0} \end{pmatrix}^{-1}$$

$$\left[\begin{pmatrix} \nabla^3 L_{\theta,\theta,\gamma_1} & \nabla^3 L_{\theta,\gamma_0,\gamma_1} \\ \nabla^3 L_{\gamma_0,\theta,\gamma_1} & \nabla^3 L_{\gamma_0,\gamma_0,\gamma_1} \end{pmatrix} + \right.$$

$$\sum_{j=1}^{n_\theta} \begin{pmatrix} \nabla^3 L_{\theta,\theta,\theta_j} & \nabla^3 L_{\theta,\gamma_0,\theta_j} \\ \nabla^3 L_{\gamma_0,\theta,\theta_j} & \nabla^3 L_{\gamma_0,\gamma_0,\theta_j} \end{pmatrix} \mathrm{ISNI}(\hat{\theta}_j) +$$

$$\left. \sum_{k=1}^{n_{\gamma_0}} \begin{pmatrix} \nabla^3 L_{\theta,\theta,\gamma_{0k}} & \nabla^3 L_{\theta,\gamma_0,\gamma_{0k}} \\ \nabla^3 L_{\gamma_0,\theta,\gamma_{0k}} & \nabla^3 L_{\gamma_0,\gamma_0,\gamma_{0k}} \end{pmatrix} \mathrm{ISNI}(\hat{\gamma}_{0k}) \right] \begin{pmatrix} \nabla^2 L_{\theta,\theta} & \nabla^2 L_{\theta,\gamma_0} \\ \nabla^2 L_{\gamma_0,\theta} & \nabla^2 L_{\gamma_0,\gamma_0} \end{pmatrix}^{-1} \begin{pmatrix} \nabla^2 L_{\theta,\gamma_1} \\ \nabla^2 L_{\gamma_0,\gamma_1} \end{pmatrix}$$

$$- \begin{pmatrix} \nabla^2 L_{\theta,\theta} & \nabla^2 L_{\theta,\gamma_0} \\ \nabla^2 L_{\gamma_0,\theta} & \nabla^2 L_{\gamma_0,\gamma_0} \end{pmatrix}^{-1}$$

$$\left[\begin{pmatrix} \nabla^3 L_{\theta,\gamma_1,\gamma_1} \\ \nabla^3 L_{\gamma_0,\gamma_1,\gamma_1} \end{pmatrix} + \begin{pmatrix} \nabla^3 L_{\theta,\theta,\gamma_1} & \nabla^3 L_{\theta,\gamma_0,\gamma_1} \\ \nabla^3 L_{\gamma_0,\theta,\gamma_1} & \nabla^3 L_{\gamma_0,\gamma_0,\gamma_1} \end{pmatrix} \right] \mathrm{ISNI} \begin{pmatrix} \hat{\theta} \\ \hat{\gamma_0} \end{pmatrix}$$

$$= - \begin{pmatrix} \nabla^2 L_{\theta,\theta} & \nabla^2 L_{\theta,\gamma_0} \\ \nabla^2 L_{\gamma_0,\theta} & \nabla^2 L_{\gamma_0,\gamma_0} \end{pmatrix}^{-1} \left[2 \begin{pmatrix} \nabla^3 L_{\theta,\theta,\gamma_1} & \nabla^3 L_{\theta,\gamma_0,\gamma_1} \\ \nabla^3 L_{\gamma_0,\theta,\gamma_1} & \nabla^3 L_{\gamma_0,\gamma_0,\gamma_1} \end{pmatrix} + \right.$$

$$\sum_{j=1}^{n_\theta} \begin{pmatrix} \nabla^3 L_{\theta,\theta,\theta_j} & \nabla^3 L_{\theta,\gamma_0,\theta_j} \\ \nabla^3 L_{\gamma_0,\theta,\theta_j} & \nabla^3 L_{\gamma_0,\gamma_0,\theta_j} \end{pmatrix} \mathrm{ISNI}(\hat{\theta}_j) +$$

$$\left. \sum_{k=1}^{n_{\gamma_0}} \begin{pmatrix} \nabla^3 L_{\theta,\theta,\gamma_{0k}} & \nabla^3 L_{\theta,\gamma_0,\gamma_{0k}} \\ \nabla^3 L_{\gamma_0,\theta,\gamma_{0k}} & \nabla^3 L_{\gamma_0,\gamma_0,\gamma_{0k}} \end{pmatrix} \mathrm{ISNI}(\hat{\gamma}_{0k}) \right] \mathrm{ISNI} \begin{pmatrix} \hat{\theta} \\ \hat{\gamma_0} \end{pmatrix}$$

$$- \begin{pmatrix} \nabla^2 L_{\theta,\theta} & \nabla^2 L_{\theta,\gamma_0} \\ \nabla^2 L_{\gamma_0,\theta} & \nabla^2 L_{\gamma_0,\gamma_0} \end{pmatrix}^{-1} \begin{pmatrix} \nabla^3 L_{\theta,\gamma_1,\gamma_1} \\ \nabla^3 L_{\gamma_0,\gamma_1,\gamma_1} \end{pmatrix}$$

When only the sensitivity of $\hat{\theta}$ is of interest and noting that $\nabla^2 L_{\theta,\gamma_0} = 0$, $\nabla^3 L_{\theta,\gamma_0,\theta_j} = 0$, $\nabla^3 L_{\gamma_0,\gamma_0,\theta_j} = 0$, $\nabla^3 L_{\theta,\theta,\gamma_{0k}} = 0$ and $\nabla^3 L_{\theta,\gamma_0,\gamma_{0k}} = 0$, we have

$$\begin{aligned}\mathrm{ISNIQ}(\hat{\theta}) = \frac{\partial^2 \theta(\hat{\gamma}_1)}{\partial \gamma_1^2} = &- \nabla^2 L_{\theta,\theta}^{-1} \left(2\nabla^3 L_{\theta,\theta,\gamma_1} + \sum_{j=1}^{n_\theta} \nabla^3 L_{\theta,\theta,\theta_j} \mathrm{ISNI}(\hat{\theta}_j) \right) \mathrm{ISNI}(\hat{\theta}) \\ &- 2\nabla^2 L_{\theta,\theta}^{-1} \nabla^3 L_{\theta,\gamma_0,\gamma_1} \mathrm{ISNI}(\hat{\gamma_0}) - \nabla^2 L_{\theta,\theta}^{-1} \nabla^3 L_{\theta,\gamma_1,\gamma_1}\end{aligned}$$

Note that the terms in the formula above are all evaluated under the MAR model (i.e., $\gamma_1 = 0$) with no need to fit any nonignorable models.

2.3 Examples

2.3.1 Univariate Normal Data

Now we consider ISNIQ of $\hat{\theta}$ for a special case of univariate normal model. Let $Y_i \sim^{\text{iid}} N(\mu, \sigma^2)$, i = 1, , N with the binary indicator variable G_i (1 indicates Y_i observed; 0 indicates Y_i missed). The selection model is simply as follows

$$\mathrm{Prob}(G_i = 1|y_i) = h(\gamma_0 + \gamma_1 y_i) = \frac{\exp(\gamma_0 + \gamma_1 y_i)}{1 + \exp(\gamma_0 + \gamma_1 y_i)}$$

The log likelihood for the above joint selection model is

$$\begin{aligned} L(\mu, \sigma^2, \gamma_0, \gamma_1) = & \sum_{i=1}^{N} \left\{ g_i \left[-\frac{1}{2} \ln((2\pi\sigma^2) e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}) + \ln \frac{\exp(\gamma_0 + \gamma_1 y_i)}{1 + \exp(\gamma_0 + \gamma_1 y_i)} \right] \right. \\ & \left. + (1 - g_i) \ln \int (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{(y_i^{mis}-\mu)^2}{2\sigma^2}} [\tfrac{1}{1+\exp(\gamma_0+\gamma_1 y_i^{mis})}] dy_i^{mis} \right\} \end{aligned}$$

Under MAR mechanism, the usual observed-data MLEs estimates of (μ, σ^2) are

$$\begin{aligned} \hat{\mu}(0) = & \frac{\sum_{i=1}^{N} g_i y_i}{\sum_{i=1}^{N} g_i} \\ \hat{\sigma}^2(0) = & \frac{\sum_{i=1}^{N} g_i (y_i - \hat{\mu}(0))^2}{\sum_{i=1}^{N} g_i} \end{aligned}$$

By the derived formula of linear ISNI (ISNIL) above, we obtain ISNIL for the outcome model parameters (μ, σ^2) as follows

$$\begin{aligned} \text{ISNIL}(\hat{\mu}) = & -\tfrac{N_m}{N} \sigma^2 \\ \text{ISNIL}(\hat{\sigma}^2) = & \quad 0 \end{aligned}$$

By the derived formula of nonlinear ISNI (ISNIQ) above, we obtain ISNIQ for the outcome model parameters (μ, σ^2) as follows

$$\begin{aligned} \mathrm{ISNIQ}(\hat{\mu}) &= \quad 0 \\ \mathrm{ISNIQ}(\hat{\sigma}^2) &= \quad 2\tfrac{N_o N_m}{N^2}(\sigma^2)^2 \end{aligned}$$

Where N is the total number of complete subjects, N_o is the number of observed subjects, and N_m is the number of missing subjects.

The nonlinear ISNI of the mean parameter estimate $(\hat{\mu})$ is zero. It indicates the linear approximation (ISNIL) is sufficient for the mean parameter estimate. But the nonlinear ISNI of the variance parameter $(\hat{\sigma}^2)$ is a non-zero value. Obviously, the linear approximation is inadequate for the variance estimate. It indicates that the selection bias for $\hat{\sigma}^2$ is highly nonlinear with quadratic form. Consequently, the sensitivity of the variance estimate is affected by the fraction of data missing. When the fraction of data missing is 0.5, the sensitivity is maximum. This is likely due to the fact that when the missing data proportion is larger than 0.5, a large part of impact is from ignorable missingness instead of nonignorability, and consequently, the sensitivity to nonignorable missingness is reduced.

The following steps show the detailed derivation:

$$
\begin{aligned}
\nabla^2 L^{-1}_{\theta,\theta} &= \begin{pmatrix} \frac{\partial^2 L}{\partial\mu\partial\mu} & \frac{\partial^2 L}{\partial\mu\partial\sigma^2} \\ \\ \frac{\partial^2 L}{\partial\sigma^2\partial\mu} & \frac{\partial^2 L}{\partial\sigma^2\partial\sigma^2} \end{pmatrix}^{-1} \\
&= \begin{pmatrix} -\left(\frac{\hat{\sigma}^2}{N_o}\right) & 0 \\ \\ 0 & -\frac{2(\hat{\sigma}^2)^2}{N_o} \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\nabla^3 L_{\theta,\theta,\gamma_1} &= -\sum_{i=1}^{N} \left\{ (1-g_i)h_i(.) \begin{pmatrix} \frac{\partial^2}{\partial\mu\partial\mu} E(y_i^{mis}) & \frac{\partial^2}{\partial\mu\partial\sigma^2} E(y_i^{mis}) \\ \frac{\partial^2}{\partial\sigma^2\partial\mu} E(y_i^{mis}) & \frac{\partial^2}{\partial\sigma^2\partial\sigma^2} E(y_i^{mis}) \end{pmatrix} \right\} \\
&= \begin{pmatrix} 0 & 0 \\ \\ 0 & 0 \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\nabla^3 L_{\theta,\gamma_0,\gamma_1} &= -\sum_{i=1}^{N} \left\{ (1-g_i)h_i(.)[1-h_i(.)]s_i \begin{pmatrix} \frac{\partial}{\partial\mu} E(y_i^{mis}) \\ \\ \frac{\partial}{\partial\sigma^2} E(y_i^{mis}) \end{pmatrix} \right\} \\
&= \begin{pmatrix} -\sum_{i=1}^{N} \{(1-g_i)h_i(.)[1-h_i(.)]\} \\ \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} -\frac{N_o N_m^2}{N^2} \\ \\ 0 \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\nabla^3 L_{\theta,\gamma_1,\gamma_1} &= -\sum_{i=1}^{N}\left\{(1-g_i)\left[h_i(.)[1-2h_i(.)]\frac{\partial}{\partial\theta}Var(y_i^{mis}|x_i) + h_i(.)[1-h_i(.)]\frac{\partial}{\partial\theta}E^2(y_i^{mis}|x_i)\right]\right\} \\
&= \begin{pmatrix} -\sum_{i=1}^{N}(1-g_i)2h_i(.)[1-h_i(.)]\mu \\ -\sum_{i=1}^{N}(1-g_i)h_i(.)[1-2h_i(.)] \end{pmatrix} \\
&= \begin{pmatrix} -2\hat{\mu}\frac{N_o N_m^2}{N^2} \\ \\ \\ \frac{N_m N_o(N_o-N_m)}{N^2} \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\nabla^3 L_{\theta,\theta,\mu} &= \begin{pmatrix} \frac{\partial^3 L}{\partial\mu\partial\mu\partial\mu} & \frac{\partial^3 L}{\partial\mu\partial\sigma^2\partial\mu} \\ \\ \\ \frac{\partial^3 L}{\partial\sigma^2\partial\mu\partial\mu} & \frac{\partial^3 L}{\partial\sigma^2\partial\sigma^2\partial\mu} \end{pmatrix} \\
&= \begin{pmatrix} 0 & \frac{N_o}{(\hat{\sigma}^2)^2} \\ \\ \\ \frac{N_o}{(\hat{\sigma}^2)^2} & 0 \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\nabla^3 L_{\theta,\theta,\sigma^2} &= \begin{pmatrix} \frac{\partial^3 L}{\partial\mu\partial\mu\partial\sigma^2} & \frac{\partial^3 L}{\partial\mu\partial\sigma^2\partial\sigma^2} \\ \\ \\ \frac{\partial^3 L}{\partial\sigma^2\partial\mu\partial\sigma^2} & \frac{\partial^3 L}{\partial\sigma^2\partial\sigma^2\partial\sigma^2} \end{pmatrix} \\
&= \begin{pmatrix} \frac{N_o}{(\hat{\sigma}^2)^2} & 0 \\ \\ \\ 0 & 0 \end{pmatrix}
\end{aligned}
$$

$$
ISNI(\hat{\mu}) = -\frac{N_m}{N}\hat{\sigma}^2
$$

2.3.2 Simple Linear Regression

Now we consider ISNIQ of $\hat{\theta}$ for a special case of linear regression model with normal errors, we have $Y = \beta_0 + \beta_1 X + \epsilon$, where $\epsilon \sim N(0, \sigma^2)$ we have $\beta = (\beta_0, \beta_1)$ and $\theta = (\beta, \sigma^2)$. The other setups are the same as univariate normal model.

The log likelihood for the above joint selection model is

$$\begin{aligned} L(\beta_0, \beta_1, \sigma^2, \gamma_0, \gamma_1) = & \sum_{i=1}^{N} \Bigg\{ g_i \left[-\frac{1}{2} \ln(2\pi\sigma^2 \exp\frac{-(y_i - x_i'\beta)^2}{2\sigma^2}) + \ln\frac{\exp(\gamma_0 + \gamma_1 y_i)}{1 + \exp(\gamma_0 + \gamma_1 y_i)} \right] \\ & + (1 - g_i) \ln \int (2\pi\sigma^2)^{-\frac{1}{2}} \exp(-\tfrac{(y_i^{mis} - x_i'\beta)^2}{2\sigma^2}) \tfrac{1}{1+\exp(\gamma_0 + \gamma_1 y_i^{mis})} dy_i^{mis} \Bigg\} \end{aligned}$$

By the derived formula of linear ISNI (ISNIL) above, we obtain ISNIL for the outcome model parameters $(\beta_0, \beta_1, \sigma^2)$ as follows

$$\begin{aligned} \mathrm{ISNIL}(\hat{\beta_0}) = & -\left[\tfrac{N_m}{N} + \tfrac{N_m}{N} \tfrac{\sum_{i=1}^{N_o} x_{oi}}{S_{x_o x_o}} (\bar{x}_o - \bar{x}_m) \right] \sigma^2 \\ \mathrm{ISNIL}(\hat{\beta_1}) = & \left[\tfrac{\frac{N_m}{N}}{\frac{S_{x_o x_o}}{N_o}} (\bar{x}_o - \bar{x}_m) \right] \sigma^2 \\ \mathrm{ISNIL}(\hat{\sigma}^2) = & \quad 0 \end{aligned}$$

By the derived formula of nonlinear ISNI (ISNIQ) above, we obtain ISNIQ for the outcome model parameters $(\beta_0, \beta_1, \sigma^2)$ as follows

$$\mathrm{ISNIQ}(\hat{\beta_0}) = \frac{2\hat{\beta}_1\hat{\sigma}^2 N_m^2(\bar{x}_o - \bar{x}_m)}{N^2} - \mathrm{ISNIQ}(\hat{\beta}_1)\bar{x}_o$$

$$\mathrm{ISNIQ}(\hat{\beta_1}) = \frac{-2\hat{\beta}_1\hat{\sigma}^2 N_o N_m(\sum x_{mi}^2 - N_m(\bar{x}\bar{x}_m - \bar{x}\bar{x}_o + \bar{x}_o\bar{x}_m))}{N^2 s_{x_o x_o}}$$

$$\mathrm{ISNIQ}(\hat{\sigma}^2) = 2\frac{N_o N_m}{N^2}(\sigma^2)^2\left[1 + \frac{N_m(\bar{x}_m - \bar{x}_o)^2}{s_{x_o x_o}}\right]$$

Where $s_{x_o x_o} = \sum_{i=1}^{N_o}(x_i - \bar{x}_o)^2$; x_o and x_m are the vectors of predictors for subjects with observed y and missing y, respectively; N_o and N_m are the numbers of observed subjects and missing subjects, respectively.

The following steps show the detailed derivation:

$$
\begin{aligned}
\nabla^2 L^{-1}_{\theta_y,\theta_y} &= -\sigma^2 \begin{bmatrix} \frac{1}{n_0} + \frac{\bar{x}_0^2}{s_{xoxo}} & -\frac{\bar{x}_o}{s_{xoxo}} & 0 \\ -\frac{\bar{x}_o}{s_{xoxo}} & \frac{1}{s_{xoxo}} & 0 \\ 0 & 0 & \frac{2\sigma^2}{n_0} \end{bmatrix} \\
\nabla^3 L_{\theta_y,\theta_y,\gamma_1} &= 0 \\
\nabla^3 L_{\theta_y,\theta_y,\beta_0} &= \begin{bmatrix} 0 & 0 & \frac{N_o}{(\sigma^2)^2} \\ 0 & 0 & \frac{N_o \bar{x}_0}{(\sigma^2)^2} \\ \frac{N_o}{(\sigma^2)^2} & \frac{N_o \bar{x}_0}{(\sigma^2)^2} & 0 \end{bmatrix} \\
\nabla^3 L_{\theta_y,\theta_y,\beta_1} &= \begin{bmatrix} 0 & 0 & \frac{N_o \bar{x}_0}{(\sigma^2)^2} \\ 0 & 0 & \frac{\sum_i x_{io}^2}{(\sigma^2)^2} \\ \frac{N_o \bar{x}_0}{(\sigma^2)^2} & \frac{\sum_i x_{io}^2}{(\sigma^2)^2} & 0 \end{bmatrix} \\
\nabla^3 L_{\theta_y,\gamma_0,\gamma_1} &= \begin{bmatrix} -\frac{N_o (N_m)^2}{N^2} \\ -\frac{N_o (N_m)^2}{N^2} \bar{x}_o \\ 0 \end{bmatrix} \\
\mathrm{ISNI}(\hat{\gamma}_0) &= -\beta_0 - \beta_1 \bar{x}_o \\
\nabla^3 L_{\theta_y,\gamma_1,\gamma_1} &= \begin{bmatrix} -\frac{2N_o (N_m)^2}{N^2} (\beta_0 + \beta_1 \bar{x}_o) \\ -\frac{2N_o (N_m)^2}{N^2} (\beta_0 \bar{x}_0 + \beta_1 \bar{x}_0^2) - \frac{2N_o N_m (N_m - N_o)}{N^2} \sigma_x^2 \beta_1 \\ \frac{N_o N_m (N_o - N_m)}{N^2} \end{bmatrix}
\end{aligned}
$$

2.4 Simulation Studies

2.4.1 Univariate Normal Data

To further illustrate the nonlinear sensitivity index method and demonstrate the superiority of nonlinear ISNIs, we perform the simulation study for univariate normal data. The ideal data Y_i is simulated from standard normal distribution, where $\mu = 0$, $\sigma = 1$ and $i = 1, \cdots, 100$. The response behavior follows a logistic regression model: $\mathtt{logit}(p_i) = \gamma_0 + \gamma_1 y_i$, where p is probability of y observed, $\gamma_0 = 0, 1, 2$, or 3, and $\gamma_1 = -1, -0.5, 0.5$, or 1. The values of γ_0 and γ_1 are varied so that the resulting datasets have varying amount of missingness ranging from ~5% to ~50% and varying degrees of nonignorable missingness varying from -1 to 1. We compute the ISNI-based approximate sensitivity curves as

$$
\begin{aligned}
\hat{\mu}^{\text{ISNIL}}(\gamma_1) &= \hat{\mu}(0) + \text{ISNIL} * \gamma_1 \\
\hat{\mu}^{\text{ISNIQ}}(\gamma_1) &= \hat{\mu}(0) + \text{ISNIL} * \gamma_1 + \frac{\text{ISNIQ}}{2}\gamma_1^2 \\
\widehat{\sigma^2}^{\text{ISNIL}}(\gamma_1) &= \widehat{\sigma^2}(0) + \text{ISNIL} * \gamma_1 \\
\widehat{\sigma^2}^{\text{ISNIQ}}(\gamma_1) &= \widehat{\sigma^2}(0) + \text{ISNIL} * \gamma_1 + \frac{\text{ISNIQ}}{2}\gamma_1^2
\end{aligned}
$$

The comparison shows that the non-linear ISNI is as good as the linear ISNI for mean estimate. For variance estimate, the use of the higher-order ISNI improves the approximation by capturing the curvature of the sensitivity curve around the MAR model.

We then repeat the simulation 100 times and present the pooled analysis in Table I and Table II. We further vary the parameter settings γ_0 and γ_1. The variation of model parameters

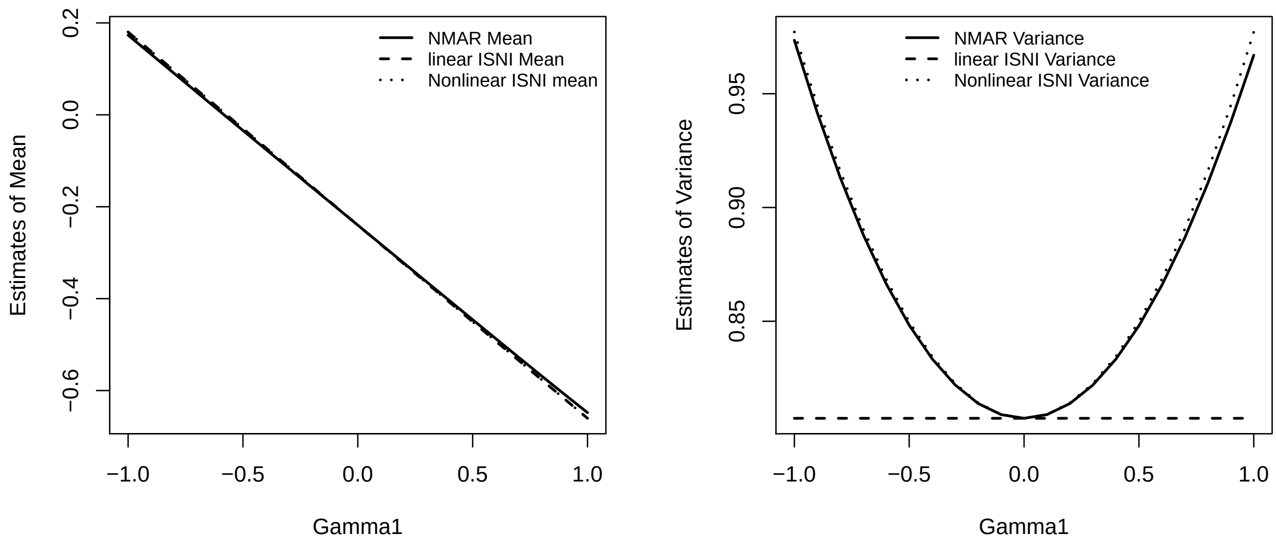

Figure 1: ISNI approximation to mean and variance estimates for univariate normal

give us a range of missing proportions as shown in Table I and Table II. Accuracy of ISNI approximation can be evaluated by

$|\hat{\mu}^{\mathrm{ISNI}}(\gamma_1) - \hat{\mu}(\gamma_1)|$ and $|\widehat{\sigma^2}^{\mathrm{ISNI}}(\gamma_1) - \widehat{\sigma^2}(\gamma_1)|$ The overall performance is measured by the average approximation error over the 100 simulated datasets. The results shows that the nonlinear sensitivity substantially improve the selection bias approximation, as the linear approximation cannot capture the local sensitivity.

Figure 1 shows the analysis based on one simulated dataset with $\gamma_0 = 0$ and $\gamma_1 = -1$ and the proportion of missingness being 0.51. The solid curve shows the exact sensitivity curves of mean: $\hat{\mu}(\gamma_1)$ and variance: $\widehat{\sigma^2}(\gamma_1)$, which were computed by maximizing the log-likelihood of

TABLE I: A SIMULATION STUDY OF THE ACCURACY OF ISNI TO APPROXIMATE MLE OF THE MEAN μ CHANGES LOCALLY FOR UNIVARIATE NORAML DATA

			Average				Approximation Error	
γ_0	γ_1	Prop_M	$\hat{\mu}(0)$	$\hat{\mu}(\gamma_1)$	$\hat{\mu}(\gamma_1)$ (ISNIL)	$\hat{\mu}(\gamma_1)$ (ISNIQ)	$\hat{\mu}(\gamma_1)$ (ISNIL)	$\hat{\mu}(\gamma_1)$ (ISNIQ)
0	-1	49.48%	-0.432	-0.022	-0.012	-0.012	0.01	0.006
0	-0.5	49.84%	-0.26	-0.028	-0.023	-0.023	0.005	0.003
0	0.5	50.16%	0.237	0.003	-0.002	-0.002	0.005	0.005
0	1	50.52%	0.417	0.011	0.002	0.002	0.009	0.011
1	-1	29.95%	-0.272	-0.021	-0.012	-0.012	0.009	0.005
1	-0.5	26.95%	-0.143	-0.014	-0.011	-0.011	0.002	0.002
1	0.5	27.47%	0.115	-0.016	-0.018	-0.018	0.002	0.001
1	1	30.11%	0.25	0.002	-0.007	-0.007	0.009	0.005
2	-1	14.8%	-0.142	-0.011	-0.005	-0.005	0.005	0.003
2	-0.5	12.41%	-0.072	-0.012	-0.011	-0.011	0.001	0.001
2	0.5	12.36%	0.049	-0.011	-0.012	-0.012	0.001	0
2	1	15.25%	0.121	-0.014	-0.019	-0.019	0.005	0.003
3	-1	6.68%	-0.073	-0.011	-0.009	-0.009	0.002	0.001
3	-0.5	5.25%	-0.038	-0.012	-0.012	-0.012	0	0
3	0.5	5%	0.013	-0.012	-0.013	-0.013	0	0
3	1	6.59%	0.052	-0.009	-0.011	-0.011	0.002	0.001

the joint nonignorable selection models for a range of γ_1. This is computationally intensive and is not required in ISNI analyses. The broken line and dotted line represent the approximation using linear ISNI and non-linear ISNI (quadratic term), respectively.

2.4.2 Simple Linear Regression

Following the same process above, we perform the simulation study for simple linear regression. The ideal outcomes are simulated from a simple linear regression model: $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma_y^2)$, where $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_y^2 = 1$ and $i = 1, \cdots, 100$ and the covariate $x_i \sim N(\mu_x, \sigma_x^2)$, where $\mu_x = 0$ and $\sigma_x = 1$. The response behavior follows a logistic regression model: $\mathrm{logit}(p) = \gamma_0 + \gamma_1 y_i$, where p is probability of y observed, $\gamma_0 = 0, 1, 2$, or 3, and

TABLE II: A SIMULATION STUDY OF THE ACCURACY OF ISNI TO APPROXIMATE MLE OF THE VARIANCE σ^2 CHANGES LOCALLY FOR UNIVARIATE NORAML DATA

			Average				Approximation Error	
γ_0	γ_1	Prop_M	$\hat{\sigma}^2(0)$	$\hat{\sigma}^2(\gamma_1)$	$\hat{\sigma}^2(\gamma_1)$ (ISNIL)	$\hat{\sigma}^2(\gamma_1)$ (ISNIQ)	$\hat{\sigma}^2(\gamma_1)$ (ISNIL)	$\hat{\sigma}^2(\gamma_1)$ (ISNIQ)
0	-1	49.48%	0.848	1.007	0.848	1.033	0.159	0.026
0	-0.5	49.84%	0.952	0.989	0.952	1.01	0.037	0.021
0	0.5	50.16%	0.953	0.99	0.953	1.012	0.037	0.022
0	1	50.52%	0.82	0.97	0.82	0.993	0.151	0.023
1	-1	29.95%	0.868	0.997	0.868	1.03	0.129	0.033
1	-0.5	26.95%	0.972	1.004	0.972	1.02	0.032	0.016
1	0.5	27.47%	0.967	0.999	0.967	1.015	0.032	0.016
1	1	30.11%	0.857	0.983	0.857	1.015	0.125	0.032
2	-1	14.8%	0.922	1.005	0.922	1.031	0.083	0.026
2	-0.5	12.41%	0.983	0.997	0.983	1.01	0.014	0.013
2	0.5	12.36%	0.983	0.997	0.983	1.01	0.014	0.013
2	1	15.25%	0.918	1.002	0.918	1.029	0.085	0.027
3	-1	6.68%	0.964	1.005	0.964	1.023	0.041	0.017
3	-0.5	5.25%	0.997	0.999	0.997	1.01	0.004	0.011
3	0.5	5%	0.999	1	0.999	1.011	0.004	0.011
3	1	6.59%	0.957	0.998	0.957	1.015	0.04	0.017

$\gamma_1 = -1, -0.5, 0.5$, or 1. The values of γ_0 and γ_1 are varied so that the resulting datasets have varying amount of missingness ranging from ~5% to ~50% and varying degrees of nonignorable missingness. Figure 2 shows the analysis based on one simulated dataset with $\gamma_0 = 0$ and $\gamma_1 = -1$, with the proportion of missingness being 0.48. The solid curve shows the exact sensitivity curves of $\hat{\beta}_0(\gamma_1)$ and $\hat{\beta}_1(\gamma_1)$, which were computed by maximizing the the log-likelihood of the joint nonignorable selection models for a range of values for γ_1. The broken line and dotted line represent the approximation using linear ISNI and non-linear ISNI (quadratic term), respectively. We compute the ISNI-based approximate sensitivity curves as

$$
\begin{aligned}
\widehat{\beta_0}^{\text{ISNIL}}(\gamma_1) &= \widehat{\beta_0}(0) + \text{ISNIL} * \gamma_1 \\
\widehat{\beta_0}^{\text{ISNIQ}}(\gamma_1) &= \widehat{\beta_0}(0) + \text{ISNIL} * \gamma_1 + \frac{\text{ISNIQ}}{2}\gamma_1^2 \\
\widehat{\beta_1}^{\text{ISNIL}}(\gamma_1) &= \widehat{\beta_1}(0) + \text{ISNIL} * \gamma_1 \\
\widehat{\beta_1}^{\text{ISNIQ}}(\gamma_1) &= \widehat{\beta_1}(0) + \text{ISNIL} * \gamma_1 + \frac{\text{ISNIQ}}{2}\gamma_1^2
\end{aligned}
$$

The comparison shows that the existing linear ISNI computes the tangent of the sensitivity curve at MAR as expected. Furthermore, the use of the higher-order ISNI improves the approximation by capturing the curvature of the sensitivity curve around the MAR model, and the broken curve follows the exact sensitivity curve closer than the linear approximation.

We then repeat the simulation 100 times and present the pooled analysis in Table III and Table IV. The results show that the nonlinear sensitivity substantially improves the selection bias approximation, as the linear approximation cannot capture the local sensitivity.

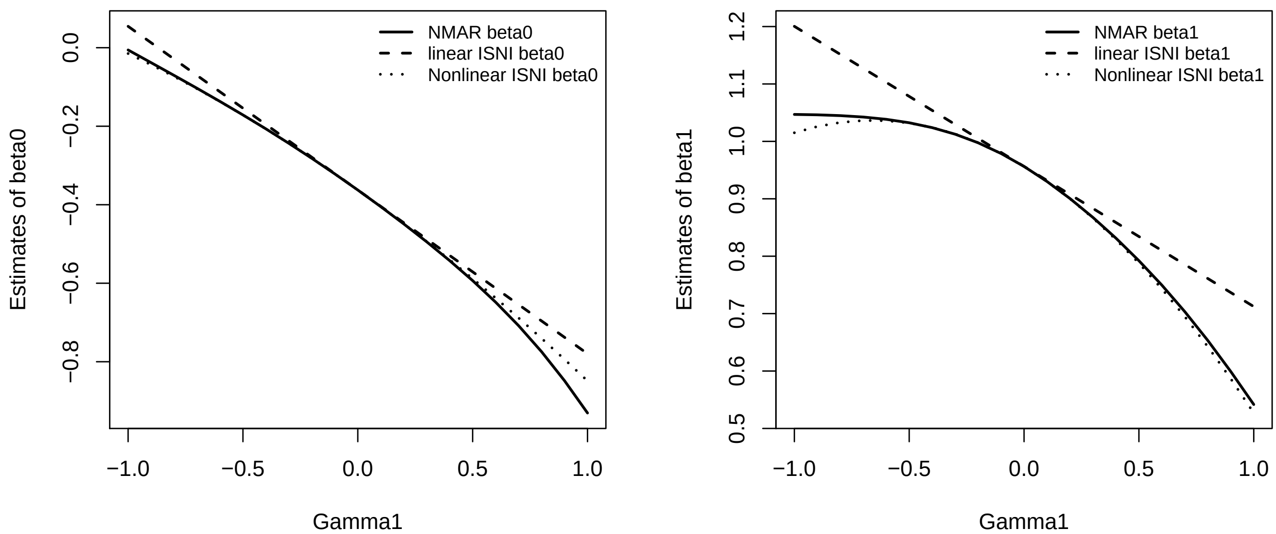

Figure 2: ISNI approximation to β_0 and β_1 estimates for simple linear regression model

TABLE III: A SIMULATION STUDY OF THE ACCURACY OF ISNI TO APPROXIMATE MLE OF THE INTERCEPT β_0 CHANGES LOCALLY FOR SIMPLE LINEAR REGRESSION MODEL

			Average				Approximation Error	
γ_0	γ_1	Prop_M	$\widehat{\beta_0}(0)$	$\widehat{\beta_0}(\gamma_1)$	$\widehat{\beta_0}(\gamma_1)$ (ISNIL)	$\widehat{\beta_0}(\gamma_1)$ (ISNIQ)	$\widehat{\beta_0}(\gamma_0)$ (ISNIL)	$\widehat{\beta_0}(\gamma_1)$ (ISNIQ)
0	-0.5	50.88%	-0.226	0.007	0.035	0.005	0.028	0.003
0	0.5	49.12%	0.24	0.018	-0.006	0.019	0.024	0
0	1	49.46%	0.411	0.021	-0.103	0.037	0.125	0.006
1	-1	33.22%	-0.259	0.007	0.08	0.009	0.073	0.003
1	-0.5	29.49%	-0.129	0.005	0.014	0.005	0.01	0
1	0.5	28.61%	0.146	0.016	0.007	0.016	0.009	0.003
1	1	32.01%	0.272	0.014	-0.051	0.011	0.065	0.008
2	-1	18.85%	-0.145	0.008	0.04	0.015	0.032	0.006
2	-0.5	13.96%	-0.049	0.016	0.019	0.017	0.002	0
2	0.5	14.03%	0.075	0.009	0.006	0.008	0.003	0.001
2	1	18.37%	0.162	0.013	-0.018	0.006	0.031	0.018
3	-1	9.08%	-0.066	0.011	0.022	0.015	0.011	0.004
3	-0.5	5.83%	-0.014	0.014	0.015	0.014	0.001	0
3	0.5	6.2%	0.041	0.011	0.01	0.011	0.001	0
3	1	9.51%	0.09	0.01	-0.004	0.005	0.014	0.008

TABLE IV: A SIMULATION STUDY OF THE ACCURACY OF ISNI TO APPROXIMATE MLE OF THE SLOPE β_1 CHANGES LOCALLY FOR SIMPLE LINEAR REGRESSION MODEL

γ_0	γ_1	Prop_M	Average				Approximation Error	
			$\hat{\beta_1}(0)$	$\hat{\beta_1}(\gamma_1)$	$\hat{\beta_1}(\gamma_1)$ (ISNIL)	$\hat{\beta_1}(\gamma_1)$ (ISNIQ)	$\hat{\beta_1}(\gamma_1)$ (ISNIL)	$\hat{\beta_1}(\gamma_1)$ (ISNIQ)
0	-1	50.54%	0.843	0.989	1.207	0.955	0.218	0.039
0	-0.5	50.88%	0.942	0.997	1.062	0.993	0.065	0.007
0	0.5	49.12%	0.953	1.007	1.063	1.004	0.056	0.001
0	1	49.46%	0.868	1.006	1.198	0.974	0.193	0.009
1	-1	33.22%	0.866	0.984	1.103	0.982	0.119	0.015
1	-0.5	29.49%	0.95	0.988	1.012	0.988	0.025	0.001
1	0.5	28.61%	0.96	1.001	1.024	1.002	0.023	0.005
1	1	32.01%	0.881	0.999	1.106	0.999	0.106	0.006
2	-1	18.85%	0.912	0.993	1.049	1.002	0.056	0.009
2	-0.5	13.96%	0.982	1.001	1.007	1.002	0.007	0
2	0.5	14.03%	0.978	1.008	1.016	1.01	0.007	0.001
2	1	18.37%	0.918	1.006	1.058	1.016	0.052	0.019
3	-1	9.08%	0.943	0.995	1.015	1.003	0.02	0.012
3	-0.5	5.83%	0.988	0.998	0.999	0.998	0.001	0
3	0.5	6.2%	0.985	1.002	1.004	1.002	0.002	0
3	1	9.51%	0.939	1.002	1.026	1.011	0.024	0.008

CHAPTER 3

NONLINEAR SENSITIVITY INDEX FOR MISSINGNESS IN BOTH THE OUTCOME AND COVARIATES

(Previously published as Gao, W., Hedeker, D., Mermelstein, R., Xie, H. (2016). A scalable approach to measuring the impact of nonignorable nonresponse with an EMA application. Statistics in Medicine. DOI: 10.1002/sim.7078.)

3.1 Motivation: An EMA Study on Adolescent Smoking Behaviors

In the previous chapter, we had developed ISNI under the condition that only the outcome Y is subject to missingness while the covariates X are fully observed. This chapter develops nonlinear sensitivity indices for selective missingness in both the outcome and covariates, which is motivated by the data collected in the EMA study as described in Example 2 in the Introduction. One important study aim in the aforementioned EMA study is to investigate the mood-link relationship for adolescents, and regression models are commonly employed for this purpose. In these regression models, the covariates in X may include other control variables for the purpose of studying mediation or controlling for confounding in addition to the main independent variable smoking. The collected EMA data has the more complex missing data pattern that when study participants do not respond to the random prompts, the answers to the questions regarding mood (the outcome Y) and some covariates in X are simultaneously missing. The simultaneous missingness in both the outcome variable and important covariates that would be collected concurrently during those nonresponded prompts is a common problem in EMA studies. Standard analysis ignores these missing prompts and analyzes the resulting data as if these random prompts were never sent out. It can be important to know the credibility of such standard analysis by examining how the standard inference could be affected if the missing random prompts are systematically different from the responded prompts. Note that such data is not restricted to EMA data only, but can happen to other kinds of studies when there exists unit nonresponses to group of questions in the survey. This raises new statistical and computational challenges. The current ISNI formulas are not computable in this situation 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. In this chapter, we extend the local sensitivity analysis to allow for the missing data pattern of simultaneous missingness in outcomes and covariates and to capture the more complex nonlinear impact of nonignorability.

3.2 Selection Model

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 Chapter 2 to include an covariate model for missing covariates. 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. We consider here models for cross-sectional data, although our index approach is general and can be extended to more complex models.

The joint selection model for missingness in both the outcome and covariates consists of the following three components:

First is outcome model. Let the outcome Y have a density function $f_{\theta_y}(Y|X, W)$ for independent subjects $i = 1, \cdots, N$ from the population of interest. The outcome Y_i is independently drawn from the exponential family that includes normal, binomial, Poisson and Gamma distributions as special cases.

Second is covariate model. If X consists of a single covariate, a generalized linear model can be used for $f_{\theta_x}(X|W)$ depending on the nature of the covariate. If X consists of multiple covariates $(X_1, X_2, \cdots X_p)$, we can consider using a product conditional model where $f_{\theta_x}(X|W) = f_{\theta_{x_1}}(X_1|W) f_{\theta_{x_2}}(X_2|W, X_1) \cdots f_{\theta_{x_p}}(X_p|W, X_1, \cdots, X_{p-1})$. In the product conditional model, each of $f_{\theta_{x_j}}(X_j|W, \tilde{X}_j), j = 1, \cdots, p$ can be modeled using a generalized linear model, where $\tilde{X}_j = (X_1, \cdots X_{j-1})$ when $j > 1$ and $\tilde{X}_j$ is null when $j = 1$. X_i can be continuous or categorical variables.

Last is missing data model. To investigate the potential impact of nonignorable prompt nonresponse, we further assume the following model for describing the prompt response behavior for subject i:

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

Where $G_i = 1(0)$ if subject's outcome is observed (subject's outcome is missing or unobserved), $h(\cdot)$ is the specified monotonic link function, e.g., the logit or probit, s_i includes a set of observed predictors for prompt response, and $\gamma_1 = (\gamma_{1x}, \gamma_{1y})^T$ is a vector of nonignorability parameters that associate the probability of missingness with the outcome and the covariates that were concurrently missing. When the nonignorability parameter $\gamma_1 (i.e., \gamma_{1x}, \gamma_{1y}$ departs from zero, the prompt nonresponse depends on the potentially unobserved mood outcome (e.g. mood) and covariate values. Consequently, the observed mood outcomes become a selective subset of the orginal planned outcomes and cause selection bias in the MAR estimates of the outcome model parameters. The general missing data model includes the following two special cases:

Special Case 1:

Y-dependent-only Nonignorability, where $\gamma_{1x}^T = 0$, and $P(G_i = 1) = h(\gamma_0^T s_i + \gamma_{1y} y_i)$.

Special Case 2:

X-dependent-only Nonignorability, where $\gamma_{1y} = 0$, and $P(G_i = 1) = h(\gamma_0^T s_i + \gamma_{1x}^T x_i)$.

3.3 Linear and Nonlinear Sensitivity Index Development

The log likelihood for the above joint model is

$$\begin{aligned} L(\theta_y, \theta_x, \gamma_0, \gamma_1) \quad &= \textstyle\sum_{i:g_i=1} \ln f_{\theta_y}(y_i|x_i, w_i) + \sum_{i:g_i=1} \ln f_{\theta_x}(x_i|w_i) + \sum_{i:g_i=1} \ln f_{\gamma}(g_i|s_i, y_i, x_i) + \\ &\textstyle\sum_{i:g_i=0} \ln \left(\int_{\Omega_{Y_i}} \int_{\Omega_{X_{1i}}} \cdots \int_{\Omega_{X_{Pi}}} f_{\gamma}(g_i|s_i, y_i, x_i) f_{\theta_y}(y_i|x_i, w_i) f_{\theta_x}(x_i|w_i) dy_i, dx_{1i} \cdots dx_{Pi} \right) \end{aligned} \quad (3.1)$$

Where $i = 1, \cdots, N$, Ω_Y is the sample space of Y and Ω_{X_j} is the sample space of $X_j, j = 1, \cdots, P$. $f_{\theta_y}(y_i|x_i, w_i), f_{\theta_x}(x_i|w_i)$ and $f_{\gamma}(g_i|s_i, y_i, x_i)$ denote the density function for the outcome, covariates and prompt response behaviors, respectively.

As compared with the likelihood in Equation 2.1 which involves missingness only in the outcome, the missing data problem in EMA data becomes considerably more difficult to handle because of simultaneous missingness in both the outcome and covariates. The high-dimensional missing data problem caused by more complex missing data patterns and reasons can increase further if we consider more complex longitudinal data analysis. For example, in an EMA study that has on average 40 planned observations per subject and a nonresponse rate of 25%, this will lead to a minimum of an additional 10-dimension integration per subject on average beyond the integrals with respect to missing covariates and/or random effects! These difficult issues hinder the use of principled methods in EMA studies to quantify the impact of nonrandom missingness, and call for methods scalable to these new types of 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. Consequently, there is a need to generalize nonlinear ISNI to make it applicable in EMA studies.

Upon the log-likelihood defined in Equation 3.1, we have $\hat{\theta}(\gamma_1) = \left(\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1), \hat{\gamma}_0(\gamma_1)\right)$ be the MLEs, given γ_1 and $\hat{\theta}^k(\gamma_1)$ be the kth element of $\hat{\theta}(\gamma_1)$. By approximating the log-likelihood with a Taylor series expansion of the MLEs of the model parameters as a function of γ_1, we have the following:

$$\hat{\theta}^k(\gamma_1) \overbrace{\approx}^{\mathrm{Taylor Expansion}} \hat{\theta}^k(0) + \gamma_1^T \overbrace{\left.\frac{\partial \hat{\theta}^k(\gamma_1)}{\partial \gamma_1}\right|_{\gamma_1=0}}^{\text{ISNIL}} + \frac{1}{2}\gamma_1^T \overbrace{\left.\frac{\partial^2 \hat{\theta}^k(\gamma_1)}{\partial \gamma_1 \gamma_1^T}\right|_{\gamma_1=0}}^{\text{ISNIQ}} \gamma_1 \tag{3.2}$$

We follow two steps to develop the nonlinear sensitivity indices 1) extending linear ISNI; 2) developing measures of higher-order sensitivity.

First, we extend the linear ISNI (ISNIL) to account for missing linear covariates. For each fixed value of γ_1, the conditional estimates $\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1)$, and $\hat{\gamma}_0(\gamma_1)$ satisfy the following condition

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

Where $L(\theta_y, \theta_x, \gamma_0, \gamma_1)$ is the log likelihood for the selection model as defined in Equation 3.1. Taking the first derivative of both sides of the above equation with respect to γ_1 and noting that $\hat{\theta}_y(\gamma_1)$, $\hat{\theta}_x(\gamma_1)$, and $\hat{\gamma}_0(\gamma_1)$ are implicit functions of γ_1, we have

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

Thus, for any γ_1 value, we have

$$\frac{\partial(\hat{\theta}_y^T(\gamma_1), \hat{\theta}_x^T(\gamma_1), \hat{\gamma}_0^T(\gamma_1))^T}{\partial\gamma_1^T} = -\left[\frac{\partial^2 L(\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)}{\partial(\theta_y^T, \theta_x^T, \gamma_0^T)^T \partial(\theta_y^T, \theta_x^T, \gamma_0^T)}\right]^{-1} \frac{\partial^2 L(\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)^T}{\partial(\theta_y^T, \theta_x^T, \gamma_0^T)^T \partial\gamma_1^T} \tag{3.3}$$

In our local sensitivity analysis, the primary interest is to investigate the sensitivity around the standard MAR model, i.e. γ_1. This local sensitivity can be captured by the derivatives at this point. In particular, we define the first order derivative evaluated at γ_1 as ISNIL

$$\text{ISNIL} = \begin{bmatrix} \frac{\partial \hat{\theta}_y(\gamma_1^T)}{\partial \gamma_1} \\ \frac{\partial \hat{\theta}_x(\gamma_1^T)}{\partial \gamma_1} \\ \frac{\partial \hat{\gamma_0}(\gamma_1^T)}{\partial \gamma_1} \end{bmatrix}_{\gamma_1=0} = -\begin{bmatrix} \nabla^2 L_{\theta_y,\theta_y} & \nabla^2 L_{\theta_y,\theta_x} & \nabla^2 L_{\theta_y,\gamma_0} \\ \nabla^2 L_{\theta_x,\theta_y} & \nabla^2 L_{\theta_x,\theta_x} & \nabla^2 L_{\theta_x,\gamma_0} \\ \nabla^2 L_{\gamma_0,\theta_y} & \nabla^2 L_{\gamma_0,\theta_x} & \nabla^2 L_{\gamma_0,\gamma_0} \end{bmatrix}^{-1} \begin{bmatrix} \nabla^2 L_{\theta_y,\gamma_{1y}} & \nabla^2 L_{\theta_y,\gamma_{1x}} \\ \nabla^2 L_{\theta_x,\gamma_{1y}} & \nabla^2 L_{\theta_x,\gamma_{1x}} \\ \nabla^2 L_{\gamma_0,\gamma_{1y}} & \nabla^2 L_{\gamma_0,\gamma_{1x}} \end{bmatrix},$$

Where for arguments a, b, $\nabla^2 L_{a,b} = \frac{\partial^2 L(\hat{\theta}_y(\gamma_1),\hat{\theta}_x(\gamma_1),\hat{\gamma}_0(\gamma_1),\gamma_1)}{\partial a \partial b}\Big|_{\hat{\theta}_y(0),\hat{\theta}_x(0),\hat{\gamma}_0(0),0}$. Under MAR and the case Y and X are missing simultaneously, we have $\nabla^2 L_{\theta_y,\theta_x} = 0, \nabla^2 L_{\theta_y,\gamma_0} = 0$, and $\nabla^2 L_{\theta_x,\gamma_0} = 0$, and thus the ISNIL for $\hat{\theta}_y$, the parameter estimates of primary interest, is

$$\left.\frac{\partial \hat{\theta}_y(\gamma_1)}{\partial \gamma_1^T}\right|_{\gamma_1=0} = -\left[\nabla^2 L^{-1}_{\theta_y,\theta_y} \nabla^2 L_{\theta_y,\gamma_{1y}}, \nabla^2 L^{-1}_{\theta_y,\theta_y} \nabla^2 L_{\theta_y,\gamma_{1x}}\right] = \left[\text{ISNIL}_y(\hat{\theta}_y), \text{ISNIL}_x(\hat{\theta}_y)\right],$$

where

$$\begin{aligned} \nabla^2 L_{\theta_y,\theta_y} &= \sum_{i:g_i=1} \left.\frac{\partial \ln f_{\theta_y}(y_i^{obs}|x_i^{obs}, w_i)}{\partial \theta_y \partial \theta_y^T}\right|_{\gamma_1=0}, \\ \nabla^2 L_{\theta_y,\gamma_{1y}} &= -\sum_{i:g_i=0} h_i \cdot E_{x_i^{mis}|w_i}\left.\left(\frac{\partial E(y_i^{mis}|x_i^{mis}, w_i)}{\partial \theta_y}\right)\right|_{\gamma_1=0}, \\ \nabla^2 L_{\theta_y,\gamma_{1x}} &= -\sum_{i:g_i=0} h_i \cdot \left.\frac{\partial E(x_i^{mis}|w_i)}{\partial \theta_y}\right|_{\gamma_1=0} = 0 \end{aligned}$$

$h_i = \mathsf{Prob}(G_i = 1)$ denoting the probability of being observed under the MAR model. The superscript obs (mis) indicates that the corresponding data element is observed (or missing). Therefore we have

$$\begin{aligned}
\text{ISNIL}_y(\hat{\theta}_y) &= \nabla^2 L^{-1}_{\theta_y,\theta_y} \nabla^2 L_{\theta_y,\gamma_{1y}}, \\
\text{ISNIL}_x(\hat{\theta}_y) &= \nabla^2 L^{-1}_{\theta_y,\theta_y} \nabla^2 L_{\theta_y,\gamma_{1x}} = 0
\end{aligned}$$

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. 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, it follows from (40) 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.

Next, we develop measures of higher-order sensitivity. The current linear ISNI approach loses its effectiveness when the impact of nonignorability takes a highly nonlinear shape around the MAR model in the situation of missingness in both of the outcome and covariates. Thus, there is a need to generalize nonlinear ISNI (ISNIQ) to make it applicable to this kind of situation. For notational simplicity, we rewrite $\gamma_1 = (\gamma_{1x}, \gamma_{1y}) = (\gamma_{11}, \gamma_{12}, \cdots, \gamma_{1P}, \gamma_{1y})$. We let γ_{1p} and γ_{1q} denote two generic elements in the γ_1 vector. Taking the second derivative of Equation 3.3 with respect to γ_1, we have

$$\begin{aligned}
&\frac{\partial^2\left(\hat{\theta}_y^T(\gamma_1), \hat{\theta}_x^T(\gamma_1), \hat{\gamma}_0^T(\gamma_1)\right)^T}{\partial\gamma_{1p}\partial\gamma_{1q}} = \\
&\left[\frac{\partial^2 L(\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)}{\partial(\theta_y^T, \theta_x^T, \gamma_0^T)^T\partial(\theta_y^T, \theta_x^T, \gamma_0^T)}\right]^{-1} \frac{\partial}{\partial\gamma_{1q}}\left[\frac{\partial^2 L(\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)}{\partial(\theta_y^T, \theta_x^T, \gamma_0^T)^T\partial(\theta_y^T, \theta_x^T, \gamma_0^T)}\right] \\
&\cdot\left[\frac{\partial^2 L(\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)}{\partial(\theta_y^T, \theta_x^T, \gamma_0^T)^T\partial(\theta_y^T, \theta_x^T, \gamma_0^T)}\right]^{-1}\left[\frac{\partial^2 L(\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)}{\partial(\theta_y^T, \theta_x^T, \gamma_0^T)^T\partial(\theta_y^T, \theta_x^T, \gamma_0^T)}\right] \\
&-\left[\frac{\partial^2 L(\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)}{\partial(\theta_y^T, \theta_x^T, \gamma_0^T)^T\partial(\theta_y^T, \theta_x^T, \gamma_0^T)}\right]^{-1} \frac{\partial}{\partial\gamma_{1q}}\left[\frac{\partial^2 L(\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)}{\partial(\theta_y^T, \theta_x^T\gamma_0^T)^T\partial\gamma_{1p}}\right],
\end{aligned} \tag{3.4}$$

Where

$$\begin{aligned}
&\frac{\partial}{\partial\gamma_{1q}}\left[\frac{\partial^2 L(\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)}{\partial(\theta_y^T, \theta_x^T, \gamma_0^T)^T\partial(\theta_y^T, \theta_x^T, \gamma_0^T)}\right] = \\
&\quad \frac{\partial^3 L(\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)}{\partial(\theta_y^T, \theta_x^T, \gamma_0^T)^T\partial(\theta_y^T, \theta_x^T, \gamma_0^T)\partial\gamma_{1q}} + \sum_{j=1}^{n_{\theta_y}} \frac{\partial^3 L(\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)}{\partial(\theta_y^T, \theta_x^T, \gamma_0^T)^T\partial(\theta_y^T, \theta_x^T, \gamma_0^T)\partial\theta_y^j}\frac{\partial\hat{\theta}_y^j(\gamma_1)}{\partial\gamma_{1q}} \\
&+\sum_{j=1}^{n_{\theta_x}} \frac{\partial^3 L(\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)}{\partial(\theta_y^T, \theta_x^T, \gamma_0^T)^T\partial(\theta_y^T, \theta_x^T, \gamma_0^T)\partial\theta_x^j}\frac{\partial\hat{\theta}_x^j(\gamma_1)}{\partial\gamma_{1q}} + \sum_{j=1}^{n_{\gamma_0}} \frac{\partial^3 L(\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)}{\partial(\theta_y^T, \theta_x^T, \gamma_0^T)^T\partial(\theta_y^T, \theta_x^T, \gamma_0^T)\partial\gamma_0^j}\frac{\partial\hat{\gamma}_0^j(\gamma_1)}{\partial\gamma_{1q}},
\end{aligned}$$

$n_{\theta_y}, n_{\theta_x}$ and n_{γ_0} are the length of θ_y, θ_x and γ_0 respectively, $\hat{\theta}_y^j, \hat{\theta}_x^j$ and $\hat{\gamma}_0^j$ are the jth element of $\hat{\theta}_y, \hat{\theta}_x$ and $\hat{\gamma}_0$, respectively, and

$$\frac{\partial}{\partial \gamma_{1q}}\left[\frac{\partial^2 L(\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)}{\partial(\theta_y^T, \theta_x^T, \gamma_0^T)^T \partial \gamma_{1p}}\right] =$$
$$\frac{\partial^3 L(\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)}{\partial(\theta_y^T, \theta_x^T, \gamma_0^T)^T \partial(\theta_y^T, \theta_x^T, \gamma_0^T) \partial \gamma_{1p}} \frac{\partial(\theta_y^T, \theta_x^T, \gamma_0^T)^T}{\partial \gamma_{1q}} + \frac{\partial^3 L(\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1), \hat{\gamma}_0(\gamma_1), \gamma_1)}{\partial(\theta_y^T, \theta_x^T, \gamma_0^T)^T \partial \gamma_{1p} \gamma_{1q}}$$

We call the second derivative of the $(\hat{\theta}_y(\gamma_1), \hat{\theta}_x(\gamma_1), \hat{\gamma_0}(\gamma_1))$ evaluated at the MAR model, i.e., $\gamma_1 = 0$ as Index of Sensitivity to NonIgnorability in Quadratic form (ISNIQ). The form is as below

$$\mathrm{ISNIQ}_{pq} = \begin{bmatrix} \frac{\partial^2 \hat{\theta}_y(\gamma_1)}{\partial \gamma_{1p}\gamma_{1q}} \\ \frac{\partial^2 \hat{\theta}_x(\gamma_1)}{\partial \gamma_{1p}\gamma_{1q}} \\ \frac{\partial^2 \hat{\gamma}_0(\gamma_1)}{\partial \gamma_{1p},\gamma_{1q}} \end{bmatrix}_{\gamma_1=0}$$

$$= -\begin{bmatrix} \nabla^2 L_{\theta_y,\theta_y} & \nabla^2 L_{\theta_y,\theta_x} & \nabla^2 L_{\theta_y,\gamma_0} \\ \nabla^2 L_{\theta_x,\theta_y} & \nabla^2 L_{\theta_x,\theta_x} & \nabla^2 L_{\theta_x,\gamma_0} \\ \nabla^2 L_{\gamma_0,\theta_y} & \nabla^2 L_{\gamma_0,\theta_x} & \nabla^2 L_{\gamma_0,\gamma_0} \end{bmatrix}^{-1} \left(\left(\begin{bmatrix} \nabla^3 L_{\theta_y,\theta_y,\gamma_{1q}} & \nabla^3 L_{\theta_y,\theta_x,\gamma_{1q}} & \nabla^3 L_{\theta_y,\gamma_0,\gamma_{1q}} \\ \nabla^3 L_{\theta_x,\theta_y,\gamma_{1q}} & \nabla^3 L_{\theta_x,\theta_x,\gamma_{1q}} & \nabla^3 L_{\theta_x,\gamma_0,\gamma_{1q}} \\ \nabla^3 L_{\gamma_0,\theta_y,\gamma_{1q}} & \nabla^3 L_{\gamma_0,\theta_x,\gamma_{1q}} & \nabla^3 L_{\gamma_0,\gamma_0,\gamma_{1q}} \end{bmatrix} + \right.\right.$$

$$\sum_{j=1}^{n_{\theta_y}} \begin{bmatrix} \nabla^3 L_{\theta_y,\theta_y,\theta_y^j} & \nabla^3 L_{\theta_y,\theta_x,\theta_y^j} & \nabla^3 L_{\theta_y,\gamma_0,\theta_y^j} \\ \nabla^3 L_{\theta_x,\theta_y,\theta_y^j} & \nabla^3 L_{\theta_x,\theta_x,\theta_y^j} & \nabla^3 L_{\theta_x,\gamma_0,\theta_y^j} \\ \nabla^3 L_{\gamma_0,\theta_y,\theta_y^j} & \nabla^3 L_{\gamma_0,\theta_x,\theta_y^j} & \nabla^3 L_{\gamma_0,\gamma_0,\theta_y^j} \end{bmatrix} \mathrm{ISNIL}_m(\hat{\theta}_y^j) + \sum_{j=1}^{n_{\theta_x}} \begin{bmatrix} \nabla^3 L_{\theta_y,\theta_y,\theta_x^j} & \nabla^3 L_{\theta_y,\theta_x,\theta_x^j} & \nabla^3 L_{\theta_y,\gamma_0,\theta_x^j} \\ \nabla^3 L_{\theta_x,\theta_y,\theta_x^j} & \nabla^3 L_{\theta_x,\theta_x,\theta_x^j} & \nabla^3 L_{\theta_x,\gamma_0,\theta_x^j} \\ \nabla^3 L_{\gamma_0,\theta_y,\theta_x^j} & \nabla^3 L_{\gamma_0,\theta_x,\theta_x^j} & \nabla^3 L_{\gamma_0,\gamma_0,\theta_x^j} \end{bmatrix} \mathrm{ISNIL}_m(\hat{\theta}_x^j)$$

$$\left.\left. + \sum_{j=1}^{n_{\gamma_0}} \begin{bmatrix} \nabla^3 L_{\theta_y,\theta_y,\gamma_0^j} & \nabla^3 L_{\theta_y,\theta_x,\gamma_0^j} & \nabla^3 L_{\theta_y,\gamma_0,\gamma_0^j} \\ \nabla^3 L_{\theta_x,\theta_y,\gamma_0^j} & \nabla^3 L_{\theta_x,\theta_x,\gamma_0^j} & \nabla^3 L_{\theta_x,\gamma_0,\gamma_0^j} \\ \nabla^3 L_{\gamma_0,\theta_y,\gamma_0^j} & \nabla^3 L_{\gamma_0,\theta_x,\gamma_0^j} & \nabla^3 L_{\gamma_0,\gamma_0,\gamma_0^j} \end{bmatrix} \mathrm{ISNIL}_m(\hat{\gamma}_0^j) \right) \mathrm{ISNIL}_k \left(\begin{bmatrix} \hat{\theta}_y \\ \hat{\theta}_x \\ \hat{\gamma}_0 \end{bmatrix} \right)\right)$$

$$- \begin{bmatrix} \nabla^2 L_{\theta_y,\theta_y} & \nabla^2 L_{\theta_y,\theta_x} & \nabla^2 L_{\theta_y,\gamma_0} \\ \nabla^2 L_{\theta_x,\theta_y} & \nabla^2 L_{\theta_x,\theta_x} & \nabla^2 L_{\theta_x,\gamma_0} \\ \nabla^2 L_{\gamma_0,\theta_y} & \nabla^2 L_{\gamma_0,\theta_x} & \nabla^2 L_{\gamma_0,\gamma_0} \end{bmatrix}^{-1}$$

$$\cdot \left(\begin{bmatrix} \nabla^3 L_{\theta_y,\gamma_{1p},\gamma_{1q}} \\ \nabla^3 L_{\theta_x,\gamma_{1p},\gamma_{1q}} \\ \nabla^3 L_{\gamma_0,\gamma_{1p},\gamma_{1q}} \end{bmatrix} + \begin{bmatrix} \nabla^3 L_{\theta_y,\theta_y,\gamma_{1p}} & \nabla^3 L_{\theta_y,\theta_x,\gamma_{1p}} & \nabla^3 L_{\theta_y,\gamma_0,\gamma_{1p}} \\ \nabla^3 L_{\theta_x,\theta_y,\gamma_{1p}} & \nabla^3 L_{\theta_x,\theta_x,\gamma_{1p}} & \nabla^3 L_{\theta_x,\gamma_0,\gamma_{1p}} \\ \nabla^3 L_{\gamma_0,\theta_y,\gamma_{1p}} & \nabla^3 L_{\gamma_0,\theta_x,\gamma_{1p}} & \nabla^3 L_{\gamma_0,\gamma_0,\gamma_{1p}} \end{bmatrix} \mathrm{ISNIL}_m \left(\begin{bmatrix} \hat{\theta}_y \\ \hat{\theta}_x \\ \hat{\gamma}_0 \end{bmatrix} \right)\right)$$

All the terms in the right hand side of the above equation are evaluated using the readily-available MAR estimates. Under the MAR and that Y and X are subject to simultaneously missingness, we further have $\nabla^2 L_{\theta_y,\gamma_0} = 0$, $\nabla^2 L_{\theta_x,\gamma_0} = 0$, $\nabla^2 L_{\theta_y,\theta_x} = 0$, $\nabla^3 L_{\theta_y,\gamma_0,\theta_y^j} = 0$, $\nabla^3 L_{\gamma_0,\gamma_0,\theta_y^j} = 0$, $\nabla^3 L_{\theta_y,\theta_x,\theta_y^j} = 0$, $\nabla^3 L_{\theta_x,\theta_x,\theta_y^j} = 0$, $\nabla^3 L_{\theta_x,\gamma_0,\theta_y^j} = 0$, $\nabla^3 L_{\theta_y,\gamma_0,\theta_x^j} = 0$, $\nabla^3 L_{\gamma_0,\gamma_0,\theta_x^j} = 0$, $\nabla^3 L_{\theta_y,\theta_x,\theta_x^j} = 0$, $\nabla^3 L_{\theta_y,\theta_y,\theta_x^j} = 0$, $\nabla^3 L_{\theta_x,\gamma_0,\theta_x^j} = 0$, $\nabla^3 L_{\theta_x,\theta_x,\gamma_0^j} = 0$, $\nabla^3 L_{\theta_y,\gamma_0,\gamma_0^j} = 0$,

$\nabla^3 L_{\theta_y,\theta_x,\gamma_0^j} = 0$, $\nabla^3 L_{\theta_y,\theta_y,\gamma_0^j} = 0$, $\nabla^3 L_{\theta_x,\gamma_0,\gamma_0^j} = 0$. Thus we have ISNIQ for the outcome model parameter $\hat{\theta}_y$, the parameter of primary interest, as

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

Where for arguments a, b, c, $\nabla^3 L_{a,b,c} = \left.\frac{\partial^3 L(\hat{\theta}_y(\gamma_1),\hat{\theta}_x(\gamma_1),\hat{\gamma}_0(\gamma_1),\gamma_1)}{\partial a\partial b\partial c}\right|_{\hat{\theta}_y(0),\hat{\theta}_x(0),\hat{\gamma}_0(0),0}$, $\mathrm{ISNIL}_p(u) = \frac{\partial u}{\partial\gamma_{1p}}$ and $\mathrm{ISNIL}_q(u) = \frac{\partial u}{\partial\gamma_{1q}}$ for the argument u, and $\hat{\theta}_y^j$ denotes the jth element of $\hat{\theta}$.

As shown above, $\mathrm{ISNIQ}(\hat{\theta}_y)$ 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 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

$$
\begin{aligned}
\mathrm{ISNIQ}_{yy}(\hat{\theta}_y) &= \left.\frac{\partial^2\hat{\theta}_y(\gamma_1)}{\partial\gamma_{1y}^2}\right|_{\gamma_1=0} \\
&= -2\nabla^2 L^{-1}_{\theta_y,\theta_y}\nabla^3 L_{\theta_y,\theta_y,\gamma_{1y}}\mathrm{ISNIL}_y(\hat{\theta}_y) \\
&\quad -\nabla^2 L^{-1}_{\theta_y,\theta_y}(\sum_{j=1}^{n_{\theta_y}}\nabla^3 L_{\theta_y,\theta_y,\theta_y^j}\mathrm{ISNIL}_y(\hat{\theta}_y^j))\mathrm{ISNIL}_y(\hat{\theta}_y) \\
&\quad -2\nabla^2 L^{-1}_{\theta_y,\theta_y}\nabla^3 L_{\theta_y,\theta_x,\gamma_{1y}}\mathrm{ISNIL}_y(\hat{\theta}_x) \\
&\quad -2\nabla^2 L^{-1}_{\theta_y,\theta_y}\nabla^3 L_{\theta_y,\gamma_0,\gamma_{1y}}\mathrm{ISNIL}_y(\hat{\gamma}_0) \\
&\quad -\nabla^2 L^{-1}_{\theta_y,\theta_y}\nabla^3 L_{\theta_y,\gamma_{1y},\gamma_{1y}}
\end{aligned}
$$

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

$$
\begin{aligned}
\mathrm{ISNIQ}_{pq}(\hat{\theta}_y) &= \left.\frac{\partial^2\hat{\theta}_y(\gamma_1)}{\partial\gamma_{1p}\partial\gamma_{1q}}\right|_{\gamma_1=0} \\
&= 0, \qquad \text{for } \gamma_{1p}\in\gamma_{1x}=(\gamma_{11},\cdots,\gamma_{1P}) \text{ and } \gamma_{1q}\in\gamma_{1x}=(\gamma_{11},\cdots,\gamma_{1P})
\end{aligned}
$$

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

Assuming that we have the logit link for missing data model, the following steps show the detailed derivation for the individual terms in ISNIL and ISNIQ.

$$
\begin{aligned}
\nabla^2 L_{\theta_y,\theta_y} &= \sum_{i:g_i=1} \left.\frac{\partial^2 \ln f_{\theta_y}(y_i^{obs}|x_i^{obs}, w_i)}{\partial\theta_y \partial\theta_y^T}\right|_{\gamma_1=0} \\
\nabla^2 L_{\theta_x,\theta_x} &= \sum_{i:g_i=1} \left.\frac{\partial^2 \ln f_{\theta_x}(x_i^{obs}|w_i)}{\partial\theta_x \partial\theta_x^T}\right|_{\gamma_1=0} \\
\nabla^2 L_{\theta_y,\theta_x} &= 0 \\
\nabla^2 L_{\theta_y,\gamma_0} &= 0 \\
\nabla^2 L_{\theta_x,\gamma_0} &= 0 \\
\nabla^2 L_{\theta_y,\gamma_{1y}} &= -\sum_{i:g_i=0} h_i \frac{\partial}{\partial\theta_y} E_{x_i|w_i}(E(y_i^{mis}|x_i^{mis}, w_i))\bigg|_{\gamma_1=0} \\
\nabla^2 L_{\theta_x,\gamma_{1y}} &= -\sum_{i:g_i=0} h_i \frac{\partial}{\partial\theta_x} E_{x_i|w_i}(E(y_i^{mis}|x_i^{mis}, w_i))\bigg|_{\gamma_1=0} \\
\nabla^2 L_{\gamma_0,\gamma_0} &= \frac{\partial^2}{\partial\gamma_0\partial\gamma_0^T} \sum_{i=1}^{N} \ln f_\gamma(g_i|y_i, x_i, s_i)\bigg|_{\gamma_1=0} \\
\nabla^2 L_{\gamma_0,\gamma_{1y}} &= -\left[\sum_{i:g_i=1}(h_i(1-h_i))y_i s_i + \sum_{i:g_i=0}(h_i(1-h_i))E_{(x_i,y_i)|w_i}(y_i^{mis}|x_i^{mis}, w_i)s_i)\right]_{\gamma_1=0} \\
\nabla^2 L_{\theta_x,\gamma_{1q}} &= -\sum_{i:g_i=0} h_i \frac{\partial}{\partial\theta_x} E_{x_{iq}|w_i}(x_{iq}^{mis}|w_i)\bigg|_{\gamma_1=0} \\
\nabla^2 L_{\gamma_0,\gamma_{1q}} &= -\left[\sum_{i:g_i=1}(h_i(1-h_i))x_i s_i + \sum_{i:g_i=0}(h_i(1-h_i))E_{x_{iq}|w_i}(x_{iq}^{mis}|w_i)s_i)\right]_{\gamma_1=0} \\
\nabla^3 L_{\theta_y,\theta_y,\gamma_{1y}} &= -\sum_{i:g_i=0} h_i \frac{\partial^2}{\partial\theta_y\partial\theta_y^T} E_{(x_i,y_i)|w_i}(y_i^{mis}|x_i^{mis}, w_i)\bigg|_{\gamma_1=0}
\end{aligned}
$$

$$\begin{aligned}
\nabla^3 L_{\theta_y,\theta_y,\theta_y^j} &= \sum_{i:g_i=1} \left.\frac{\partial^3 \ln f_{\theta_y}(y_i^{obs}|x_i^{obs}, w_i)}{\partial\theta_y \partial\theta_y^T \partial\theta_y^j}\right|_{\gamma_1=0} \\
\nabla^3 L_{\theta_y,\theta_x,\gamma_{1y}} &= -\sum_{i:g_i=0} h_i \left.\frac{\partial^2}{\partial\theta_y \partial\theta_x^T} E_{(x_i,y_i)|w_i}(y_i^{mis}|x_i^{mis}, w_i)\right|_{\gamma_1=0} \\
\nabla^3 L_{\theta_y,\gamma_0,\gamma_{1y}} &= -\sum_{i:g_i=0} (1-h_i)h_i \cdot \left.\frac{\partial}{\partial\theta_y} E_{(x_i,y_i)|w_i}(y_i^{mis}|x_i^{mis}, w_i) \cdot s_i^T\right|_{\gamma_1=0} \\
\nabla^3 L_{\theta_y,\gamma_{1y},\gamma_{1y}} &= -\sum_{i:g_i=0} \frac{\partial}{\partial\theta_y}\left[h_i(1-h_i)E^2(y_i^{mis}|w_i) + h_i(1-2h_i)var(y_i^{mis}|w_i)\right]_{\gamma_1=0} \\
\nabla^3 L_{\theta_y,\gamma_{1y},\gamma_{1q}} &= -\sum_{i:g_i=0} \frac{\partial}{\partial\theta_y}\left[h_i(1-h_i)E(y_i^{mis}|w_i)E(x_{iq}^{mis}|w_i) + h_i(1-2h_i)Cov(y_i^{mis}, x_{iq}^{mis}|w_i)\right]_{\gamma_1=0}
\end{aligned}$$

3.4 Examples

3.4.1 Simple Linear Regression with Both Outcome and Covariates Following a Normal Distribution

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

$$\left[\text{ISNIL}_x(\hat\beta_0), \text{ISNIL}_y(\hat\beta_0)\right] = \left[0, -\hat\sigma^2(0)\frac{N_m}{N}\right], \qquad \left[\text{ISNIL}_x(\hat\beta_1), \text{ISNIL}_y(\hat\beta_1)\right] = [0, 0],$$

And for the nonlinear components (ISNIQ) we have for $\gamma_1 = (\gamma_{1x}, \gamma_{1y})$

$$\mathrm{ISNIQ}(\hat{\beta}_0) = \begin{bmatrix} 0 & -\hat{\sigma}^2(0)\frac{N_o N_m}{N^2}\frac{N_o\hat{\sigma}_x^2(0)}{s_{xoxo}}\bar{x}_o \\ -\hat{\sigma}^2(0)\frac{N_o N_m}{N^2}\frac{N_o\hat{\sigma}_x^2(0)}{s_{xoxo}}\bar{x}_o & -2\hat{\beta}_1(0)\hat{\sigma}^2(0)\frac{N_o N_m}{N^2}\frac{N_o\hat{\sigma}_x^2(0)}{s_{xoxo}}\bar{x}_o \end{bmatrix}$$

$$\mathrm{ISNIQ}(\hat{\beta}_1) = \begin{bmatrix} 0 & \hat{\sigma}^2(0)\frac{N_o N_m}{N^2}\frac{N_o\hat{\sigma}_x^2(0)}{s_{xoxo}} \\ \hat{\sigma}^2(0)\frac{N_o N_m}{N^2}\frac{N_o\hat{\sigma}_x^2(0)}{s_{xoxo}} & 2\hat{\beta}_1(0)\hat{\sigma}^2(0)\frac{N_o N_m}{N^2}\frac{N_o\hat{\sigma}_x^2(0)}{s_{xoxo}} \end{bmatrix}$$

Where $s_{x_o x_o} = \sum_{i:g_i=1}(x_i - \bar{x}_o)^2$ 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, N_m is the number of missing observations and $N_o = N - N_m$.

All these entries in the ISNIQ matrices are evaluated using the MAR estimates only. 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 $\mathrm{ISNIL}_x(\hat{\beta}_1(0))$ and $\mathrm{ISNIL}_y(\hat{\beta}_1(0))$ are zeros) whereas the ISNIQ matrix for $\hat{\beta}_1(0)$ has non-zero entries, indicating the importance of examining the higher order sensitivity.

The following steps show the detailed derivation of individual terms in ISNIL and ISNIQ.

$$\nabla^2 L^{-1}_{\theta_y,\theta_y} = -\sigma^2 \begin{bmatrix} \frac{1}{N_o} + \frac{\bar{x}_0^2}{s_{xoxo}} & -\frac{\bar{x}_o}{s_{xoxo}} & 0 \\ -\frac{\bar{x}_o}{s_{xoxo}} & \frac{1}{s_{xoxo}} & 0 \\ 0 & 0 & \frac{2\sigma^2}{N_o} \end{bmatrix}$$

$$\nabla^3 L_{\theta_y,\theta_y,\gamma_{1y}} = 0$$

$$\nabla^3 L_{\theta_y,\theta_y,\beta_0} = \begin{bmatrix} 0 & 0 & \frac{N_o}{(\sigma^2)^2} \\ 0 & 0 & \frac{N_o \bar{x}_o}{(\sigma^2)^2} \\ \frac{N_o}{(\sigma^2)^2} & \frac{N_o \bar{x}_o}{(\sigma^2)^2} & 0 \end{bmatrix}$$

$$
\begin{aligned}
\nabla^3 L_{\theta_y,\theta_y,\beta_1} &= \begin{bmatrix} 0 & 0 & \frac{N_o \bar{x}_o}{(\sigma^2)^2} \\ 0 & 0 & \frac{\sum_i x_{io}^2}{(\sigma^2)^2} \\ \frac{N_o \bar{x}_o}{(\sigma^2)^2} & \frac{\sum_i x_{io}^2}{(\sigma^2)^2} & 0 \end{bmatrix} \\
\nabla^3 L_{\theta_y,\theta_x,\gamma_{1y}} &= \begin{bmatrix} 0 & 0 \\ -\frac{N_o N_m}{N} & 0 \\ 0 & 0 \end{bmatrix} \\
\mathrm{ISNIL}_y(\hat{\theta}_x) &= \begin{bmatrix} -\frac{N_m}{N}\sigma_x^2 \beta_1 \\ 0 \end{bmatrix} \\
\mathrm{ISNIL}_x(\hat{\theta}_x) &= \begin{bmatrix} -\frac{N_m}{N}\sigma_x^2 \\ 0 \end{bmatrix} \\
\nabla^3 L_{\theta_y,\gamma_0,\gamma_{1y}} &= \begin{bmatrix} -\frac{N_o (N_m)^2}{N^2} \\ -\frac{N_o (N_m)^2}{N^2}\bar{x}_o \\ 0 \end{bmatrix} \\
\mathrm{ISNIL}_y(\hat{\gamma}_0) &= -\beta_0 - \beta_1 \bar{x}_o \\
\mathrm{ISNIL}_x(\hat{\gamma}_0) &= -\bar{x}_o \\
\nabla^3 L_{\theta_y,\gamma_{1y},\gamma_{1y}} &= \begin{bmatrix} -\frac{2N_o (N_m)^2}{N^2}(\beta_0 + \beta_1 \bar{x}_o) \\ -\frac{2N_o (N_m)^2}{N^2}(\beta_0 \bar{x}_o + \beta_1 \bar{x}_o^2) - \frac{2N_o N_m (N_m - N_o)}{N^2}\sigma_x^2 \beta_1 \\ \frac{N_o N_m (N_o - N_m)}{n^2} \end{bmatrix} \\
\nabla^3 L_{\theta_y,\gamma_{1y},\gamma_{1x}} &= \begin{bmatrix} -\frac{N_o (N_m)^2}{N^2}\bar{x}_o \\ -\frac{N_o (N_m)^2}{N^2}\bar{x}_o^2 - \frac{N_o N_m (N_m - N_o)}{N^2}\sigma_x^2 \\ 0 \end{bmatrix},
\end{aligned}
$$

Where the model parameters in the above are replaced with MLEs under MAR when computing NISNIs.

3.4.2 Simple Linear Regression with a Normally Distributed Outcome and a Bernoulli Distributed Covariate

We consider another example where the outcome model is $Y_i \sim N(\beta_0+\beta_1 x_i, \sigma^2), i = 1, \cdots, N$, the covariate x_i is binary data following with Bernoulli distribution $X \sim B(1, p)$ and $p = \frac{\exp(\alpha X)}{1+\exp(\alpha X)}$. To preserve simplicity, we have $\text{logit}(P(X = 1)) = \alpha_0$. Therefore $p = \frac{\exp(\alpha_0)}{1+\exp(\alpha_0)}$. The prompt response model $\text{logit}(P(G_i = 1)) = \gamma_0 + \gamma_{1x} x_i + \gamma_{1y} y_i$. Using the general formula of ISNIL and ISNIQ above, we can derive linear components (ISNIL) as following

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

And nonlinear components (ISNIQ) as following with $\gamma_1 = (\gamma_{1x}, \gamma_{1y})$

$$\begin{aligned}
\text{ISNIQ}(\hat{\beta}_0) &= \begin{bmatrix} 0 & -\hat{\sigma}^2(0)\frac{N_o N_m}{N^2}\frac{N_o \hat{\sigma}_x^2(0)}{s_{x_o x_o}}\hat{p} \\ -\hat{\sigma}^2(0)\frac{N_o N_m}{N^2}\frac{N_o \hat{\sigma}_x^2(0)}{s_{x_o x_o}}\hat{p} & -2\hat{\beta}_1(0)\hat{\sigma}^2(0)\frac{N_o N_m}{N^2}\frac{N_o \hat{\sigma}_x^2(0)}{s_{x_o x_o}}\hat{p} \end{bmatrix} \\
\text{ISNIQ}(\hat{\beta}_1) &= \begin{bmatrix} 0 & \hat{\sigma}^2(0)\frac{N_o N_m}{N^2}\frac{N_o \hat{\sigma}_x^2(0)}{s_{x_o x_o}} \\ \hat{\sigma}^2(0)\frac{N_o N_m}{N^2}\frac{N_o \hat{\sigma}_x^2(0)}{s_{x_o x_o}} & 2\hat{\beta}_1(0)\hat{\sigma}^2(0)\frac{N_o N_m}{N^2}\frac{N_o \hat{\sigma}_x^2(0)}{s_{x_o x_o}} \end{bmatrix}
\end{aligned}$$

Where $s_{x_o x_o} = \sum_{i:g_i=1}(x_i - \hat{p})^2$ 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, N_m is the number of missing observations and $N_o = N - N_m$.

The following steps show the detailed derivation:

$$
\begin{aligned}
\nabla^2 L^{-1}_{\theta_y,\theta_y} &= -\sigma^2 \begin{bmatrix} \frac{1}{N_o} + \frac{\hat{p}^2}{s_{xoxo}} & -\frac{\hat{p}}{s_{xoxo}} & 0 \\ -\frac{\hat{p}}{s_{xoxo}} & \frac{1}{s_{xoxo}} & 0 \\ 0 & 0 & \frac{2\sigma^2}{N_o} \end{bmatrix} \\
\nabla^2 L_{\theta_y,\gamma_{1y}} &= \begin{bmatrix} -\frac{N_o N_m}{N} \\ -\frac{N_o N_m}{N}\hat{p} \\ 0 \end{bmatrix} \\
\nabla^3 L_{\theta_y,\theta_y,\gamma_{1y}} &= 0 \\
\nabla^3 L_{\theta_y,\theta_y,\beta_0} &= \begin{bmatrix} 0 & 0 & \frac{N_o}{(\sigma^2)^2} \\ 0 & 0 & \frac{N_o\hat{p}}{(\sigma^2)^2} \\ \frac{N_o}{(\sigma^2)^2} & \frac{N_o\hat{p}}{(\sigma^2)^2} & 0 \end{bmatrix} \\
\nabla^3 L_{\theta_y,\theta_y,\beta_1} &= \begin{bmatrix} 0 & 0 & \frac{N_o\hat{p}}{(\sigma^2)^2} \\ 0 & 0 & \frac{\sum_i^{N_o} x_{io}^2}{(\sigma^2)^2} \\ \frac{N_o\hat{p}}{(\sigma^2)^2} & \frac{\sum_i^{N_o} x_{io}^2}{(\sigma^2)^2} & 0 \end{bmatrix} \\
\nabla^3 L_{\theta_y,\theta_x,\gamma_{1y}} &= \begin{bmatrix} 0 \\ -\frac{N_o N_m}{N}\hat{\sigma_x}^2 \\ 0 \end{bmatrix} \\
ISNIL_y(\hat{\theta}_x) &= -\hat{\beta}_1 \frac{N_m}{N} \\
ISNIL_x(\hat{\theta}_x) &= -\frac{N_m}{N} \\
\nabla^3 L_{\theta_y,\gamma_0,\gamma_{1y}} &= \begin{bmatrix} -\frac{N_o N_m^2}{N^2} \\ -\frac{N_o N_m^2}{N^2}\hat{p} \\ 0 \end{bmatrix} \\
ISNIL_y(\hat{\gamma}_0) &= -\beta_0 - \beta_1\hat{p} \\
ISNIL_x(\hat{\gamma}_0) &= -\hat{p}
\end{aligned}
$$

$$\nabla^3 L_{\theta_y,\gamma_{1y},\gamma_{1y}} = \begin{bmatrix} -\frac{2N_oN_m^2}{N^2}(\hat{\beta}_0+\hat{\beta}_1\hat{p}) \\ -\frac{2N_oN_m^2}{N^2}(\hat{\beta}_0\hat{p}+\hat{\beta}_1\hat{p}^2) - \frac{2N_oN_m(N_m-N_o)}{N^2}\hat{\sigma_x}^2\hat{\beta_1} \\ \frac{N_oN_m(N_o-N_m)}{N_2} \end{bmatrix}$$

$$\nabla^3 L_{\theta_y,\gamma_{1y},\gamma_{1x}} = \begin{bmatrix} -\frac{N_oN_m^2}{N^2}\hat{p} \\ -\frac{N_oN_m^2}{N^2}\hat{p}^2 - \frac{N_oN_m(N_m-N_o)}{N^2}\hat{\sigma}_x^2 \\ 0 \end{bmatrix},$$

Where $\hat{\sigma}_x^2 = \frac{\exp(\hat{\alpha}_0)}{(1+\exp(\hat{\alpha}_0))^2}$, $\hat{p} = \frac{\exp(\hat{\alpha}_0)}{1+\exp(\hat{\alpha}_0)}$ and $s_{xoxo} = \sum_{i:g_i=1}(x_i - \hat{p})^2$

3.5 Index Calibration

3.5.1 Y-dependent Nonignorability

The ISNI value depends on the scale of the outcome Y and the scale of 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 $\gamma_1 = (\gamma_{1y})$, which effectively sets $\gamma_{1x} = \mathbf{0}$. 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 jth element of θ) we have $\hat{\theta}_j(\gamma_1) - \hat{\theta}_j(0) \approx \gamma_1 \text{ISNIL}_y(\hat{\theta}_j) + \frac{\gamma_1^2}{2}\text{ISNIQ}_{yy}(\hat{\theta}_j)$.

From the above equation, we compute the smallest value of γ_1, denoted as $\widetilde{\gamma}_1$, that causes the right hand side of the above equation to be same as one standard error. To put the $\widetilde{\gamma}_1$ value in the scale of standardized magnitude of nonignorability defined above, we define a sensitivity transformation c statistic as $c = |\widetilde{\gamma}_1 * \sigma_Y|$. 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 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 (6), we suggest using $c = 1$ as a general cutoff value for important sensitivity. Note that when the ISNIQ=0, we have $c = \left|\frac{\text{S.E.} * \sigma_Y}{\text{ISNI}}\right|$ which reduces to the c statistics defined in (6) for the case of linear ISNI only.

3.5.2 Y-and-X-dependent Nonignorability

We next consider the situation of y-and-x-dependent nonignorability where $\gamma_1 = (\gamma_{11}, \cdots \gamma_{1P}, \gamma_{1y})$. For continuous outcome and covariates, one can hypothetically standardize these continuous variables to all have a SD of one, and consider the missing data model $\text{Prob}(G_i = 1) = h(s_i^T \gamma_0 + (y_i^*, x_i^*)^T \gamma_1^*)$, where $(x_i^*, y_i^*) = (\frac{x_{1i}}{\sigma_{x_1}}, \cdots, \frac{x_{Pi}}{\sigma_{x_P}}, \frac{y_i}{\sigma_Y})$. Extending the idea for the above univariate nonignorability case to the multi-dimensional 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

$$\|\gamma_1^*\| \quad = \quad \sqrt{\gamma_{11}^2\sigma_{x_1}^2 + \cdots + \gamma_{1P}^2\sigma_{x_P}^2 + \gamma_{1y}^2\sigma_y^2}$$

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 (maximum - minimum) of changes in estimates achievable in the hyperball as

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

Note that by setting $\gamma_{1x} = (\gamma_{11}, \cdots, \gamma_{1P}) = 0$ 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, obtaining the range, in general, 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 $\|\gamma_1^*\|$ for which the maximal absolute change in the parameter estimate $\hat{\theta}_j$ is the same as its standard error.

3.6 Simulation Studies

3.6.1 Simple Linear Regression with Both Outcome and Covariates Following Normal Distribution

To illustrate the use of the nonlinear ISNI and demonstrate its superiority, we perform another simulation study for simple linear regression. The ideal outcomes are simulated from a simple linear regression model: $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma_y^2)$ where $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_y^2 = 1$, the covariate $x_i \sim N(\mu_x, \sigma_x^2)$, where $\mu_x = 0$, $\sigma_x = 1$, and $i = 1, \cdots, 1000$. The response behavior follows a logistic regression model: $\mathtt{logit}(\mu_{ij}^G) = \gamma_0 + \gamma_{1y} y_i + \gamma_{1x} x_i$, where the binary indicator $G = 1$ (responded) and $G = 0$ (non-responded 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, varying degrees, and types of nonignorable missingness.

We first consider Y-dependent-only nonignorable missingness, generating twelve random samples with $\gamma_{1x} = 0$, $\gamma_0 = 1, 2,$ or $5, \gamma_{1y} = 1, 0.5, 0.5,$ or 1. Figure 3 shows the analysis based on one 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 curves of $\hat{\beta}_1(\gamma_{1y})$, which were computed by maximizing the the log-likelihood of the joint nonignorable selection models for a range of values for γ_{1y} with γ_{1x} fixed at zero. This is computationally intensive and is not required in ISNI analysis. We compute the ISNI-based approximate sensitivity curves as

$$\begin{aligned}\widehat{\beta_1}^{\text{ISNIL}}(\gamma_{1y}) &= \widehat{\beta_1}(0) + \text{ISNIL}_y * \gamma_{1y} \\ \widehat{\beta_1}^{\text{ISNIQ}}(\gamma_{1y}) &= \widehat{\beta_1}(0) + \text{ISNIL}_y * \gamma_{1y} + \frac{\text{ISNIQ}_{yy}}{2}\gamma_{1y}^2 \end{aligned} \tag{3.5}$$

Figure 3 demonstrates the superiority of the higher-order ISNI approximation (broken line) over the first-order ISNI approximation (dotted line). In Figure 3, 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.

We summarize the simulation results in Table V. The computation of ISNI-adjusted estimates only requires knowing ISNIL_y 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 V 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 (e.g., $5e$-16 and 0.292, respectively) are very close to the ISNIL and ISNIQ values (e.g. 0 and 0.290, respectively), which demonstrates 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.

We next consider the outcome-and-covariate-dependent nonignorable nonresponse, generating twenty-four random datasets with $\gamma_0 = 1, 2$ or 5, and (a) $\gamma_{1x} = \sqrt{0.5}$ or $-\sqrt{0.5}, \gamma_{1y} = \sqrt{0.5}$ or $-\sqrt{0.5}$ or (b) $\gamma_{1x} = \sqrt{0.25}$ or $-\sqrt{0.25}, \gamma_{1y} = \sqrt{0.25}$ or $-\sqrt{0.25}$. In (a) and (b), we have $\|\gamma_1\| = 1$ and 0.5, respectively, the same as that in the above $Y-$dependent-only case. We use the following equations to compute the ISNI-adjusted estimates.

$$
\begin{aligned}
\hat{\beta_1}^{\text{ISNIL}}(\gamma_1) &= \hat{\beta_1}(0) + \text{ISNIL}^T * \gamma_1, \\
\hat{\beta_1}^{\text{ISNIQ}}(\gamma_1) &= \hat{\beta_1}(0) + \text{ISNIL}^T * \gamma_1 + \gamma_1^T \frac{\text{ISNIQ}}{2} \gamma_1
\end{aligned} \tag{3.6}
$$

Where $\gamma_1 = (\gamma_{1x}, \gamma_{1y})$.

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 Equation 3.1 for the joint nonignorable selection models at a grid of values for $(\gamma_{1x}, \gamma_{1y})$. 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. It is clear that the surface formed by $\hat{\beta_1}^{\text{ISNIQ}}$ provides much better approximation to the sensitivity surface of $\hat{\beta}_1(\gamma_1)$ than that using $\hat{\beta_1}^{\text{ISNIL}}$.

We then summarize the simulation results in Table VI. In the simulated dataset, the numerically evaluated derivatives are $(-5e{-}06, 7e{-}05)$ 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. 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 V: AN APPLICATION OF NISNI TO SIMULATED DATA FOR SIMPLE LINEAR REGRESSION WITH BOTH OUTCOME AND COVARIATES FOLLOWING NORMAL DISTRIBUTION FOR Y-DEPENDENT-ONLY NONIGNORABLE NONRESPONSE

γ_0	γ_{1y}	Prop_M	$\hat{\beta}_1(0)$(S.E.)	$\frac{\Delta\hat{\beta}_1}{\Delta\gamma_{1y}}$	$\frac{\Delta^2\hat{\beta}_1}{\Delta\gamma_{1y}^2}$	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	0.010	5.65

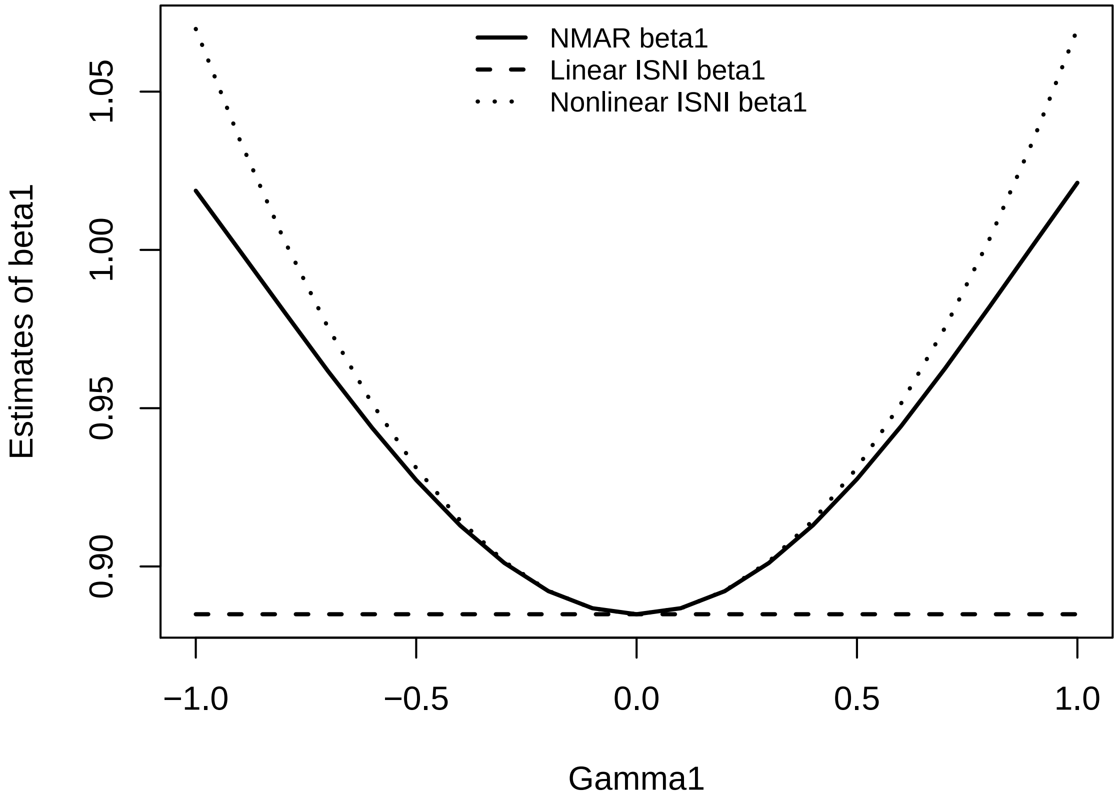

Figure 3: ISNI approximation to β_1 estimates for simple linear regression model with both missing continuous outcome and covariates for Y-dependent-only nonignorable nonresponse

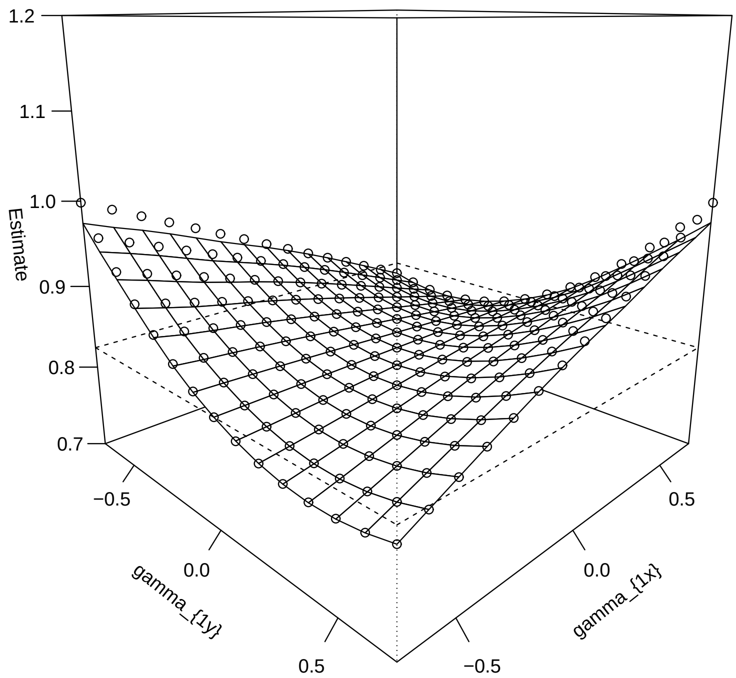

Figure 4: ISNI approximation to β_1 estimates for simple linear regression model with both outcome and covariates following normal distribution for Y-and-X-dependent nonignorable nonresponse

TABLE VI: AN APPLICATION OF NISNI TO SIMULATED DATA FOR SIMPLE LINEAR REGRESSION WITH BOTH OUTCOME AND COVARIATES FOLLOWING NORMAL DISTRIBUTION FOR Y-AND-X-DEPENDENT NONIGNORABLE NONRESPONSE

γ_0	γ_{1y}	γ_{1x}	Prop_M	$\hat{\beta}_1(0)$(S.E.)	$\frac{\Delta\hat{\beta}_1}{\Delta\gamma_{1y}}$	$\frac{\Delta^2\hat{\beta}_1}{\Delta\gamma_{1y}^2}$	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
	$\sqrt{0.25}$	$\sqrt{0.25}$	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
	$-\sqrt{0.25}$	$-\sqrt{0.25}$	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
	$\sqrt{0.5}$	$-\sqrt{0.5}$	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
	$\sqrt{0.25}$	$-\sqrt{0.25}$	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
	$-\sqrt{0.25}$	$\sqrt{0.25}$	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
	$\sqrt{0.25}$	$\sqrt{0.25}$	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
	$-\sqrt{0.5}$	$-\sqrt{0.5}$	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
	$-\sqrt{0.25}$	$-\sqrt{0.25}$	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
	$\sqrt{0.5}$	$-\sqrt{0.5}$	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
	$\sqrt{0.25}$	$-\sqrt{0.25}$	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
	$-\sqrt{0.25}$	$\sqrt{0.25}$	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
5	$\sqrt{0.5}$	$\sqrt{0.5}$	1.6%	0.98 (0.03)	(-6e-06, 5e-06)	(0.032, 0.016, 2e-03)	(0, 0)	(0.027, 0.014, 0)	2.10
	$\sqrt{0.25}$	$\sqrt{0.25}$	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
	$-\sqrt{0.5}$	$-\sqrt{0.5}$	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
	$-\sqrt{0.25}$	$-\sqrt{0.25}$	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
	$\sqrt{0.25}$	$-\sqrt{0.25}$	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
	$-\sqrt{0.25}$	$\sqrt{0.25}$	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

3.6.2 Simple Linear Regression with a Normally Distributed Outcome and a Bernoulli Distributed Covariate

Following the same process above, we perform a simulation study for simple linear regression with a normally distributed outcome and a Bernoulli distributed covariate. The ideal outcomes are simulated from a simple linear regression model: $Y_i \sim N(\beta_0+\beta_1 x_i, \sigma_y^2)$ where $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_y^2 = 1$, the covariate $x_i \sim B(1, p)$, where $p = 1/(1+\exp(-\alpha_0))$ and $\alpha_0 = 1$, and $i = 1, \cdots, 1000$. The response behavior follows a logistic regression model: $\mathrm{logit}(\mu_{ij}^G) = \gamma_0+\gamma_{1y} y_i+\gamma_{1x} x_i$, where the binary indicator $G = 1$ (responded) and $G = 0$ (nonresponded with x and y both missing). $\gamma_0 = 0, 1, 2,$ or 3, $\gamma_1 = -1, -0.5, 0.5,$ or 1. The values of γ_0 and γ_1 are varied so that the resulting datasets have varying amount of missingness ranging from ~5% to ~50% and varying degrees of nonignorable missingness.

We use the equation defined in Equation (Equation 3.5) to compute the ISNI-based approximate sensitivity curves for Y dependent-only Nonignorable Nonresponse. Figure 5 shows the analysis based on one simulated dataset with $\gamma_0 = 0$ and $\gamma_1 = -1$ and with the proportion of missingness being 0.48. We use the equations defined in Equation (Equation 3.6) to compute the ISNI-adjusted estimates for Y-and-X-dependent Nonignorable Nonresponse. Figure 6 plots the two ISNI-based approximate sensitivity surfaces with $\gamma_0 = 0$ and $\gamma_1 = (\sqrt{0.5}, \sqrt{0.5})$ and with the proportion of missingness being 0.36.

We then summarize the simulation results in Table VII for Y dependent-only Nonignorable Nonresponse and Table VIII for Y-and-X-dependent Nonignorable Nonresponse. Again, they demonstrate that 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 VII: AN APPLICATION OF NISNI TO SIMULATED DATA FOR SIMPLE LINEAR REGRESSION WITH NORMAL DISTRIBUTED OUTCOME AND A BERNOULLI COVARIATE FOR Y-DEPENDENT-ONLY NONIGNORABLE NONRESPONSE

γ_0	γ_{1y}	Prop_M	$\hat{\beta}_1(0)$(S.E.)	$\frac{\Delta\hat{\beta}_1}{\Delta\gamma_{1y}}$	$\frac{\Delta^2\hat{\beta}_1}{\Delta\gamma_{1y}^2}$	ISNIL	ISNIQ	c
1	1	21.1%	0.97 (0.08)	0.0033	0.288	0	0.300	0.68
	0.5	22.7%	0.94 (0.09)	0.0032	0.329	0	0.342	0.73
	-1	44.5%	0.85 (0.09)	0.0007	0.388	0	0.395	0.62
	-0.5	34.4%	1.01 (0.09)	0.0027	0.474	0	0.493	0.65
2	1	10.1%	0.88 (0.08)	0.0023	0.156	0	0.161	0.99
	0.5	11.2%	0.95 (0.08)	0.0029	0.195	0	0.202	0.95
	-1	26.4%	0.92 (0.08)	0.0022	0.337	0	0.348	0.65
	-0.5	19.8%	0.98 (0.08)	0.0026	0.302	0	0.316	0.72
5	1	0.5%	0.99 (0.07)	0.0002	0.010	0	0.011	3.93
	0.5	0.5%	1.00 (0.07)	0.0002	0.011	0	0.011	3.91
	-1	3.0%	0.99 (0.07)	0.0010	0.058	0	0.059	1.62
	-0.5	1.5%	1.01 (0.07)	0.0006	0.031	0	0.031	2.32

Table V, Table VI, Table VII, and Table VIII summarize the simulation results for all simulated datasets for both outcome and covariates following normal distributions and for outcome following a normal distribution and covariate following a Bernoulli distribution. 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. Fur-

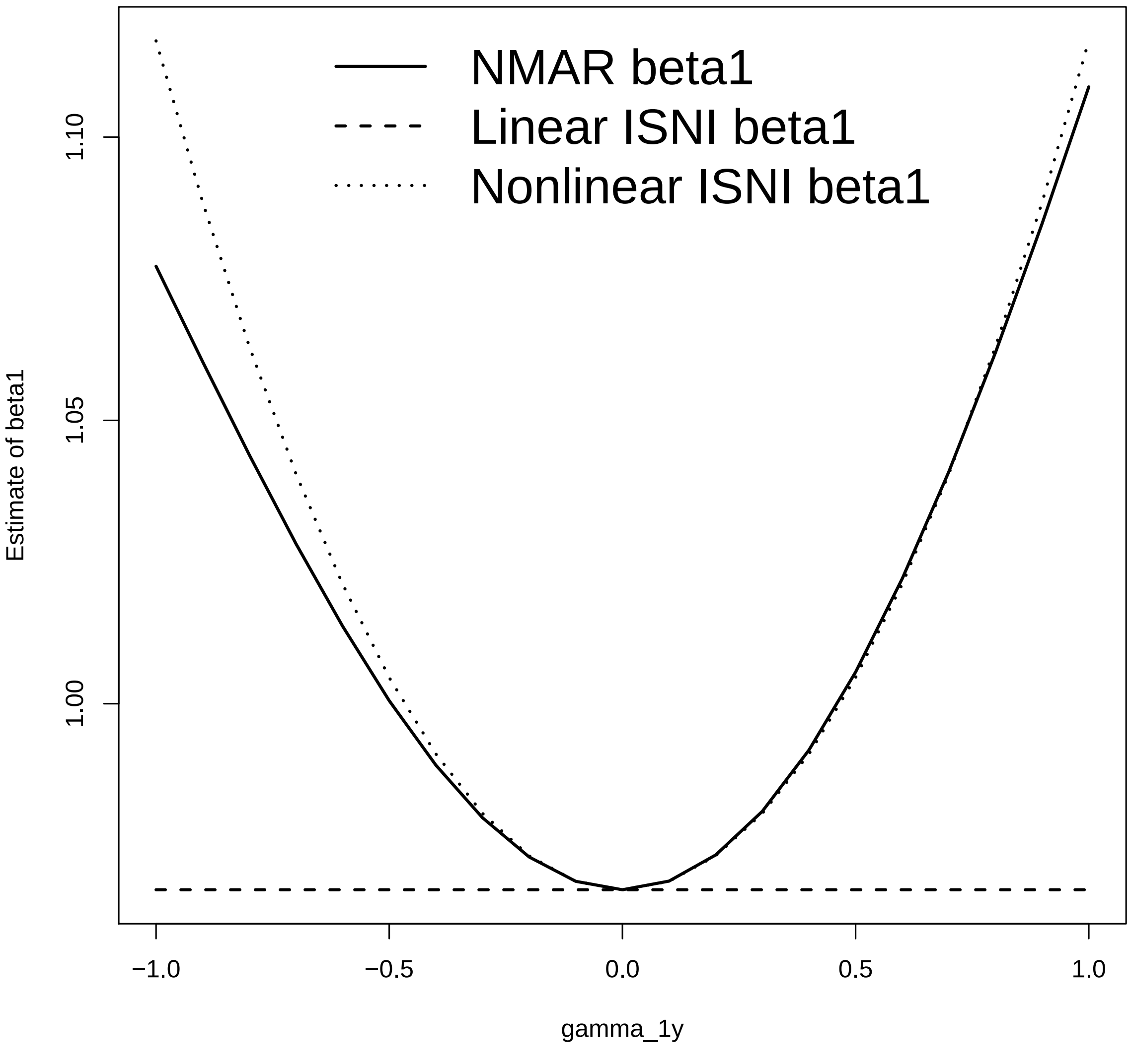

Figure 5: ISNI approximation to β_1 estimates for simple linear regression model with normal distributed outcome and a Bernoulli distributed covariate for Y-dependent-only Nonignorable Nonresponse

TABLE VIII: AN APPLICATION OF NISNI TO SIMULATED DATA FOR SIMPLE LINEAR REGRESSION WITH NORMAL DISTRIBUTED OUTCOME AND A BERNOULLI COVARIATE FOR Y-AND-X-DEPENDENT NONIGNORABLE NONRESPONSE

γ_0	γ_{1y}	γ_{1x}	Prop_M	$\hat{\beta}_1(0)$(S.E.)	$\frac{\Delta\hat{\beta}_1}{\Delta\gamma_{1y}}$	$\frac{\Delta^2\hat{\beta}_1}{\Delta\gamma_{1y}^2}$	ISNIL	ISNIQ	c
1	$\sqrt{0.5}$	$\sqrt{0.5}$	15.9%	0.90 (0.09)	(2e-04, 2e-04)	(0.252, 0.140, 0.000)	(0, 0)	(0.252, 0.141, 0)	0.62
	$\sqrt{0.25}$	$\sqrt{0.25}$	17.7%	0.91 (0.09)	(3e-04, 2e-04)	(0.276, 0.152, 0.000)	(0, 0)	(0.277, 0.153, 0)	0.60
	$-\sqrt{0.5}$	$-\sqrt{0.5}$	48.7%	0.80 (0.09)	(-7e-05, 2e-04)	(0.419, 0.258, 0.002)	(0, 0)	(0.413, 0.257, 0)	0.49
	$-\sqrt{0.25}$	$-\sqrt{0.25}$	43.1%	0.92 (0.09)	(7e-05, -2e-05)	(0.493, 0.268, -0.002)	(0, 0)	(0.496, 0.269, 0)	0.47
	$\sqrt{0.5}$	$-\sqrt{0.5}$	30.6%	1.06 (0.08)	(2e-04, 2e-04)	(0.422, 0.199, 0.000)	(0, 0)	(0.424, 0.200, 0)	0.48
	$\sqrt{0.25}$	$-\sqrt{0.25}$	30.1%	1.00 (0.09)	(2e-04, 2e-04)	(0.415, 0.208, 0.000)	(0, 0)	(0.417, 0.209, 0)	0.50
	$-\sqrt{0.5}$	$\sqrt{0.5}$	29.6%	1.00 (0.09)	(2e-04, 2e-04)	(0.423, 0.212, -0.003)	(0, 0)	(0.426, 0.213, 0)	0.50
	$-\sqrt{0.25}$	$\sqrt{0.25}$	30.0%	1.00 (0.09)	(2e-04, 1e-04)	(0.434, 0.216, 0.000)	(0, 0)	(0.434, 0.217, 0)	0.50
2	$\sqrt{0.5}$	$\sqrt{0.5}$	7.7%	0.90 (0.08)	(2e-04, 1e-04)	(0.132, 0.074, 0.001)	(0, 0)	(0.133, 0.074, 0)	0.81
	$\sqrt{0.25}$	$\sqrt{0.25}$	8.7%	0.94 (0.08)	(2e-04, 1e-04)	(0.158, 0.084, 0.000)	(0, 0)	(0.159, 0.084, 0)	0.76
	$-\sqrt{0.5}$	$-\sqrt{0.5}$	29.8%	0.92 (0.08)	(2e-04, 1e-04)	(0.398, 0.216, -0.001)	(0, 0)	(0.399, 0.217, 0)	0.49
	$-\sqrt{0.25}$	$-\sqrt{0.25}$	26.0%	0.95 (0.08)	(2e-04, 1e-04)	(0.376, 0.198, 0.000)	(0, 0)	(0.377, 0.199, 0)	0.50
	$\sqrt{0.5}$	$-\sqrt{0.5}$	15.0%	0.96 (0.08)	(2e-04, 2e-04)	(0.248, 0.129, 0.000)	(0, 0)	(0.248, 0.113, 0)	0.61
	$\sqrt{0.25}$	$-\sqrt{0.25}$	14.1%	0.98 (0.08)	(2e-04, 2e-04)	(0.251, 0.128, 0.000)	(0, 0)	(0.251, 0.128, 0)	0.61
	$-\sqrt{0.5}$	$\sqrt{0.5}$	15.1%	0.99 (0.08)	(2e-04, 1e-04)	(0.254, 0.128, -0.001)	(0, 0)	(0.253, 0.128, 0)	0.60
	$-\sqrt{0.25}$	$\sqrt{0.25}$	14.8%	1.00 (0.08)	(2e-04, 2e-04)	(0.253, 0.126, 0.000)	(0, 0)	(0.254, 0.126, 0)	0.61
5	$\sqrt{0.5}$	$\sqrt{0.5}$	0.3%	1.00 (0.07)	(1e-06, -4e-06)	(0.006, 0.003, 0.000)	(0, 0)	(0.006, 0.003, 0)	3.63
	$\sqrt{0.25}$	$\sqrt{0.25}$	0.5%	1.00 (0.07)	(3e-05, 9e-06)	(0.010, 0.005, 0.000)	(0, 0)	(0.011, 0.005, 0)	2.81
	$-\sqrt{0.5}$	$-\sqrt{0.5}$	3.3%	0.99 (0.07)	(8e-05, 6e-05)	(0.064, 0.033, 0.000)	(0, 0)	(0.065, 0.033, 0)	1.12
	$-\sqrt{0.25}$	$-\sqrt{0.25}$	2.2%	1.00 (0.07)	(5e-05, 3e-05)	(0.046, 0.023, 0.000)	(0, 0)	(0.045, 0.022, 0)	1.36
	$\sqrt{0.5}$	$-\sqrt{0.5}$	0.6%	1.00 (0.07)	(1e-05, 1e-05)	(0.012, 0.006, 0.000)	(0, 0)	(0.013, 0.006, 0)	2.57
	$\sqrt{0.25}$	$-\sqrt{0.25}$	0.7%	1.00 (0.07)	(4e-05, 2e-05)	(0.014, 0.008, 0.000)	(0, 0)	(0.015, 0.007, 0)	2.38
	$-\sqrt{0.5}$	$\sqrt{0.5}$	1.0%	1.02 (0.07)	(3e-05, 3e-05)	(0.022, 0.010, 0.0003)	(0, 0)	(0.021, 0.010, 0)	1.99
	$-\sqrt{0.25}$	$\sqrt{0.25}$	1.1%	1.01 (0.07)	(4e-05, 3e-05)	(0.023, 0.012, 0.000)	(0, 0)	(0.023, 0.012, 0)	1.90

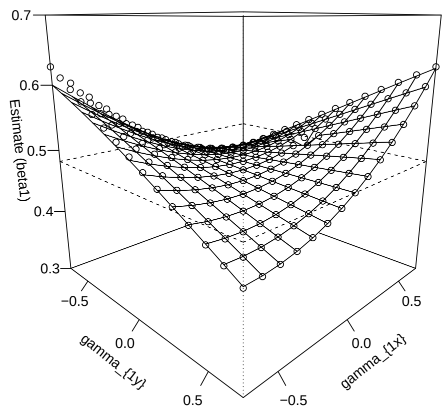

Figure 6: ISNI approximation to β_1 estimates for simple linear regression model with normal distributed outcome and a bernoulli distributed covariate for Y-and-X-dependent nonignorable nonresponse

thermore, 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.

CHAPTER 4

APPLICATONS

(Previously published as Gao, W., Hedeker, D., Mermelstein, R., Xie, H. (2016). A scalable approach to measuring the impact of nonignorable nonresponse with an EMA application. Statistics in Medicine. DOI: 10.1002/sim.7078.)

4.1 Example 1: Crossover in a Clinical Trial of Multiple Sclerosis

The primary endpoint for this multiple sclerosis (MS) trial is AD25, an assay of immune function defined as the antibody-dependent cellular cytotoxicity at an effector:target ratio of 25:1. There were no crossovers from placebo to treatment arm because subjects were not allowed to switch to active treatment. However, as shown in Table IX, there were 3 out of 11 subjects randomized to the treatment arm who later self-switched to the placebo arm. Because these 3 crossovers may differ from the others in their AD25 values, the remaining 8 subjects form a potentially biased representation of the original group randomized to the treatment arm. For illustrative purposes, in the analysis below we focus on the analysis of data from the treatment group only.

Models of the nonlinear ISNI (NISNI) formula as described in Section 2.3.1 are applied to evaluate the potential impact of nonignorable selective crossovers on the estimation of the AD25 distribution in the treatment arm, with results presented in TTable X. The distributional features considered in the analysis include mean, variance, inter-quartile range (IQR) and the tail percentiles. The rationale for considering measures other than the mean parameter is obvious. In clinical trials, researchers may also be interested in understanding the treatment effect on the spread of the outcome, in addition to the center of the distribution. Furthermore, investigators may be interested in percentile comparisons (e.g., as a way to investigate treatment effect heterogeneity). Therefore, it is practically relevant to investigate the impact of nonignorability on the broader set of distributional parameters, which is the aim of our ISNI analysis. As shown in Table X, ISNIQ is zero for $\hat{\mu}(0)$, the MAR estimate of the AD25 mean in the treatment arm. As a result, both the linear ISNI and nonlinear ISNI analyses have a c statistic of 1.3, which is slightly above 1. Although the higher-order nonlinear index does not affect the sensitivity assessment of the mean as compared with the linear index, it does affect the estimates of the variance $\hat{\sigma}(0)$. Specifically, the linear index ISNIL($\hat{\sigma}(0)$) = 0 and consequently c = 1, suggesting no local sensitivity for the variance MAR estimate when using the linear index alone. However, in fact, there is a significant nonlinear component because ISNIQ($\hat{\sigma}(0)$) = 25188 with c = 1.6, a substantial difference from c = 1 when using the linear index alone. The c statistic of 1.6 exceeds the cutoff value of 1 for important selection bias relative to the size of sampling error. Furthermore, the c value could fall below 1 in a different sample (e.g. with a larger sample size and thus a smaller standard error) while as shown in Section 2.3.1, the c statistic will always be 1 if using the linear index alone and thus won't be able to detect sensitivity when it exists.

The nonlinear impact of nonignorability can also occur in respect to other higher-order distributional features aside from the mean. The examples of these features include percentiles

Q_α and IQR, which in the univariate normal case are $\mu + p_\alpha \sigma$ and 1.35σ, respectively, where α is the tail probability and p_α is the α-percentile for the standard normal distribution. The NISNI values for these parameters can be computed as follows. Define a continuously differentiable scalar function $f(\theta)$. According to the chain rule of differentiation, we have

$$\begin{aligned} \text{ISNIL}(f(\hat{\theta})) &= \frac{\partial f(\hat{\theta})}{\partial \theta} \frac{\partial \theta^T}{\partial \gamma_1} = f'(\hat{\theta}).\text{ISNIL}(\hat{\theta}), \\ \text{ISNIQ}(f(\hat{\theta})) &= f''(\hat{\theta}).\text{ISNIL}^2(\hat{\theta}) + f'(\hat{\theta}).\text{ISNIQ}(\hat{\theta}), \end{aligned}$$

Where $f'(.)$ and $f''(.)$ denote the first and second derivatives of the function $f(.)$, respectively. Consider a special case where $f(\theta) = \alpha^T\theta$, i.e. a linear combination of model parameters where α is a vector of constant. Then we have

$$\begin{aligned} \text{ISNIL}(\alpha^T\hat{\theta}) &= \alpha^T\text{ISNIL}(\hat{\theta}), \\ \text{ISNIQ}(\alpha^T\hat{\theta}) &= \alpha^T\text{ISNIQ}(\hat{\theta}), \end{aligned}$$

Applying the above results to Q_α and IQR, which are functions of the model parameter θ, we obtain their ISNIL and ISNIQ values as listed in Table X. We note again that when using the linear ISNI alone, we either cannot detect sensitivity ($c = 1$ for IQR) or considerably underestimate the sensitivity ($c = 2.24$ for using ISNIL alone .vs. $c = 1.3$ for additionally using ISNIQ for $Q_{97.5}$).

TABLE IX: THE AD25 VALUES IN THE TREATMENT ARM IN THE MS CLINICAL TRIAL

Subjid	1	2	3	4	5	6	7	8	9	10	11
AD25	2	3	3	3	21	25	27	49	*	*	*

TABLE X: SENSITIVITY ANALYSIS IN THE MS DATA

Parameter	MAR Est	SE	ISNIL	ISNIQ	Linear Correction		Nonlinear Correction	
					Est ($\gamma_1 = \pm\frac{1}{\sigma_y}$)	c	Est ($\gamma_1 = \pm\frac{1}{\sigma_y}$)	c
μ	16.65	5.61	-68.72	0	[12.30, 20.95]	1.30	[12.30, 20.95]	1.30
σ^2	251.98	126.00	0	25188.55	251.98	∞	[251.98, 301.96]	1.58
$Q_{97.5}$	47.77	9.71	-68.72	1555.44	[43.34, 52.1]	2.24	[46.53, 55.19]	1.31
IQR	21.42	5.34	0	1071.35	21.42	∞	[21.42, 23.56]	2.50

4.2 Example 2: Nonresponses in EMA Studies

Our application considers the nonresponse issue in Ecological Momentary Assessment (EMA) studies. As a real-time data capturing method, EMA has become increasingly important in health studies (2). 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, in (41), the studying question is "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 (smart phones). 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. An important feature of the EMA data is the more complex missing data pattern, where response and covariates are subject to simultaneous missingness. 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 Chapter 2. Therefore, we apply our new nonlinear index method developed in Chapter 3 to quantify the potential bias from such missing data in EMA studies.

According to (41), "data comes from a longitudinal study of the natural history of smoking among adolescents" (PO1CA098262, PI R. Mermelstein). The Electronic Diary study involved sampling novice smokers as 9th and 10th graders; 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. (41) has described "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 and nonsmoking 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). Thus, we will be able to compare the subjective and objective contexts surrounding the smoke and random times.

The smoking-mood relationship has been well studied among adults, in particular those trying to stop smoking, but much less is known about how mood is associated with smoking among adolescents. Understanding more about the mood-smoking relationship in adolescents can help to identify who is most at risk for smoking escalation and developing nicotine dependency, as well as providing insights into intervention development. Therefore, we will conduct an analysis to investigate the smoking-mood relationship among adolescents. As compared with more traditional data collection methods, "EMA provides an excellent window into the lives of adolescents"(41). In smoking-related studies, "EMA captures subtle variations in mood as they occur, and can do so more accurately than other measurement modalities" (41) and (42). "With the use of random assessments that are independent from the occurrence of specific events, such as smoking, EMA can provide useful comparison information about background moods, and allow us to address several critical questions about mood-smoking relationships among adolescents, including the following: 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 moderator variables may also help in the prediction of smoking escalation among relatively novice smokers"(41).

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.

In the EMA data, the outcome Mood is the negative mood prior to prompt signal or smoking event, with a higher value meaning a worse mood. Mood is the average of a subject's evaluation of the following five items before the prompt signal: 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. As a result, its average value is considered a nearly continuous value. The primary covariate is smoking status (1=smoking event vs 0=random prompts). There are two mediators/confounders we are interested in. The first is Social, which denotes a measurement for social isolation prior to prompt signal or smoking event and is the average of a subject's evaluation of the following three items before the prompt signal: I felt ignored, I felt left out, I felt lonely, where each item is on a rating scale from 1 to 10 with a higher value meaning worse. It takes nearly continuous values. The second is Comp, which denotes dichotomous measurement of companionship prior to prompt signal or smoking event. It is coded as 0=alone and 1=alone-others nearby/with others. Our analysis scheme is firstly to conduct simple linear regression model to examine whether the mediator affects mood in random prompts, and then to conduct multiple linear regression models to examine the mood-smoking relationship in young adults after controlling the mediator in the model.

4.2.1 Analysis 1: Considering Social as Mediator

The simple linear regression model to examine whether the social isolation affects mood in random prompts is as following

$$\mathsf{Mood}_i = (\mathsf{Intercept}, \mathsf{Social})_i^T \beta + \epsilon_i,$$

Missing data arises because of nonresponses to the random prompts, which leads to the mood outcome being missing, as well as Social associated with the missed random prompts. They are missing simultaneously. However, if mood affects adolescents' compliance with answering data collection prompts, then our conclusions about the mood-social isolation relationship may be biased. 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: logit$\mathsf{P}(\mathsf{G}_i = 1) = \gamma_0 + \gamma_{1x} * \mathtt{social}_i + \gamma_{1y}\mathtt{mood}_i$, for the mood-social isolation model. We assume the Social follows univariate normal distribution.

We first conducted the MAR analysis and then conducted both linear and nonlinear ISNI analyses, assuming Mood-dependent-only nonignorability by fixing $\gamma_{1x} = 0$ in the above prompt nonresponse model. The MAR analysis results for the mood-social isolation regression model are summarized in Table XI and show that social isolation is associated with higher negative mood at random background times, and this relationship is statistically significant with a p-value $< .0001$. These MAR estimates are potentially biased with nonignorable nonresponse to random prompts. 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 apply our new proposed NISNI method 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 XI. The c statistic for the Social using ISNIQ is 0.96, while it is ∞ when using the ISNIL alone. So the c statistic values indicate 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 the impact of nonignorability properly.

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 XII. 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).

TABLE XI: NISNI ANALYSIS OF SOCIAL ISOLATION AND MOOD RELATIONSHIP USING EMA DATA ASSUMING Y-DEPENDENT-ONLY NONIGNORABLE NONRESPONSE

Parameter	MAR Est	SE	ISNILy	ISNIQyy	Linear Correction		Nonlinear Correction	
					Est ($\gamma_{1y} = \pm\frac{1}{\sigma_y}$)	c	Est ($\gamma_{1y} = \pm\frac{1}{\sigma_y}$)	c
Intercept	1.461	0.148	-0.648	-1.780	[1.068, 1.861]	0.599	[0.740, 1.579]	0.479
Social	0.747	0.049	0	0.732	0.747	∞	[0.747, 0.886]	0.960

TABLE XII: NISNI ANALYSIS OF SOCIAL ISOLATION AND MOOD RELATIONSHIP USING EMA DATA ASSUMING Y-AND-X-DEPENDENT NONIGNORABLE NONRESPONSE

Parameter	MAR Est	SE	ISNIL	ISNIQ	Linear Correction		Nonlinear Correction	
					Est ($\gamma_{1y} = \pm\frac{1}{\sigma_y}$)	c	Est ($\gamma_{1y} = \lVert\gamma_1^*\rVert$)	c
Intercept	1.461	0.148	(-0.648, 0)	(-1.779, -1.191, 0)	[1.061, 1.861]	0.599	[0.654, 1.779]	0.475
Social	0.747	0.049	(0, 0)	(0.732, 0.490 0)	0.747	∞	[0.708, 0.926]	0.525

Next we investigate the mood-smoking relationship in young adults with the covariate Social in the model. This covariate can be a mediator to mediate the mood-smoking relationship or be a confounder to confounding contextual variable that simultaneously affects mood and smoking behavior but not in the mood-smoking causal pathway. The following linear regression models for the ideal outcome are considered:

$$\mathsf{Mood}_i \quad = \quad (\mathsf{Intercept}, \mathsf{SmkEvent}, \mathsf{Social})_i^T \beta + \epsilon_i$$

Where SmkEvent denotes a variable indicating whether the observation is from a random prompt(=0) or a smoking event(=1). Upon this analysis, we will examine wheter moods prior to smoking are different than the moods during random prompts. For this purpose, the regression coefficient for SmkEvent compares the mood outcome prior to smoking with that at random background times controlling Social in the model.

To investigate the potential impact of nonignorable nonresponse, we assume the following prompt nonresponse model: logit$P(G_i = 1) = s_i\gamma_0 + \gamma_{1x} * \mathsf{social}_i + \gamma_{1y}\mathsf{mood}_i$, where $s_i = (\mathsf{Intercept}, \mathsf{SmkEvent})_i$. The ideal data on the covariate Social is assumed to have a simple linear regression model, which is $\mathsf{Social}_i = (\mathsf{Intercept}, \mathsf{SmkEvent})_i^T \delta + e_i$.

We first conducted the MAR analysis and then conducted both linear and nonlinear ISNI analyses, assuming Mood-dependently-only nonignorability by fixing $\gamma_{1x} = 0$ in the above prompt nonresponse model. The MAR analysis results for the Mood outcome regression model are summarized in Table XIII 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. The results on both the ISNIL and ISNIQ values, as well as the calibrated range of parameter changes and c statistics in Table XIII, are used to gauge the sensitivity of these MAR estimates to nonignorable nonresponse, which avoids fitting any complicated nonignorable models and only requires readily-available MAR estimates. The column "Est($\gamma_{1y} = \pm\frac{1}{\sigma_y}$)" 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 nonignorability 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 nonignorability. 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. Again, it is critical to use nonlinear indices to properly measure the impact of nonignorability.

We further conduct a nonlinear ISNI analysis that permits Mood-and-Social-dependent nonignorability where the nonresponse probabilities are allowed to depend on both mood and social isolation, with results reported in Table XIV. 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).

TABLE XIII: NISNI ANALYSIS OF SMOKING-MOOD RELATIONSHIP WITH SOCIAL ISOLATION AS A COVARIATE USING EMA DATA ASSUMING Y-DEPENDENT-ONLY NONIGNORABLE NONRESPONSE

Parameter	MAR Est	SE	ISNIL_y	ISNIQ_{yy}	Linear Correction		Nonlinear Correction	
					Est ($\gamma_{1y} = \pm\frac{1}{\sigma_Y}$)	c	Est ($\gamma_{1y} = \pm\frac{1}{\sigma_Y}$)	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 XIV: NISNI ANALYSIS OF SMOKING-MOOD RELATIONSHIP WITH SOCIAL ISOLATION AS A COVARIATE USING EMA DATA ASSUMING Y-AND-X-DEPENDENT NONIGNORABLE NONRESPONSE

Parameter	MAR Est	SE	ISNIL	ISNIQ	Linear Correction		Nonlinear Correction	
					Est ($\gamma_{1y} = \pm\frac{1}{\sigma_Y}$)	c	Est ($\|\gamma_1^*\| = 1$)	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	∞	[0.738, 0.823]	0.88

4.2.2 Analysis 2: Considering Comp as Mediator

The simple linear regression model to examine whether the companionship status affects mood in random prompts is as following

$$\mathsf{Mood}_i \;=\; (\mathsf{Intercept}, \mathsf{Comp})_i^T \beta + \epsilon_i,$$

We assume the following prompt nonresponse model:

$$\text{logit}P(G_i = 1) = \gamma_0 + \gamma_{1x} * \mathsf{comp}_i + \gamma_{1y}\mathsf{mood}_i,$$

For the mood-companionship model. We also assume that Comp follows a bernoulli distribution.

Table XV summarized Comp-Mood relationship in random prompts assuming Y-Dependent-only nonignorable nonresponse. The MAR analysis results for the mood-companionship regression model showed that the companionship is not statistically significant associated with negative mood at random with a p-value of 0.79. The c statistic for the Comp using ISNIQ is 2.11, while it is ∞ when using the ISNIL alone. Hence, the c statistic values indicate that the linear ISNI analysis is unable to detect any impact of nonignorability on the coefficient of the covariate Comp. In this case, it is critical to use nonlinear indices to measure properly the impact of nonignorability. Table XVI summarized the Comp-Mood relationship in random prompts assuming Y-and-X Dependent nonignorable nonresponse. 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-Comp-dependency *vs.* Mood-dependency-only).

TABLE XV: NISNI ANALYSIS OF COMPANIONSHIP AND MOOD RELATIONSHIP USING EMA DATA ASSUMING Y-DEPENDENT-ONLY NONIGNORABLE NONRESPONSE

Parameter	MAR Est	SE	ISNILy	ISNIQyy	Linear Correction		Nonlinear Correction	
					Est ($\gamma_{1y} = \pm 1$)	c	Est ($\gamma_{1y} = \pm 1$)	c
Intercept	3.229	0.215	-1.098	-0.081	[2.131, 4.327]	0.196	[2.090, 4.286]	0.194
Companionship	0.069	0.254	0	0.114	0.069	∞	[0.069, 0.126]	2.111

TABLE XVI: NISNI ANALYSIS OF COMPANIONSHIP AND MOOD RELATIONSHIP USING EMA DATA ASSUMING Y-AND-X-DEPENDENT NONIGNORABLE NONRESPONSE

Parameter	MAR Est	SE	ISNIL	ISNIQ	Linear Correction		Nonlinear Correction	
					Est ($\gamma_{1y} = \pm 1$)	c	Est ($\gamma_{1y} = \lVert\gamma_1^*\rVert$)	c
Intercept	3.229	0.215	(-1.098, 0)	(-0.081, -0.591, 0)	[2.131, 4.327]	0.196	[2.641, 3.802]	0.406
Companionship	0.069	0.254	(0, 0)	(0.114, 0.829, 0)	0.069	∞	[-0.121, 0.272]	1.117

Next we will investigate the mood-smoking relationship in young adults with the covariates Comp in the model. The following linear regression models for the ideal outcome are considered:

$$\mathsf{Mood}_i \quad = \quad (\mathsf{Intercept}, \mathsf{SmkEvent}, \mathsf{Comp})_i^T \beta + \epsilon_i$$

The regression coefficient for SmkEvent compares the mood outcome prior to smoking with that at random background times controlling Comp in the model. To investigate the potential impact of nonignorable nonresponse, we assume the following prompt nonresponse model:

$$\text{logit}P(G_i = 1) = s_i\gamma_0 + \gamma_{1x}\mathsf{comp}_i + \gamma_{1y}\mathsf{mood}_i$$

where $s_i = (\mathsf{Intercept}, \mathsf{SmkEvent})_i$ and a logistic regression model for the ideal data on the covariate Comp is assumed as $\mathsf{logit}(\mathsf{Comp}_i = 1) = (\mathsf{Intercept}, \mathsf{SmkEvent})_i^T \delta + e_i$.

Table XVII summarized smoking-mood relationship with companionship as an covariate assuming Y-dependent-only nonignorable nonresponse by fixing $\gamma_{1x} = 0$ in the above prompt nonresponse model. The MAR analysis results show that companionship is not statistically significantly associated with higher negative mood at random background times with a p-value of 0.072. The c statistic for the Comp using ISNIQ is 3.33, while it is ∞ when using the ISNIL alone. Thus, the c statistic values indicate that the linear ISNI analysis is unable to detect any impact of nonignorability on the coefficient of the covariate Comp, whereas there are, in fact, significant changes in parameter estimates relative to the sampling error. Table XVIII summarized the smoking-mood relationship with companionship as an covariate, assuming Y-and-X-dependent nonignorable nonresponse in the above prompt nonresponse model. As compared with results assuming Mood−dependently-only nonignorability, the c statistics are somewhat smaller. However, the sensitivity analysis results remain qualitatively unchanged with respect to the type of nonignorability (Mood-and-Social-dependency *vs.* Mood-dependency-only).

TABLE XVII: NISNI ANALYSIS OF SMOKING-MOOD RELATIONSHIP WITH COMPANIONSHIP AS A COVARIATE USING EMA DATA ASSUMING Y-DEPENDENT-ONLY NONIGNORABLE NONRESPONSE

Parameter	MAR Est	SE	$\text{ISNIL}_{\mathsf{y}}$	$\text{ISNIQ}_{\mathsf{yy}}$	Linear Correction		Nonlinear Correction	
					Est ($\gamma_{1\mathsf{y}} = \pm 1$)	c	Est ($\gamma_{1\mathsf{y}} = \pm 1$)	c
Intercept	2.175	0.676	-2.203	0.031	[-0.027, 4.378]	0.31	[-0.012, 4.393]	0.31
SmkEvent	1.123	0.623	1.101	0.002	[0.022, 3.000]	0.57	[0.022, 2.225]	0.57
Companionship	-0.028	0.251	0	-0.045	-0.028	∞	[-0.051, -0.028]	3.33

TABLE XVIII: NISNI ANALYSIS OF SMOKING-MOOD RELATIONSHIP WITH COMPANIONSHIP AS A COVARIATE USING EMA DATA ASSUMING Y-AND-X-DEPENDENT NONIGNORABLE NONRESPONSE

Parameter	MAR Est	SE	ISNIL	ISNIQ	Linear Correction		Nonlinear Correction	
					Est ($\gamma_{1y} = \pm 1$)	c	Est ($\gamma_{1y} = \lVert\gamma_1^*\rVert$)	c
Intercept	2.175	0.676	(-2.203, 0)	0.031	[-0.027, 4.378]	0.31	[1.094, 3.264]	0.635
SmkEvent	1.123	0.623	(1.101, 0)	0.002	[0.022, 3.000]	0.57	[0.601, 1.348]	1.192
Companionship	-0.028	0.251	(0, 0)	-0.045	-0.028	∞	[-0.056, -0.028]	1.663

CHAPTER 5

DISCUSSION

An untestable *i*gnorable missingness assumption is often used in reality. Assessing the impact of nonignorability in standard analyses results is necessary so that researchers can judge when such nonignorable missingness may be a concern and require attention. Prior research has focused on the linear sensitivity index to measure the sensitivity to nonignorability. It demonstrates their usefulness in a range of important statistical applications. In this work, we relax this linearity assumption and developed more general nonlinear sensitivity index measures: 1) nonlinear sensitivity idex for missingness in the outcome only, and 2) nonlinear sensitivity index for missingness in both the outcome and covariates. In 2), we also developed a more complex nonignorability missing model, depending on the outcome only and depending on both the outcome and the covariates. These nonlinear index measures maintain the computational simplicity of the linear sensitivity index measures and avoid fitting complicated nonignorable models and thus are well suited for use in big data and data-rich environment nowadays. We identify situations where nonlinear sensitivity indices are most useful and can lead to qualitatively different conclusions regarding the impact of nonignorability on the MAR estimates. The proposed nonlinear sensitivity measures can effectively detect the impact of nonignorability comparing to the linear index measures in some important situations. These situations include when the parameters of interest are concerned with finer distributional features such as variance and tail percentiles, as well as when the outcome and covariates in a regression model are subject to simultaneous missingness (e.g. EMA studies).

In practical applications, our sensitivity index methods can be useful in the following ways. First, the investigators can use the indices to incorporate sensitivity analysis results into the primary reporting. Our sensitivity index methods can be used as a computationally feasible tool for this purpose. Tables in our application chapters provide examples of how our sensitivity analysis results can be incorporated into the primary reporting and be used to evaluate the impact of nonignorability on the parameter estimates relative to sampling errors. Second, our sensitivity index methods can be used for informing more efficient efforts to collect additional data. An investigator may choose to collect additional data 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 undoubtedly require additional resources and time in collecting data, as well as constructing and fitting more complicated models. If an investigator chooses this option, our method can be useful for quickly screening datasets in which the impact of nonignorable missingness is important, before deciding to invest a great deal of effort and valuable resources to collect additional data and perform arduous modeling. Our developed NISNI method not only can apply to modern electronic data capturing methods, such as the EMA methods, but also can be applied in the more traditional types of data to quantify the potential impact of nonignorability.

One potential complexity of applying the proposed index method occurs when the outcome does not follow Gaussian distribution. We need to further develop the closed-form formulas for other data distributions. The other limitation of our work is that we restrict our analysis to the cross-sectional data. The analysis does not exploit intensive longitudinal information within the multiple measurements 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 computational work, which we plan to present elsewhere.

APPENDICES

Appendix A

R CODE FOR ISNI ANALYSIS FOR EMA DATA WITH BOTH OUTCOME AND COVARIATES FOLLOWING NORMAL DISTRIBUTION

```
library (Rsolnp)

### Apply to real data (EMA) Simple Linear regression###
### for Both Outcome and Covariates Following Normal Distribution ###

## Read in the function.
source('C:\\Users\\gaowh\\Documents\\Thesis_2013\\Thesis_gao\\nisni R program\\nisniglm.R')

#read ema data

ema = read.csv("C:\\Users\\gaowh\\Documents\\Thesis_2013\\Hui Papers\\EMA data\\emabase.csv",
                header = TRUE)

#check missing data
sapply(ema, function(x) sum(is.na(x)))

###only analyze non-smoking subjects###
ema2 = ema[ema$smoking==1,]
attach(ema2)

## compute isnil and isniq
sim.isni <- glm(mood~sociso)
out<- summary(sim.isni)
coef <- cbind(round(out$coefficients[ , 1],3), round(out$coefficients[ , 2],3))

## get the residual variance
ig.sigma2 <- summary(sim.isni)$dispersion*(length(mood[g==1])-length(coef(sim.isni)))/length(mood[g==1])

## verify with explicit formula
n=length(mood)
```

Appendix A (Continued)

```
nm= sum(is.na(mood))
no=length(mood)-nm
xobar=mean(sociso, na.rm=T)
sigma.x2=(sd(sociso, na.rm=T))^2
sxoxo= sum((sociso[g==1]-xobar)^2)
sd.y = sqrt(ig.sigma2)
sd.x=sqrt(sigma.x2)


####################################################
### For Y-Dependent-only Nonignorable Nonresponse ###
####################################################
isnil.beta1  = 0
isniq.beta1  = round(2*ig.sigma2*((no^2*nm)/n^2)*sigma.x2*sim.isni$coef[2]*(1/sxoxo), 3)
isnil.beta0  = round(-ig.sigma2*(nm/n), 3)
isniq.beta0  = round(-isniq.beta1*xobar, 3)

isnil<-rbind(isnil.beta0, isnil.beta1)
isniq<-rbind(isniq.beta0, isniq.beta1)
isni<-cbind(isnil, isniq)

#calculate calibrated estimates
quad<- function(se, isnil, isniq, sdy){
if (isniq!=0){
if(isniq >0){
z <- matrix(c(-se,isnil,0.5*isniq), ncol=1)
          } else
          {
z <- matrix(c(-se,isnil,abs(0.5*isniq)), ncol=1)
          }
root<-polyroot(z)
mini<-min(c(Re(abs(root[1])), Re(abs(root[2]))))
            }else
            {
mini<-abs(se/isnil)
            }
res<- round(abs(mini*sdy),3)
}
c.l.beta0<-quad(coef[1,2],isnil.beta0,0, ig.sigma2)
c.q.beta0<-quad(coef[1,2],isnil.beta0,isniq.beta0, ig.sigma2)
```

Appendix A (Continued)

```
c.l.beta1<-quad(coef[2,2],isnil.beta1,0, ig.sigma2)
c.q.beta1<-quad(coef[2,2],isnil.beta1,isniq.beta1, ig.sigma2)



gamma1l=round(1/sd.y,3)
gamma1u=-round(1/sd.y,3)

gamma1 <- seq(gamma1u, gamma1l, 0.018)
beta0.isnill = min(round(sim.isni$coef[1] + isnil.beta0*gamma1,3))
beta0.isnilu = max(round(sim.isni$coef[1] + isnil.beta0*gamma1,3))
beta0.isniql = min(round(sim.isni$coef[1] + isnil.beta0*gamma1 + (isniq.beta0/2)*(gamma1^2), 3))
beta0.isniqu = max(round(sim.isni$coef[1] + isnil.beta0*gamma1 + (isniq.beta0/2)*(gamma1^2), 3))
r.beta0.l<- paste0("(", round(beta0.isnill,3), ", ", beta0.isnilu,")")
r.beta0.q<- paste0("(", round(beta0.isniql,3), ", ", beta0.isniqu,")")



beta1.isnill = min(round(sim.isni$coef[2] + isnil.beta1*gamma1,3))
beta1.isnilu = max(sim.isni$coef[2] + isnil.beta1*gamma1,3)
beta1.isniql = min(round(sim.isni$coef[2] + isnil.beta1*gamma1 + (isniq.beta1/2)*(gamma1^2), 3))
beta1.isniqu = max(round(sim.isni$coef[2] + isnil.beta1*gamma1 + (isniq.beta1/2)*(gamma1^2), 3))
r.beta1.l<- paste0("(", round(beta1.isnill,3), ", ", beta1.isnilu,")")
r.beta1.q<- paste0("(", round(beta1.isniql,3), ", ", beta1.isniqu,")")



isni.beta0 <- cbind(isnil.beta0, isniq.beta0, r.beta0.l, c.l.beta0, r.beta0.q, c.q.beta0)
isni.beta1 <- cbind(isnil.beta1, isniq.beta1, r.beta1.l, c.l.beta1, r.beta1.q, c.q.beta1)
isni.beta  <- rbind(isni.beta0, isni.beta1)

isni.nn.ydep<- cbind(coef, isni.beta)
colnames(isni.nn.ydep) <- c("MAR Est", "SE", "ISNILy", "ISNILyy", "Linear Correction_range",
                            "Linear Correction_c", "NonLinear Correction_range", "NonLinear Correction_c")
isni.nn.ydep



####################################################
### For Y-and-X-Dependent Nonignorable Nonresponse ###
####################################################
isnilx.beta0  = 0
isnily.beta0  = -ig.sigma2*(nm/n)
isnilx.beta1  = 0
```

Appendix A (Continued)

```
isnily.beta1  = 0

isniqxx.beta1  = 0
isniqxy.beta1  = ig.sigma2*((no*nm)/n^2)*((no*sigma.x2)/sxoxo)
isniqyy.beta1  = 2*sim.isni$coef[2]*ig.sigma2*((no*nm)/n^2)*((no*sigma.x2)/sxoxo)
isniqxx.beta0  = 0
isniqxy.beta0  = -isniqxy.beta1*xobar
isniqyy.beta0  = -isniqyy.beta1*xobar

isnil0<-paste0("(", round(isnily.beta0,3), ", ", isnilx.beta0,")")
isnil1<-paste0("(", round(isnily.beta1,3), ", ", isnilx.beta1,")")

isniq0<-paste0("(", round(isniqyy.beta0, 3), ", ", round(isniqxy.beta0, 3), ", ", isniqxx.beta0,")")
isniq1<-paste0("(", round(isniqyy.beta1, 3), ", ", round(isniqxy.beta1, 3), ", ", isniqxx.beta1,")")

#calculate calibrated estimates
#minimum value of range
fnbeta0.l.min=function(x)
{
isnilx.beta0*x[1] + isnily.beta0*x[2]
}
fnbeta0.q.min=function(x)
{
isnilx.beta0*x[1] + isnily.beta0*x[2] +
(1/2)*((x[1]*isniqxx.beta0 + x[2]*isniqxy.beta0)*x[1] +
(x[1]*isniqxy.beta0 + x[2]*isniqyy.beta0)*x[2])
}
fnbeta1.l.min=function(x)
{
isnilx.beta1*x[1] + isnily.beta1*x[2]
}
fnbeta1.q.min=function(x)
{
isnilx.beta1*x[1] + isnily.beta1*x[2] +
(1/2)*((x[1]*isniqxx.beta1 + x[2]*isniqxy.beta1)*x[1] +
(x[1]*isniqxy.beta1 + x[2]*isniqyy.beta1)*x[2])
}

#maximum value of range
fnbeta0.l.max=function(x)
```

Appendix A (Continued)

```
{
-(isnilx.beta0*x[1] + isnily.beta0*x[2])
}
fnbeta0.q.max=function(x)
{
-(isnilx.beta0*x[1] + isnily.beta0*x[2] +
(1/2)*((x[1]*isniqxx.beta0 + x[2]*isniqxy.beta0)*x[1] +
(x[1]*isniqxy.beta0 + x[2]*isniqyy.beta0)*x[2]))
}
fnbeta1.l.max=function(x)
{
-(isnilx.beta1*x[1] + isnily.beta1*x[2])
}
fnbeta1.q.max=function(x)
{
-(isnilx.beta1*x[1] + isnily.beta1*x[2] +
(1/2)*((x[1]*isniqxx.beta1 + x[2]*isniqxy.beta1)*x[1] +
(x[1]*isniqxy.beta1 + x[2]*isniqyy.beta1)*x[2]))
}

eqn1=function(x){
z1=sqrt(x[1]^2*sigma.x2 + x[2]^2*ig.sigma2)
return(c(z1))
}
x0 = c(0.5,0.5)
beta0.l.gamma1.min=solnp(x0, fun = fnbeta0.l.min, eqfun = eqn1, eqB =1)
beta0.q.gamma1.min=solnp(x0, fun = fnbeta0.q.min, eqfun = eqn1, eqB =1)

beta1.l.gamma1.min=solnp(x0, fun = fnbeta1.l.min, eqfun = eqn1, eqB =1)
beta1.q.gamma1.min=solnp(x0, fun = fnbeta1.q.min, eqfun = eqn1, eqB =1)



beta0.l.gamma1.max=solnp(x0, fun = fnbeta0.l.max, eqfun = eqn1, eqB =1)
beta0.q.gamma1.max=solnp(x0, fun = fnbeta0.q.max, eqfun = eqn1, eqB =1)
beta1.l.gamma1.max=solnp(x0, fun = fnbeta1.l.max, eqfun = eqn1, eqB =1)
beta1.q.gamma1.max=solnp(x0, fun = fnbeta1.q.max, eqfun = eqn1, eqB =1)

r.beta0.l.l=coef[1,1]+round(tail(beta0.l.gamma1.min$values, n=1),3)
r.beta0.l.u=coef[1,1]+round(-tail(beta0.l.gamma1.max$values, n=1),3)
```

Appendix A (Continued)

```
r.beta1.l.l=coef[2,1]+round(tail(beta1.l.gamma1.min$values, n=1),3)
r.beta1.l.u=coef[2,1]+round(-tail(beta1.l.gamma1.max$values, n=1),3)


r.beta0.q.l=coef[1,1]+round(tail(beta0.q.gamma1.min$values, n=1),3)
r.beta0.q.u=coef[1,1]+round(-tail(beta0.q.gamma1.max$values, n=1),3)

r.beta1.q.l=coef[2,1]+round(tail(beta1.q.gamma1.min$values, n=1),3)
r.beta1.q.u=coef[2,1]+round(-tail(beta1.q.gamma1.max$values, n=1),3)


r.beta0.l<- paste0("(", r.beta0.l.l, ", ", r.beta0.l.u,")")
r.beta0.q<- paste0("(", r.beta0.q.l, ", ", r.beta0.q.u,")")

r.beta1.l<- paste0("(", r.beta1.l.l, ", ", r.beta1.l.u,")")
r.beta1.q<- paste0("(", r.beta1.q.l, ", ", r.beta1.q.u,")")

#c statistics
fnbeta0.q.c=function(x)
{
-(isnilx.beta0*x[1] + isnily.beta0*x[2] +
(1/2)*((x[1]*isniqxx.beta0 + x[2]*isniqxy.beta0)*x[1] +
(x[1]*isniqxy.beta0 + x[2]*isniqyy.beta0)*x[2]))
}

fnbeta1.q.c=function(x)
{
-(isnilx.beta1*x[1] + isnily.beta1*x[2] +
(1/2)*((x[1]*isniqxx.beta1 + x[2]*isniqxy.beta1)*x[1] +
(x[1]*isniqxy.beta1 + x[2]*isniqyy.beta1)*x[2]))
}

eqnc=function(x){
z1=sqrt(x[1]^2*sigma.x2 + x[2]^2*ig.sigma2)
return(c(z1))
}

x0 = c(0.5,0.5)
```

Appendix A (Continued)

```
# derive c statistics for beta0
c.b0=seq(0, 1, by=0.001)
range.beta0=rep(0, length(c.b0))
for (i in 1:length(c.b0))
   {
range.beta0[i]=round(-tail(solnp(x0, fun = fnbeta0.q.c, eqfun = eqnc, eqB =c.b0[i])$values, n=1), 3)
   }

c.beta0.range = cbind(range.beta0, c.b0)
c.beta0.qx = subset(c.beta0.range, range.beta0==coef[1,2])
c.beta0.q = min(c.beta0.qx[,2])

# derive c statistics for beta1
c.b1=seq(0, 3, by=0.005)
range.beta1=rep(0, length(c.b1))
for (i in 1:length(c.b1))
   {
range.beta1[i]=round(-tail(solnp(x0, fun = fnbeta1.q.c, eqfun = eqnc, eqB =c.b1[i])$values, n=1), 3)
   }
c.beta1.range = cbind(range.beta1, c.b1)
c.beta1.qx = subset(c.beta1.range, range.beta1==coef[2,2])
c.beta1.q = min(c.beta1.qx[,2])

isnil<-rbind(isnil0, isnil1)
isniq<-rbind(isniq0, isniq1)
range.l<-rbind(r.beta0.l, r.beta1.l)
c.l=rbind(c.l.beta0, c.l.beta1)
range.q<-rbind(r.beta0.q, r.beta1.q)
c.q<-rbind(c.beta0.q, c.beta1.q)
isni.nn.yxdep<-cbind(coef, isnil, isniq, range.l, c.l, range.q, c.q)

colnames(isni.nn.yxdep) <- c("MAR Est", "SE", "ISNILy", "ISNILyy", "Linear Correction_range",
                          "Linear Correction_c", "NonLinear Correction_range", "NonLinear Correction_c")
isni.nn.yxdep
```

Appendix B

R CODE FOR ISNI ANALYSIS FOR EMA DATA WITH NORMAL DISTRIBUTED OUTCOME AND A BERNOULLI DISTRIBUTED COVARIATE

```
library (Rsolnp)

##########################################################################
############ Apply to real data (EMA) - Simple Linear regression ###########
#### for Normal Distributed Outcome And Bernoulli Distributed Covariate ####
##########################################################################


## Read in the function.
source('C:\\Users\\gaowh\\Documents\\Thesis_2013\\Thesis_gao\\nisni R program\\nisniglm.R')

#read ema data
ema = read.csv("C:\\Users\\gaowh\\Documents\\Thesis_2013\\Hui Papers\\EMA data\\emabase.csv",
               header = TRUE)

#check missing data
sapply(ema, function(x) sum(is.na(x)))

###only analyze non-smoking subjects###
ema2 = ema[ema$smoking==1,]
attach(ema2)

# recategorize comp
rcomp <- ifelse(comp == 1, 0, 1)

## compute isnil and isniq
sim.isni <- glm(mood~rcomp)
out<- summary(sim.isni)
coef <- cbind(round(out$coefficients[ , 1],3), round(out$coefficients[ , 2],3))
```

Appendix B (Continued)

```
## get the residual variance
ig.sigma2 <- round(summary(sim.isni)$dispersion*(length(mood[g==1])-length(coef(sim.isni)))/
            length(mood[g==1]),3)

## verify with explicit formula
n=length(mood)
nm= sum(is.na(mood))
no=length(mood)-nm
xobar=mean(rcomp, na.rm=T)
sigma.x2=(sd(rcomp, na.rm=T))^2
sxoxo= sum((rcomp[g==1]-xobar)^2)
sd.y = sqrt(ig.sigma2)


####################################################
### For Y-Dependent-only Nonignorable Nonresponse ###
####################################################
isnil.beta1  = 0
isniq.beta1  = round(2*ig.sigma2*((no^2*nm)/n^2)*sigma.x2*sim.isni$coef[2]*(1/sxoxo),3)

isnil.beta0  = round(-ig.sigma2*(nm/n),3)
isniq.beta0  = round(-isniq.beta1*xobar,3)

isnil<-rbind(isnil.beta0, isnil.beta1)
isniq<-rbind(isniq.beta0, isniq.beta1)
isni<-cbind(isnil, isniq)


quad<- function(se, isnil, isniq){
if (isniq!=0){
if(isniq >0){
z <- matrix(c(-se,isnil,0.5*isniq), ncol=1)
          } else
          {
z <- matrix(c(-se,isnil,abs(0.5*isniq)), ncol=1)
          }
root<-polyroot(z)
mini<-min(c(Re(abs(root[1])), Re(abs(root[2]))))
            }else
            {
```

Appendix B (Continued)

```
mini<-abs(se/isnil)
              }
res<- round(abs(mini),3)
}
c.l.beta0<-quad(coef[1,2],isnil.beta0,0)
c.q.beta0<-quad(coef[1,2],isnil.beta0,isniq.beta0)
c.l.beta1<-quad(coef[2,2],isnil.beta1,0)
c.q.beta1<-quad(coef[2,2],isnil.beta1,isniq.beta1)

#calculate calibrated estimates
gamma1l=1
gamma1u=-1

gamma1 <- seq(gamma1u, gamma1l, 0.01)
beta0.isnill = min(round(sim.isni$coef[1] + isnil.beta0*gamma1,3))
beta0.isnilu = max(round(sim.isni$coef[1] + isnil.beta0*gamma1,3))
beta0.isniql = min(round(sim.isni$coef[1] + isnil.beta0*gamma1 + (isniq.beta0/2)*(gamma1^2), 3))
beta0.isniqu = max(round(sim.isni$coef[1] + isnil.beta0*gamma1 + (isniq.beta0/2)*(gamma1^2), 3))
r.beta0.l<- paste0("(", round(beta0.isnill,3), ", ", beta0.isnilu,")")
r.beta0.q<- paste0("(", round(beta0.isniql,3), ", ", beta0.isniqu,")")

beta1.isnill = min(round(sim.isni$coef[2] + isnil.beta1*gamma1,3))
beta1.isnilu = max(sim.isni$coef[2] + isnil.beta1*gamma1,3)
beta1.isniql = min(round(sim.isni$coef[2] + isnil.beta1*gamma1 + (isniq.beta1/2)*(gamma1^2), 3))
beta1.isniqu = max(round(sim.isni$coef[2] + isnil.beta1*gamma1 + (isniq.beta1/2)*(gamma1^2), 3))
r.beta1.l<- paste0("(", round(beta1.isnill,3), ", ", beta1.isnilu,")")
r.beta1.q<- paste0("(", round(beta1.isniql,3), ", ", beta1.isniqu,")")

isni.beta0 <- cbind(isnil.beta0, isniq.beta0, r.beta0.l, c.l.beta0, r.beta0.q, c.q.beta0)
isni.beta1 <- cbind(isnil.beta1, isniq.beta1, r.beta1.l, c.l.beta1, r.beta1.q, c.q.beta1)
isni.beta  <- rbind(isni.beta0, isni.beta1)

isni.nb.ydep<- cbind(coef, isni.beta)
colnames(isni.nb.ydep) <- c("MAR Est", "SE", "ISNILy", "ISNILyy", "Linear Correction_range",
                            "Linear Correction_c", "NonLinear Correction_range", "NonLinear Correction_c")
isni.nb.ydep


####################################################
### For Y-and-X-Dependent Nonignorable Nonresponse ###
```

Appendix B (Continued)

```
####################################################

isnilx.beta0  = 0
isnily.beta0  = -ig.sigma2*(nm/n)
isnilx.beta1  = 0
isnily.beta1  = 0

isniqxx.beta1  = 0
isniqxy.beta1  = ig.sigma2*((no*nm)/n^2)*((no*sigma.x2)/sxoxo)
isniqyy.beta1  = 2*sim.isni$coef[2]*ig.sigma2*((no*nm)/n^2)*((no*sigma.x2)/sxoxo)
isniqxx.beta0  = 0
isniqxy.beta0  = -isniqxy.beta1*xobar
isniqyy.beta0  = -isniqyy.beta1*xobar

isnil0<-paste0("(", round(isnily.beta0,3), ", ", isnilx.beta0,")")
isnil1<-paste0("(", round(isnily.beta1,3), ", ", isnilx.beta1,")")

isniq0<-paste0("(", round(isniqyy.beta0, 3), ", ", round(isniqxy.beta0, 3), ", ", isniqxx.beta0,")")
isniq1<-paste0("(", round(isniqyy.beta1, 3), ", ", round(isniqxy.beta1, 3), ", ", isniqxx.beta1,")")


#calculate calibrated estimates
#minimum value of range
fnbeta0.l.min=function(x)
{
isnilx.beta0*x[1] + isnily.beta0*x[2]
}
fnbeta0.q.min=function(x)
{
isnilx.beta0*x[1] + isnily.beta0*x[2] +
(1/2)*((x[1]*isniqxx.beta0 + x[2]*isniqxy.beta0)*x[1] +
(x[1]*isniqxy.beta0 + x[2]*isniqyy.beta0)*x[2])
}
fnbeta1.l.min=function(x)
{
isnilx.beta1*x[1] + isnily.beta1*x[2]
}
fnbeta1.q.min=function(x)
{
isnilx.beta1*x[1] + isnily.beta1*x[2] +
```

Appendix B (Continued)

```
(1/2)*((x[1]*isniqxx.beta1 + x[2]*isniqxy.beta1)*x[1] +
(x[1]*isniqxy.beta1 + x[2]*isniqyy.beta1)*x[2])
}

#maximum value of range
fnbeta0.l.max=function(x)
{
-(isnilx.beta0*x[1] + isnily.beta0*x[2])
}
fnbeta0.q.max=function(x)
{
-(isnilx.beta0*x[1] + isnily.beta0*x[2] +
(1/2)*((x[1]*isniqxx.beta0 + x[2]*isniqxy.beta0)*x[1] +
(x[1]*isniqxy.beta0 + x[2]*isniqyy.beta0)*x[2]))
}
fnbeta1.l.max=function(x)
{
-(isnilx.beta1*x[1] + isnily.beta1*x[2])
}
fnbeta1.q.max=function(x)
{
-(isnilx.beta1*x[1] + isnily.beta1*x[2] +
(1/2)*((x[1]*isniqxx.beta1 + x[2]*isniqxy.beta1)*x[1] +
(x[1]*isniqxy.beta1 + x[2]*isniqyy.beta1)*x[2]))
}

eqn1=function(x){
z1=sqrt(x[1]^2 + x[2]^2*ig.sigma2)
return(c(z1))
}

x0 = c(0,0)

beta0.l.gamma1.min=solnp(x0, fun = fnbeta0.l.min, eqfun = eqn1, eqB =1)
beta0.q.gamma1.min=solnp(x0, fun = fnbeta0.q.min, eqfun = eqn1, eqB =1)
beta1.l.gamma1.min=solnp(x0, fun = fnbeta1.l.min, eqfun = eqn1, eqB =1)
beta1.q.gamma1.min=solnp(x0, fun = fnbeta1.q.min, eqfun = eqn1, eqB =1)

beta0.l.gamma1.max=solnp(x0, fun = fnbeta0.l.max, eqfun = eqn1, eqB =1)
beta0.q.gamma1.max=solnp(x0, fun = fnbeta0.q.max, eqfun = eqn1, eqB =1)
```

Appendix B (Continued)

```
beta1.l.gamma1.max=solnp(x0, fun = fnbeta1.l.max, eqfun = eqn1, eqB =1)
beta1.q.gamma1.max=solnp(x0, fun = fnbeta1.q.max, eqfun = eqn1, eqB =1)

r.beta0.l.l=coef[1,1]+round(tail(beta0.l.gamma1.min$values, n=1),3)
r.beta0.l.u=coef[1,1]+round(-tail(beta0.l.gamma1.max$values, n=1),3)

r.beta1.l.l=coef[2,1]+round(tail(beta1.l.gamma1.min$values, n=1),3)
r.beta1.l.u=coef[2,1]+round(-tail(beta1.l.gamma1.max$values, n=1),3)

r.beta0.q.l=coef[1,1]+round(tail(beta0.q.gamma1.min$values, n=1),3)
r.beta0.q.u=coef[1,1]+round(-tail(beta0.q.gamma1.max$values, n=1),3)

r.beta1.q.l=coef[2,1]+round(tail(beta1.q.gamma1.min$values, n=1),3)
r.beta1.q.u=coef[2,1]+round(-tail(beta1.q.gamma1.max$values, n=1),3)

r.beta0.l<- paste0("(", r.beta0.l.l, ", ", r.beta0.l.u,")")
r.beta0.q<- paste0("(", r.beta0.q.l, ", ", r.beta0.q.u,")")

r.beta1.l<- paste0("(", r.beta1.l.l, ", ", r.beta1.l.u,")")
r.beta1.q<- paste0("(", r.beta1.q.l, ", ", r.beta1.q.u,")")


#c statistics
fnbeta0.q.c=function(x)
{
-(isnilx.beta0*x[1] + isnily.beta0*x[2] +
(1/2)*((x[1]*isniqxx.beta0 + x[2]*isniqxy.beta0)*x[1] +
(x[1]*isniqxy.beta0 + x[2]*isniqyy.beta0)*x[2]))
}

fnbeta1.q.c=function(x)
{
-(isnilx.beta1*x[1] + isnily.beta1*x[2] +
(1/2)*((x[1]*isniqxx.beta1 + x[2]*isniqxy.beta1)*x[1] +
(x[1]*isniqxy.beta1 + x[2]*isniqyy.beta1)*x[2]))
}

eqnc=function(x){
z1=sqrt(x[1]^2 + x[2]^2*ig.sigma2)
return(c(z1))
```

Appendix B (Continued)

```
}

x0 = c(0,0)

# derive c statistics for beta0
c.b0=seq(0, 1, by=0.001)
range.beta0=rep(0, length(c.b0))
for (i in 1:length(c.b0))
   {
range.beta0[i]=round(-tail(solnp(x0, fun = fnbeta0.q.c, eqfun = eqnc, eqB =c.b0[i])$values, n=1), 3)
   }

c.beta0.range = cbind(range.beta0, c.b0)
c.beta0.qx = subset(c.beta0.range, range.beta0==coef[1,2])
c.beta0.q = min(c.beta0.qx[,2])



# derive c statistics for beta1
c.b1=seq(0, 1.5, by=0.001)
range.beta1=rep(0, length(c.b1))
for (i in 1:length(c.b1))
   {
range.beta1[i]=round(-tail(solnp(x0, fun = fnbeta1.q.c, eqfun = eqnc, eqB =c.b1[i])$values, n=1), 3)
   }
c.beta1.range = cbind(range.beta1, c.b1)
c.beta1.qx = subset(c.beta1.range, range.beta1==coef[2,2])
c.beta1.q = min(c.beta1.qx[,2])

isnil<-rbind(isnil0, isnil1)
isniq<-rbind(isniq0, isniq1)
range.l<-rbind(r.beta0.l, r.beta1.l)
c.l<-rbind(c.l.beta0, c.l.beta1)
range.q<-rbind(r.beta0.q, r.beta1.q)
c.q<-rbind(c.beta0.q, c.beta1.q)
isni.nb.yxdep<-cbind(coef, isnil, isniq, range.l, c.l, range.q, c.q)

colnames(isni.nb.yxdep) <- c("MAR Est", "SE", "ISNILy", "ISNILyy", "Linear Correction_range",
                             "Linear Correction_c", "NonLinear Correction_range", "NonLinear Correction_c")
isni.nb.yxdep
```

Appendix B (Continued)

```
#################################################################################
########################## Multiple Linear Regression #########################
#################################################################################

sim.isni <- glm(mood~smoking + rcomp)
out<- summary(sim.isni)
coef <- cbind(round(out$coefficients[ , 1],3), round(out$coefficients[ , 2],3))

## get the residual variance
ig.sigma2 <- round(summary(sim.isni)$dispersion*(length(mood[g==1])-length(coef(sim.isni)))/
             length(mood[g==1]),3)

#use function to obtain isnil and isniq
isni  <- nisniglm(dep=mood, indep.ymodel=rcomp, indep.dist="binomial", cov.ymodel=smoking,
                  indep.xmodel=smoking, indep.gmodel=smoking)
isni2<-do.call("cbind", isni)
isnil<- isni2[,1]
isniq.yy<- isni2[,2]
isniq.xy<- isni2[,3]


#################################################################################
### For Y-Dependent Only Nonignorable Nonresponse ###
#################################################################################

isnil.beta0 <- isnil[1]
isnil.beta1 <- isnil[2]
isnil.beta2 <- isnil[3]
isniq.yy.beta0 <- isniq.yy[1]
isniq.yy.beta1 <- isniq.yy[2]
isniq.yy.beta2 <- isniq.yy[3]

quad<- function(se, isnil, isniq.yy){
if (isniq.yy!=0){
if(isniq.yy>0){
z <- matrix(c(-se,isnil,0.5*isniq.yy), ncol=1)
           } else
           {
z <- matrix(c(-se,isnil,abs(0.5*isniq.yy)), ncol=1)
```

```
            }
root<-polyroot(z)
mini<-min(c(Re(abs(root[1])), Re(abs(root[2]))))
              }else
              {
mini<-abs(se/isnil)
              }
res<- round(abs(mini),3)
}

c.l.beta0<-quad(coef[1,2],isnil.beta0,0)
c.q.beta0<-quad(coef[1,2],isnil.beta0,isniq.yy.beta0)
c.l.beta1<-quad(coef[2,2],isnil.beta1,0)
c.q.beta1<-quad(coef[2,2],isnil.beta1,isniq.yy.beta1)
c.l.beta2<-quad(coef[3,2],isnil.beta2,0)
c.q.beta2<-quad(coef[3,2],isnil.beta2,isniq.yy.beta2)


#calculate calibrated estimates
gamma1l=1
gamma1u=-1

gamma1 <- seq(gamma1u, gamma1l, 0.01)
beta0.isnill = min(round(sim.isni$coef[1] + isnil.beta0*gamma1,3))
beta0.isnilu = max(round(sim.isni$coef[1] + isnil.beta0*gamma1,3))
beta0.isniq.yyl = min(round(sim.isni$coef[1] + isnil.beta0*gamma1 + (isniq.yy.beta0/2)*(gamma1^2), 3))
beta0.isniq.yyu = max(round(sim.isni$coef[1] + isnil.beta0*gamma1 + (isniq.yy.beta0/2)*(gamma1^2), 3))
r.beta0.l<- paste0("(", round(beta0.isnill,3), ", ", beta0.isnilu,")")
r.beta0.q<- paste0("(", round(beta0.isniq.yyl,3), ", ", beta0.isniq.yyu,")")


beta1.isnill = min(round(sim.isni$coef[2] + isnil.beta1*gamma1,3))
beta1.isnilu = max(sim.isni$coef[2] + isnil.beta1*gamma1,3)
beta1.isniq.yyl = min(round(sim.isni$coef[2] + isnil.beta1*gamma1 + (isniq.yy.beta1/2)*(gamma1^2), 3))
beta1.isniq.yyu = max(round(sim.isni$coef[2] + isnil.beta1*gamma1 + (isniq.yy.beta1/2)*(gamma1^2), 3))
r.beta1.l<- paste0("(", round(beta1.isnill,3), ", ", beta1.isnilu,")")
r.beta1.q<- paste0("(", round(beta1.isniq.yyl,3), ", ", beta1.isniq.yyu,")")

beta2.isnill = min(round(sim.isni$coef[3] + isnil.beta2*gamma1,3))
beta2.isnilu = max(sim.isni$coef[3] + isnil.beta2*gamma1,3)
```

Appendix B (Continued)

```
beta2.isniq.yyl = min(round(sim.isni$coef[3] + isnil.beta2*gamma1 + (isniq.yy.beta2/2)*(gamma1^2), 3))
beta2.isniq.yyu = max(round(sim.isni$coef[3] + isnil.beta2*gamma1 + (isniq.yy.beta2/2)*(gamma1^2), 3))
r.beta2.l<- paste0("(", round(beta2.isnill,3), ", ", beta2.isnilu,")")
r.beta2.q<- paste0("(", round(beta2.isniq.yyl,3), ", ", beta2.isniq.yyu,")")

isni.beta0 <- cbind(isnil.beta0, isniq.yy.beta0, r.beta0.l, c.l.beta0, r.beta0.q, c.q.beta0)
isni.beta1 <- cbind(isnil.beta1, isniq.yy.beta1, r.beta1.l, c.l.beta1, r.beta1.q, c.q.beta1)
isni.beta2 <- cbind(isnil.beta2, isniq.yy.beta2, r.beta2.l, c.l.beta2, r.beta2.q, c.q.beta2)
isni.beta  <- rbind(isni.beta0, isni.beta1, isni.beta2)

isni.nb.ydep.mult<- cbind(coef, isni.beta)
colnames(isni.nb.ydep.mult) <- c("MAR Est", "SE", "ISNILy", "ISNILyy", "Linear Correction_range",
                                 "Linear Correction_c", "NonLinear Correction_range",

isni.nb.ydep.mult


##########################################################################
### For Y-and-X-Dependent Nonignorable Nonresponse ###
##########################################################################

isnilx.beta0  = 0
isnily.beta0  = isnil[1]
isnilx.beta1  = 0
isnily.beta1  = isnil[2]
isnilx.beta2  = 0
isnily.beta2  = isnil[3]

isniqxx.beta0  = 0
isniqxy.beta0  = isniq.xy[1]
isniqyy.beta0  = isniq.yy[1]
isniqxx.beta1  = 0
isniqxy.beta1  = isniq.xy[2]
isniqyy.beta1  = isniq.yy[2]
isniqxx.beta2  = 0
isniqxy.beta2  = isniq.xy[3]
isniqyy.beta2  = isniq.yy[3]

isnil0<-paste0("(", round(isnily.beta0,3), ", ", isnilx.beta0,")")
isnil1<-paste0("(", round(isnily.beta1,3), ", ", isnilx.beta1,")")
```

Appendix B (Continued)

```
isnil2<-paste0("(", round(isnily.beta2,3), ", ", isnilx.beta2,")")

isniq0<-paste0("(", round(isniqyy.beta0, 3), ", ", round(isniqxy.beta0, 3), ", ", isniqxx.beta0,")")
isniq1<-paste0("(", round(isniqyy.beta1, 3), ", ", round(isniqxy.beta1, 3), ", ", isniqxx.beta1,")")
isniq2<-paste0("(", round(isniqyy.beta2, 3), ", ", round(isniqxy.beta2, 3), ", ", isniqxx.beta2,")")


#calculate calibrated estimates
#minimum value of range
fnbeta0.l.min=function(x)
{
isnilx.beta0*x[1] + isnily.beta0*x[2]
}
fnbeta0.q.min=function(x)
{
isnilx.beta0*x[1] + isnily.beta0*x[2] +
(1/2)*((x[1]*isniqxx.beta0 + x[2]*isniqxy.beta0)*x[1] +
(x[1]*isniqxy.beta0 + x[2]*isniqyy.beta0)*x[2])
}
fnbeta1.l.min=function(x)
{
isnilx.beta1*x[1] + isnily.beta1*x[2]
}
fnbeta1.q.min=function(x)
{
isnilx.beta1*x[1] + isnily.beta1*x[2] +
(1/2)*((x[1]*isniqxx.beta1 + x[2]*isniqxy.beta1)*x[1] +
(x[1]*isniqxy.beta1 + x[2]*isniqyy.beta1)*x[2])
}

fnbeta2.l.min=function(x)
{
isnilx.beta2*x[1] + isnily.beta2*x[2]
}
fnbeta2.q.min=function(x)
{
isnilx.beta2*x[1] + isnily.beta2*x[2] +
(1/2)*((x[1]*isniqxx.beta2 + x[2]*isniqxy.beta2)*x[1] +
(x[1]*isniqxy.beta2 + x[2]*isniqyy.beta2)*x[2])
}
```

Appendix B (Continued)

```
#maximum value of range
fnbeta0.l.max=function(x)
{
-(isnilx.beta0*x[1] + isnily.beta0*x[2])
}
fnbeta0.q.max=function(x)
{
-(isnilx.beta0*x[1] + isnily.beta0*x[2] +
(1/2)*((x[1]*isniqxx.beta0 + x[2]*isniqxy.beta0)*x[1] +
(x[1]*isniqxy.beta0 + x[2]*isniqyy.beta0)*x[2]))
}
fnbeta1.l.max=function(x)
{
-(isnilx.beta1*x[1] + isnily.beta1*x[2])
}
fnbeta1.q.max=function(x)
{
-(isnilx.beta1*x[1] + isnily.beta1*x[2] +
(1/2)*((x[1]*isniqxx.beta1 + x[2]*isniqxy.beta1)*x[1] +
(x[1]*isniqxy.beta1 + x[2]*isniqyy.beta1)*x[2]))
}
fnbeta2.l.max=function(x)
{
-(isnilx.beta2*x[1] + isnily.beta2*x[2])
}
fnbeta1.q.max=function(x)
{
-(isnilx.beta2*x[1] + isnily.beta2*x[2] +
(1/2)*((x[1]*isniqxx.beta2 + x[2]*isniqxy.beta2)*x[1] +
(x[1]*isniqxy.beta2 + x[2]*isniqyy.beta2)*x[2]))
}

eqn1=function(x){
z1=sqrt(x[1]^2 + x[2]^2*ig.sigma2)
return(c(z1))
}
x0 = c(0,0)
beta0.l.gamma1.min=solnp(x0, fun = fnbeta0.l.min, eqfun = eqn1, eqB =1)
```

Appendix B (Continued)

```
beta0.q.gamma1.min=solnp(x0, fun = fnbeta0.q.min, eqfun = eqn1, eqB =1)
beta1.l.gamma1.min=solnp(x0, fun = fnbeta1.l.min, eqfun = eqn1, eqB =1)
beta1.q.gamma1.min=solnp(x0, fun = fnbeta1.q.min, eqfun = eqn1, eqB =1)
beta2.l.gamma1.min=solnp(x0, fun = fnbeta2.l.min, eqfun = eqn1, eqB =1)
beta2.q.gamma1.min=solnp(x0, fun = fnbeta2.q.min, eqfun = eqn1, eqB =1)

beta0.l.gamma1.max=solnp(x0, fun = fnbeta0.l.max, eqfun = eqn1, eqB =1)
beta0.q.gamma1.max=solnp(x0, fun = fnbeta0.q.max, eqfun = eqn1, eqB =1)
beta1.l.gamma1.max=solnp(x0, fun = fnbeta1.l.max, eqfun = eqn1, eqB =1)
beta1.q.gamma1.max=solnp(x0, fun = fnbeta1.q.max, eqfun = eqn1, eqB =1)
beta2.l.gamma1.max=solnp(x0, fun = fnbeta2.l.max, eqfun = eqn1, eqB =1)
beta2.q.gamma1.max=solnp(x0, fun = fnbeta2.q.max, eqfun = eqn1, eqB =1)

r.beta0.l.l=coef[1,1]+round(tail(beta0.l.gamma1.min$values, n=1),3)
r.beta0.l.u=coef[1,1]+round(-tail(beta0.l.gamma1.max$values, n=1),3)


r.beta1.l.l=coef[2,1]+round(tail(beta1.l.gamma1.min$values, n=1),3)
r.beta1.l.u=coef[2,1]+round(-tail(beta1.l.gamma1.max$values, n=1),3)

r.beta2.l.l=coef[3,1]+round(tail(beta2.l.gamma1.min$values, n=1),3)
r.beta2.l.u=coef[3,1]+round(-tail(beta2.l.gamma1.max$values, n=1),3)


r.beta0.q.l=coef[1,1]+round(tail(beta0.q.gamma1.min$values, n=1),3)
r.beta0.q.u=coef[1,1]+round(-tail(beta0.q.gamma1.max$values, n=1),3)

r.beta1.q.l=coef[2,1]+round(tail(beta1.q.gamma1.min$values, n=1),3)
r.beta1.q.u=coef[2,1]+round(-tail(beta1.q.gamma1.max$values, n=1),3)

r.beta2.q.l=coef[3,1]+round(tail(beta2.q.gamma1.min$values, n=1),3)
r.beta2.q.u=coef[3,1]+round(-tail(beta2.q.gamma1.max$values, n=1),3)

r.beta0.l<- paste0("(", r.beta0.l.l, ", ", r.beta0.l.u,")")
r.beta0.q<- paste0("(", r.beta0.q.l, ", ", r.beta0.q.u,")")

r.beta1.l<- paste0("(", r.beta1.l.l, ", ", r.beta1.l.u,")")
r.beta1.q<- paste0("(", r.beta1.q.l, ", ", r.beta1.q.u,")")

r.beta2.l<- paste0("(", r.beta2.l.l, ", ", r.beta2.l.u,")")
```

Appendix B (Continued)

```
r.beta2.q<- paste0("(", r.beta2.q.l, ", ", r.beta2.q.u,")")

#c statistics
fnbeta0.q.c=function(x)
{
-(isnilx.beta0*x[1] + isnily.beta0*x[2] +
(1/2)*((x[1]*isniqxx.beta0 + x[2]*isniqxy.beta0)*x[1] +
(x[1]*isniqxy.beta0 + x[2]*isniqyy.beta0)*x[2]))
}

fnbeta1.q.c=function(x)
{
-(isnilx.beta1*x[1] + isnily.beta1*x[2] +
(1/2)*((x[1]*isniqxx.beta1 + x[2]*isniqxy.beta1)*x[1] +
(x[1]*isniqxy.beta1 + x[2]*isniqyy.beta1)*x[2]))
}

fnbeta2.q.c=function(x)
{
-(isnilx.beta2*x[1] + isnily.beta2*x[2] +
(1/2)*((x[1]*isniqxx.beta2 + x[2]*isniqxy.beta2)*x[1] +
(x[1]*isniqxy.beta2 + x[2]*isniqyy.beta2)*x[2]))
}

eqnc=function(x){
z1=sqrt(x[1]^2 + x[2]^2*ig.sigma2)
return(c(z1))
}

x0 = c(0,0)


# derive c statistics for beta0
c.b0=seq(0, 1, by=0.001)
range.beta0=rep(0, length(c.b0))
for (i in 1:length(c.b0))
   {
range.beta0[i]=round(-tail(solnp(x0, fun = fnbeta0.q.c, eqfun = eqnc, eqB =c.b0[i])$values, n=1), 3)
   }
```

Appendix B (Continued)

```
c.beta0.range = cbind(range.beta0, c.b0)
c.beta0.qx = subset(c.beta0.range, range.beta0==coef[1,2])
c.beta0.q = min(c.beta0.qx[,2])


# derive c statistics for beta1
c.b1=seq(0, 2, by=0.001)
range.beta1=rep(0, length(c.b1))
for (i in 1:length(c.b1))
   {
range.beta1[i]=round(-tail(solnp(x0, fun = fnbeta1.q.c, eqfun = eqnc, eqB =c.b1[i])$values, n=1), 3)
   }
c.beta1.range = cbind(range.beta1, c.b1)
c.beta1.qx = subset(c.beta1.range, range.beta1==coef[2,2])
c.beta1.q = min(c.beta1.qx[,2])

# derive c statistics for beta2
c.b2=seq(0, 2, by=0.001)
range.beta2=rep(0, length(c.b2))
for (i in 1:length(c.b2))
   {
range.beta2[i]=round(-tail(solnp(x0, fun = fnbeta2.q.c, eqfun = eqnc, eqB =c.b2[i])$values, n=1), 3)
   }
c.beta2.range = cbind(range.beta2, c.b2)
c.beta2.qx = subset(c.beta2.range, range.beta2==coef[2,2])
c.beta2.q = min(c.beta2.qx[,2])

isnil<-rbind(isnil0, isnil1, isnil2)
isniq<-rbind(isniq0, isniq1, isniq2)
range.l<-rbind(r.beta0.l, r.beta1.l, r.beta2.l)
c.l=rbind(c.l.beta0, c.l.beta1, c.l.beta2)
range.q<-rbind(r.beta0.q, r.beta1.q, r.beta2.q)
c.q<-rbind(c.beta0.q, c.beta1.q, c.beta2.q)
isni.nb.yxdep.mult<-cbind(coef, isnil, isniq, range.l, c.l, range.q, c.q)


colnames(isni.nb.yxdep.mult) <- c("MAR Est", "SE", "ISNILy", "ISNILyy", "Linear Correction_range",
                                  "Linear Correction_c", "NonLinear Correction_range",

isni.nb.yxdep.mult
```

Appendix C

R CODE FOR NISNIGLM FUNCTION USED IN THE ABOVE EXAMPLES

```
## Function to calculate Nonlinear ISNI, the index of sensitivity to
## nonignorability, in a generalized linear model for distributions of Gaussian.
## The dataset needs to include g varaible (indicator of observed data (g=1) and missing data (g=0))

nisniglm<-function(dep, indep.ymodel, indep.dist="gaussian", cov.ymodel=NULL, indep.xmodel=NULL,
                   indep.gmodel=NULL){
#dep: outcome varaible y, vector
#indep.dist: covariate's distribution, gaussian or binormial distribution
#indep.ymodel: indepedent variable x for y model, vector, having simutanuous missing data as y
#cov.ymodel: fully observed covariates w, vector or matrix or null
#indep.xmodel: covariates for x model, vector or matrix or null
#indep.gmodel: covariates for g model (missing data model), vector or matrix or null

#############################
# Set up needed parameters #
#############################
 y              <- dep
 yo             <- y[g==1]
 sdy            <- sd(yo)
 x.single       <- indep.ymodel
 xo.single      <- x.single[g==1]

 dist.int<-charmatch(indep.dist,c("gaussian","binomial"))

 if (is.null(cov.ymodel)==TRUE){
 cov.ymodel     <- cov.ymodel
 } else
 {
 cov.ymodel     <- as.matrix(cov.ymodel)
 cov.o.ymodel   <- cov.ymodel[g==1,]
 }
 sdx            <- sd(xo.single)
```

Appendix C (Continued)

```
  if (is.null(indep.xmodel)==TRUE){
  indep.xmodel   <- indep.xmodel
  } else
  {
  indep.xmodel   <- as.matrix(indep.xmodel)
  indep.o.xmodel   <- indep.xmodel[g==1,]
  }

  indep.gmodel   <- indep.gmodel

##############
# Run xmodel #
##############
  if (is.null(indep.xmodel)==TRUE){
  indep.o.xmodel <- rep(1, length(yo))
  data.xmodel<- data.frame(xo.single, indep.o.xmodel)
  xmodel<-switch(dist.int,
                 lm(xo.single~., data.xmodel),
                 glm(xo.single~., data.xmodel, family=binomial),
                 stop("Invalid family type")
                )
  coef.xmodel<- summary(xmodel)$coef[,1]
  xfit <-switch(dist.int,
                predict(xmodel),
                exp(coef.xmodel)/(1+exp(coef.xmodel)),
                stop("Invalid data")
               )

# Predict missing data for xmodel #
   xnew <- rep(1, length(y) - length(yo))
   expected.xm <-switch(dist.int,
                        coef.xmodel*xnew,
                        #exp(coef.xmodel)/(1+exp(coef.xmodel))*xnew,
                        exp(coef.xmodel*xnew)/(1+exp(coef.xmodel*xnew)),
                        stop("Invalid data")
                       )
    xm.single_<-expected.xm
    xm.single <-switch(dist.int,
                       xm.single_,
```

Appendix C (Continued)

```
                    rbinom(length(xm.single_), 1, xm.single_),
                    stop("Invalid data")
                    )
    x.single[g==0]<-xm.single
} else
{
 data.xmodel<- data.frame(xo.single, indep.o.xmodel)
 xmodel<-switch(dist.int,
                lm(xo.single~., data.xmodel),
                glm(xo.single~., data.xmodel, family=binomial),
                stop("Invalid family type")
               )
 coef.xmodel<- summary(xmodel)$coef[,1]
 xfit<-predict(xmodel)

# Predict missing data for xmodel #
 xnew <- data.frame(indep.o.xmodel=indep.xmodel[g==0,])
 colnames(xnew)<- names(coef.xmodel)[2:length(coef.xmodel)]

 expected.xm<-switch(dist.int,
                predict(xmodel, xnew),
                predict(xmodel, xnew, type="response"),
                stop("Invalid family type")
               )

  xm.single <- expected.xm
  x.single[g==0]<-xm.single
}


##############
# Run ymodel #
##############
  if (is.null(cov.ymodel)==TRUE){
 data.ymodel<- data.frame(yo, xo.single)
 ymodel<-lm(yo~., data.ymodel)
 coef.ymodel<-summary(ymodel)$coef[,1]
 yfit<-predict(ymodel)

 # Predict missing data for ymodel
```

Appendix C (Continued)

```
ynew <- data.frame(xo.single=xm.single)
expected.ym<-predict(ymodel, ynew)
y[g==0]<-expected.ym
ym<-y[g==0]
} else
{
data.ymodel<- data.frame(yo, cov.o.ymodel, xo.single)
ymodel<-lm(yo~., data.ymodel)
coef.ymodel<-summary(ymodel)$coef[,1]
yfit<-predict(ymodel)

# Predict missing data for ymodel
ynew <- data.frame(cov.o.ymodel=cov.ymodel[g==0,], xo.single=xm.single)
colnames(ynew)<- names(coef.ymodel)[2:length(coef.ymodel)]
expected.ym<-predict(ymodel, ynew)
y[g==0]<-expected.ym
ym<-y[g==0]
}

cc<-rep(1,length(y))


#####################################################################
# Obtain estimated probablity of missing data by missing data model #
#####################################################################
 if (is.null(indep.gmodel)==TRUE){
ho<-length(yo)/length(y)
hm<-ho
} else
{
gmodel<-glm(g~indep.gmodel,family=binomial,maxit=50)
ho<-gmodel$fitted.values[g==1]
hm<-gmodel$fitted.values[g==0]
}
ccm<-hm*cc[g==0]


####################################################################
# calculating linear isnil.y for X and Y are missing simutaunously #
####################################################################
```

Appendix C (Continued)

```
 x.ymodel      <- as.matrix(cbind(rep(1, length(y)), cov.ymodel, x.single))
 xo.ymodel     <- x.ymodel[g==1,]
 xm.ymodel     <- x.ymodel[g==0,]
 xxo.ymodel    <- t(xo.ymodel)%*%xo.ymodel
 tnorm.ymodel  <- sum(c(yo-yfit)^2)/length(yo)
 se.y          <- tnorm.ymodel^0.5*diag(solve(xxo.ymodel))^0.5

 nabla_yy      <- t(sqrt(cc[g==1])*xo.ymodel)%*%(sqrt(cc[g==1])*xo.ymodel)
 nabla_yr1y    <- t(xm.ymodel)%*%ccm
 isnil.y       <- -1*tnorm.ymodel*solve(nabla_yy,nabla_yr1y)

 nabla_yy      <- solve(t(sqrt(cc[g==1])*xo.ymodel)%*%(sqrt(cc[g==1])*xo.ymodel)/tnorm.ymodel)
 nabla_yr1y    <- t(xm.ymodel)%*%ccm
 isnil.y       <- -1*nabla_yy%*%nabla_yr1y

###############################################
# calculating linear isnil.x for X is missing #
###############################################
 x.xmodel       <- as.matrix(cbind(rep(1, length(y)), indep.xmodel))
 xo.xmodel      <- x.xmodel[g==1,]
 xm.xmodel      <- x.xmodel[g==0,]
 xxo.xmodel     <- t(xo.xmodel)%*%xo.xmodel
 tnorm.xmodel   <- sum(c(xo.single-xfit)^2)/length(xo.single)
 se.x           <- tnorm.xmodel^0.5*diag(solve(xxo.xmodel))^0.5

 beta.x.ymodel <- coef.ymodel[length(coef.ymodel)]

 nabla_xx       <- t(sqrt(cc[g==1])*xo.xmodel)%*%(sqrt(cc[g==1])*xo.xmodel)
 nabla_xr1y     <- beta.x.ymodel*t(xm.xmodel)%*%ccm
 isnil.x        <- -1*tnorm.xmodel*solve(nabla_xx,nabla_xr1y)

#############################################################
# calculating linear isnil.x.x for X is missing depending x #
#############################################################
 nabla_xr1x     <- t(xm.xmodel)%*%ccm
 isnil.x.x      <- -1*tnorm.xmodel*solve(nabla_xx,nabla_xr1x)

###############################
# calculating linear isnil.r0 #
###############################
```

Appendix C (Continued)

```
if (is.null(indep.gmodel)==TRUE){
cco.gmodel <- ho*(1-ho)*cc[g==1]
ccm.gmodel <- hm*(1-hm)*cc[g==0]
nabla.r0r0 <- t(sqrt(cco.gmodel))%*%sqrt(cco.gmodel) + t(sqrt(ccm.gmodel))%*%sqrt(ccm.gmodel)
nabla.r0r1 <- t(y[g==1])%*%cco.gmodel+t(y[g==0])%*%ccm.gmodel
} else
{
cco.gmodel <- ho*(1-ho)
ccm.gmodel <- hm*(1-hm)
x.gmodel   <- as.matrix(cbind(rep(1, length(y)), indep.gmodel))
xo.gmodel  <- x.gmodel[g==1,]
xm.gmodel  <- x.gmodel[g==0,]
nabla.r0r0 <-t(sqrt(cco.gmodel)*xo.gmodel)%*%
             (sqrt(cco.gmodel)*xo.gmodel)+ t(sqrt(ccm.gmodel)*xm.gmodel)%*%
                                            (sqrt(ccm.gmodel)*xm.gmodel)
nabla.r0r1 <- t(xo.gmodel)%*%(cco.gmodel*y[g==1])+t(xm.gmodel)%*%(ccm.gmodel*y[g==0])
}
isnil.r0   <- -1*solve(nabla.r0r0, nabla.r0r1)

#################################
# calculating linear isnil.r0.x #
#################################
if (is.null(indep.gmodel)==TRUE){
cco.gmodel <- ho*(1-ho)*cc[g==1]
ccm.gmodel <- hm*(1-hm)*cc[g==0]
nabla.r0r0 <-t(sqrt(cco.gmodel))%*%sqrt(cco.gmodel) + t(sqrt(ccm.gmodel))%*%
             sqrt(ccm.gmodel)
nabla.r0r1.x <- t(x.single[g==1])%*%cco.gmodel+t(x.single[g==0])%*%ccm.gmodel
} else
{
cco.gmodel <- ho*(1-ho)
ccm.gmodel <- hm*(1-hm)
x.gmodel   <- as.matrix(cbind(rep(1, length(y)), indep.gmodel))
xo.gmodel  <- x.gmodel[g==1,]
xm.gmodel  <- x.gmodel[g==0,]
nabla.r0r0 <- t(sqrt(cco.gmodel)*xo.gmodel)%*%
              (sqrt(cco.gmodel)*xo.gmodel)+ t(sqrt(ccm.gmodel)*xm.gmodel)%*%
                                                (sqrt(ccm.gmodel)*xm.gmodel)
nabla.r0r1.x <- t(xo.gmodel)%*%(cco.gmodel*x.single[g==1])+t(xm.gmodel)%*%
                (ccm.gmodel*x.single[g==0])
```

Appendix C (Continued)

```
 }
 isnil.r0.x    <- -1*solve(nabla.r0r0, nabla.r0r1.x)

#########################
# calcuating nabla_yxr1y #
#########################
 nabla_yxr1y                          <- matrix(0, length(coef.ymodel), length(coef.xmodel))
 if (is.null(indep.xmodel)==TRUE){
 nabla_yxr1y[length(coef.ymodel)]   <- sum(hm*xm.xmodel)
 } else
 {
 nabla_yxr1y[length(coef.ymodel),] <- colSums(hm*xm.xmodel)
 }
 nabla_yxr1y                          <- -nabla_yxr1y


##########################
# calcuating nabla_yr0r1y #
##########################
# hm2=(1-2*hm)*hm
 hm2=(1-hm)*hm
 ccm2=hm2*cc[g==0]
 if (is.null(indep.gmodel)==TRUE){
 nabla_yr0r1y<- -t(xm.ymodel)%*%ccm2
 } else
 {
 nabla_yr0r1y    <- -t(hm2*xm.ymodel)%*%xm.gmodel
 }


###########################
# calcuating nabla_yr1yr1y #
###########################
 hm1.yr1yr1y       <- 2*(1-hm)*hm
 hm2.yr1yr1y       <- 2*(1-2*hm)*hm
 ccm2.yr1yr1y      <- hm2.yr1yr1y*cc[g==0]

 nabla_yr1yr1y.1   <- t(xm.ymodel)%*%(hm1.yr1yr1y*ym)
 nabla_yr1yr1y.2x <- as.vector(c(rep(0, length(coef.ymodel)-1), coef.ymodel
                              [length(coef.ymodel)]*tnorm.xmodel))
```

Appendix C (Continued)

```
 nabla_yr1yr1y.2  <- matrix(rep(nabla_yr1yr1y.2x, length(xm.ymodel[,1])), length(coef.ymodel))
                        %*%ccm2.yr1yr1y
 nabla_yr1yr1y    <- - (nabla_yr1yr1y.1 + nabla_yr1yr1y.2)

########################
# calcuating nabla_yyyj #
########################
yyy<- function(j){
       xo.ymodel.j=xo.ymodel[,j]
       nabla_yyyj.11=matrix(0, length(coef.ymodel), length(coef.ymodel))
       nabla_yyyj.12=t(xo.ymodel)%*%xo.ymodel.j
       nabla_yyyj.21= t(xo.ymodel.j)%*%xo.ymodel
       #nabla_yyyj.22=2(t(yo)%*%xo.ymodel.j - t(xo.ymodel.j)%*%xo.ymodel%*%coef)
       nabla_yyyj.22=0
       nabla_yyyj.x=cbind(rbind(nabla_yyyj.11, nabla_yyyj.21), rbind(nabla_yyyj.12, nabla_yyyj.22))
       nabla_yyyj=isnil.y[j]*nabla_yyyj.x
       return(nabla_yyyj)
}

j=length(coef.ymodel)
nabla_yyy.x<-vector("list", length=j)
for (i in 1:j){
nabla_yyy.x[[i]]<-yyy(i)
}
nabla_yyy<-Reduce('+', nabla_yyy.x)%*% append(isnil.y, 0, length(coef.ymodel))
nabla_yyy2<- as.vector(nabla_yyy[1:length(coef.ymodel)])


###########################
# calcuating nabla_yr1yr1x #
###########################
 hm3           <- (1-hm)*hm
 ccm3          <- hm3*cc[g==0]
 hm4           <- (1-2*hm)*hm
 ccm4          <- hm4*cc[g==0]

 nabla_yr1yr1x.1 <- -t(hm3*xm.ymodel)%*%(x.single[g==0])
 nabla_yr1yr1x.2 <- matrix ( c(rep(0, length(xm.ymodel[1,])-1), -sum(ccm4*tnorm.xmodel)),
                                length(xm.ymodel[1,]))
 nabla_yr1yr1x <- nabla_yr1yr1x.1 + nabla_yr1yr1x.2
```

Appendix C (Continued)

```
  nabla_yy.inv<- -1*nabla_yy
  isniq.yy<- -nabla_yy.inv%*%
             (nabla_yyy2 + 2*nabla_yxr1y%*%isnil.x + 2*nabla_yr0r1y%*%isnil.r0 + nabla_yr1yr1y)
  isniq.yx<- -nabla_yy.inv%*%
             (nabla_yxr1y%*%isnil.x.x + nabla_yr0r1y%*%isnil.r0.x + nabla_yr1yr1x)

  list1.func<-list(isnil.y, isniq.yy, isniq.yx)
  names(list1.func)=c("isnil.y", "isniq.yy", "isniq.yx")
  return (list1.func)
}
```

Appendix D

PERMISSION FOR REUSED OWN ARTICLE

Dear Customer,

Thank you for your enquiry.

General permissions requests

We are pleased to advise that permission for the majority of our journal content, and for an increasing number of book publications, may be cleared quickly by using the RightsLink service via Wileyś websites http://onlinelibrary.wiley.com and www.wiley.com. Please see this link for instructions and further details.

Alternative services

copyright.com offers an alternative means for clearing permission for Wiley content.

PLSclear offers a UK based permissions clearance service for Wiley book and journal content.

RightsPortal offers an Australian and Asia Pacific based permissions clearance service for Wiley journal content.

Please note that we kindly request all permissions requests to be processed using RightsLink or one of the alternative services where available, which will provide you with a downloadable permissions license once your transaction is complete.

Accessibility requests

If you are making a request for an accessible copy of a Wiley text for those with a print disability, please note that we have received your email and no further action is required from you at this time. We will be in contact shortly.

Wiley authors

If you are a Wiley author wishing to republish extracts of your own published work, include your article in your thesis or use copies for your internal teaching purposes, please refer to the gratis reuse rights that you retain when you signed your copyright transfer agreement. Alternatively, if you still require a formal permission license please locate you work on Wiley Online Library, click "request permission", select "Author of this Wiley work" and your appropriate reuse rights to download your permissions license.

For details on how to self-archive your work visit here and our author compliance tool here. For article sharing policies, including information on sharing articles on Scholarly Collaboration Networks (SCNs), such as ResearchGate please view our handy infographic here. Please note that ResearchGate has not yet signed the STM sharing principles.

Institutional customers

If you are a college or university instructor wishing to obtain evaluation materials, desk copies or gain access to instructor resources, please visit here to locate your sales rep to discuss your requirements further.

For requests from overseas publishers to translate whole books please redirect your request to translationrights@wiley.com

For non-permissions enquiries (submitting a manuscript, access to content online etc.) please redirect your request to your local Wiley customer service team, by visiting www.wiley.com and selecting "Contact us".

If you still require assistance from the permissions team directly (content is not available online, no RightsLink "request permissions" links etc.) please resubmit your request, noting clearly "NON RIGHTSLINK" in the subject bar. We will then endeavour to get back to you as soon as possible, but a response may take up to 15 working days in peak times.

With best wishes,

The Wiley Global Permissions team

CITED LITERATURE

1. Heitjan, D. J.: Causal inference in a clinical trial: A comparative example. Controlled Clinical Trials, 20:309–318, 1999.

2. Stone, A. A., Shiffman, S., Atienza, A., and Nebeling, L.: The Science of Real-time Data Capture. New York, Oxford University Press, 2007.

3. Little, R. J. A. and Rubin, D. B.: Statistical Analysis with Missing Values. New York, John Wiley, 2002.

4. Rubin, D. B.: Multiple Imputation for Nonresponse in Surveys. New York, John Wiley and Sons, Inc, 1987.

5. Rubin, D. B.: Inference and missing data (with discussion). Biometrika, 63:581–592, 1976.

6. Troxel, A. B., Ma, G., and Heitjan, D. F.: An index of local sensitivity to nonignorability. Statistica Sinica, 14:1221–1237, 2004.

7. Heitjan, D. F. and Rubin, D. B.: Ignorability and coarse data. Annals of Statistics, 19:2244–2253, 1991.

8. Little, R. J. A.: Modeling the drop-out mechanism in longitudinal studies. Journal of the American Statistical Association, 90:1112–1121, 1995.

9. Little, R. J. A.: Nonresponse adjustments in longitudinal surveys: Models for categorical data. Bulletin of the International Statistical Institute, Proceedings of the 45th Session: Invited Papers, Section 15.1:1–18, 1985.

10. Little, R. J. A.: A note about models for selectivity bias. Econometrica, 53:1469–1474, 1985.

11. Kenward, M. G.: Selection models for repeated measurements with non-random dropout: an illustration of sensitivity. Statistics in Medicine, 17:2723–2732, 1998.

12. Wu, M. C. and Carroll, R. J.: Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. Biometrics, 44:175–188, 1988.

13. Wu, M. C. and Bailey, K. R.: Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model. Biometrics, 45:939–955, 1989.

14. Have, T. R. T., Kunselman, A. R., Pulkstenis, E. P., and Landis, J. R.: Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. Biometrics, 54:367–383, 1998.

15. Gad, A. M. and Darwish, N. M. M.: A shared parameter model for longitudinal data with missing values. American Journal of Applied Mathematics and Statistics, 1:30–35, 2013.

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

17. Council, N. R.: The Prevention and Treatment of Missing Data in Clinical Trials. Washington, D.C., The National Academies Press, 2010.

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

19. Scharfstein, D. O., Rotnitzky, A., and Robins, J. M.: Adjusting for nonignorable dropout using semiparametric models (with discussion). Journal of the American Statistical Association, 94:1096–1146, 1999.

20. Little, R. J. A. and Wang, Y.: Pattern-mixture models for multivariate incomplete data with covariates. Biometrics, 52:98–111, 1996.

21. Daniels, M. J. and Hogan, J. W.: Reparameterizing the pattern mixture model of sensitivity analyses under informative dropout. Biometrics, 95:1241–1248, 2000.

22. Copas, J. B. and Li, H. G.: Inference for non-random samples (with discussion). Journal of the Royal Statistical Society B, 59:55–95, 1997.

23. Copas, J. B. and Eguchi, S.: Local sensitivity approximations for selectivity bias. Journal of the Royal Statistical Society B, 63:871–895, 2001.

24. Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E., and Kenward, M. G.: Sensitivity analysis for nonrandom dropout: A local influence approach. Biometrics, 57:7–14, 2001.

25. Cook, R. D.: Assessment of local influence. Journal of the Royal Statistical Society Series B, 48:133–169, 1986.

26. Xie, H. and Heitjan, D. F.: Sensitivity analysis of causal inference in a clinical trial subject to crossover. Clinical Trials, 1:21–30, 2004.

27. Zhang, J. and Heitjan, D. F.: Nonignorable censoring in randomized clinical trials. Clinical Trials, 2:488–496, 2005.

28. Ma, G., Troxel, A. B., and Heitjan, D. F.: An index of local sensitivity to nonignorable dropout in longitudinal modeling. Statistics in Medicine, 24:2129–2150, 2005.

29. Xie, H.: A local sensitivity analysis approach to longitudinal non-gaussian data with non-ignorable dropout. Statistics in Medicine, 27:3155–3177, 2008.

30. Xie, H.: Adjusting for nonignorable missingness when estimating generalized additive models. Biometrical Journal, 52:186–200, 2010.

31. Xie, H. and Heitjan, D. F.: Local sensitivity to nonignorability: Dependence on the assumed dropout mechanism. Statistics in Biopharmaceutical Research, 29:243–257, 2009.

32. Aranda-Ordaz, F. J.: On two families of transformations to additivity for binary response data. Biometrika, 68:357–363, 1981.

33. Xie, H., Qian, Y., and Qu, L.: A semiparametric approach for analyzing nonignorable missing data. Statistica Sinica, 21:1881–1899, 2011.

34. Xie, H.: Analyzing longitudinal clinical trial data with nonignorable missingness and unknown missingness reasons. Computational Statistics and Data Analysis, 56:1287–1300, 2012.

35. Xie, H. and Qian, Y.: Measuring the impact of nonignorability in panel data with non-monotone nonresponse. Journal of Applied Econometrics, 27:129–159, 2012.

36. Vach, W. and Blettner, M.: Logistic regression with incompletely observed categorical covariates investigating the sensitivity against violation of the missing at random assumption. Statistica in Medicine, 14:1315–1329, 1995.

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

38. Zhang, J. and Heitjan, D. F.: A simple local sensitivity analysis tool for nonignorable coarsening: Application to dependent censoring. Biometrics, 62:1260–1268, 2006.

39. Zhang, J. and Heitjan, D. F.: Impact of nonignorable coarsening on bayesian inference. Biostatistics, 8:722–743, 2007.

40. Little, R. J. A.: Regression with missing x's: a review. Journal of the American Statistical Association, 87:1127–1137, 1992.

41. Mermelstein, R., Hedeker, D., Flay, B., and Shiffman, S.: Real-time data capture and adolescent cigarette smoking: Moods and smoking. In The Science of Real-time Data Capture: Self-report in Health Research, eds. A. Stone, S. Shiffman, A. Atienza, and L. Nebeling, pages 117–135. Oxford, Oxford University Press, 2007.

42. Whalen, C. K., Jamner, L. D., Henker, B., and Delfino, R. J.: Smoking and moods in adolescents with depressive and aggressive dispositions: Evidence from surveys and electronic diaries. Health Psychology, 20:99–111, 2001.

VITA

EDUCATION:	BS, JiLin Institute of Technology, Changchun, China, 1998
	MS, Biostatistics, School of Public Health, University of Illinois at Chicago, 2004
	PhD, Biostatistics, School of Public Health, University of Illinois at Chicago, 2018
TEACHING:	Teaching Assistant, BSTT 400 Biostatistics, School of Public Health, University of Illinois at Chicago, 2003
	Teaching Assistant, BSTT 505 Logistic Regression and Survival Analysis, School of Public Health, University of Illinois at Chicago, 2014
RESEARCH:	Research Assistant, School of Public Health, University of Illinois at Chicago, 2015 - 2016

PUBLICATIONS:

Gao, W., Hedeker, H., Mermelstein, R., Xie, H. (2016). A Scalable Approach to Measuring the Impact of Nonignorable Nonresponse with an EMA Application. Statistics in Medicine, 35(30):5579-5602.

Gao, W., Xing, B., Hedeker, H., Mermelstein, R., Xie, H. ISNI: An R Package to Measure the Impact of Nonignorable Missing Data. 2016 JSM contributed poster

Xie, H., Gao, W., Xing, B., Heitjian, D., Hedeker, D., Yuan, C. Measuring the Impact of Nonignorable Missingness Using the R Package ISNI. Computer Methods and Programs in Biomedicine, in press