Measuring the Impact of Nonignorability in Panel Data with Non-Monotone Nonresponse

Yi Qian*

Kellogg School of Management

Northwestern University

and

Hui Xie

Department of Epidemiology & Biostatistics, University of Illinois

Chicago, IL 60612, U.S.A.

September 20, 2009

Summary

The analysis of panel data with non-monotone nonresponse often relies on the critical and untestable assumption of *ignorable* missingness. It is important to assess the consequences of departures from the ignorability assumption. Non-monotone nonresponse, however, can often make such sensitivity analysis infeasible because the likelihood functions for alternative models involve high-dimensional and difficult-to-evaluate integrals with respect to missing outcomes. We develop an extension of the local sensitivity method that overcomes computational difficulty and completely avoids fitting alternative models and evaluating these high-dimensional integrals. The proposed method is ap-

*The authors contributed equally and are listed in alphabetical order. Corresponding Author: Dr. Yi Qian, 2001 Sheridan Road, Evanston, IL 60208. Phone: (847)491-7113. Fax: (847)491-2498. Email: yiqian@kellogg.northwestern.edu.

plicable to a wide range of panel outcomes. We apply the method to a Smoking Trend dataset where we relax the standard ignorability assumption and evaluate how smoking-trend estimates in different groups of U.S. young adults are affected by alternative assumptions about the missing-data mechanism. The main finding is that the standard estimate in the black-male group is sensitive to nonignorable missingness but those in other groups are reasonably robust.

KEY WORDS: Generalized Least Squares; Generalized Linear Mixed Model; Multinomial Logit Model; Nonignorable; Nonlinear Integer Programming; Sensitivity Analysis; Panel Tobit Model

JEL Classifications: C01, C23, C33

1. Introduction

Panel data with missing values are prevalent in social sciences and health studies. Missing data could be due to either attrition (also known as *dropout*) or intermittent missingness. A missing value in a panel is a *dropout* if all of the subsequent values are missing; otherwise, the missing value is *intermittent*. Intermittent missing values lead to a non-monotone pattern of missingness, which is generally more difficult to analyze than if all missing values are caused solely by dropouts. In practice, a dropout may occur if a subject has an adverse event (*e.g.*, death) or formally withdraws from the study; intermittent missingness may occur if a subject misses a visit due to an illness, forgetfulness or unwillingness to make the trip. Both types of nonresponse occur frequently in household panel surveys, longitudinal experimental studies, and even scanner databases. The subjects' behavior underlining their nonresponse often reflects their attitude toward the treatment and their characteristics relevant to the outcome. This nonresponse behavior, if not properly accounted for, would lead to self-selection bias (Heckman and Hotz, 1989).

Thus, an important step in analyzing data subject to missingness is deciding whether the missing data is *ignorable*. A necessary condition for ignorability is *missing at random* (MAR). When the probability of missingness is independent of the unobserved values in the data matrix given the observed values of all the variables, the missingness mechanism is MAR (Rubin 1976, Heitjan and Rubin 1991). Under MAR and parameter distinctness (Rubin 1976), missing data is ignorable and one can perform a valid likelihood-based inference without positing a model for the missing data process. In practice, many panel datasets with non-monotone missingness have been analyzed under the ignorability assumption using imputation or likelihood-based methods. Because the MAR assumption is a critical assumption that is untestable without using additional data or introducing other untestable assumptions (Little 1995; Nicoletti 2006), it is important to assess the credibility of the MAR analysis by evaluating its sensitivity to departures from the MAR assumption.

In this paper, we provide a tractable approach to evaluating the maximum sensitivity of MAR inferences within a class of nonignorable selection models for panel data with an arbitrary pattern of missingness. This class of nonignorable models combines a model for the complete-data-generating process with a transitional multinomial logit model for the missing-data process. In the models, the unobservable is related to the probability of missingness through the nonignorability parameter γ_1. When γ_1 is zero, the missing data mechanism is MAR. Without using additional data or imposing other untestable assumptions, γ_1 generally cannot be reliably estimated together with other model parameters. Thus, we follow a more careful approach, which compares the inferences over a plausible range of γ_1 values, and investigates how sensitive the inference is to departures from $\gamma_1 = 0$. Although conceptually straightforward, such an analysis is very difficult to conduct with non-monotone missingness. For any fixed non-zero γ_1, the likelihood function often involves high-dimensional and difficult-to-evaluate integrals with respect to missing data. Moreover, these integrals need

to be numerically evaluated many times due to the need to repeatedly fit the nonignorable models. We adopt a local sensitivity approach that uses a Taylor-series approximation to the likelihood. Our approach completely avoids fitting the nonignorable models and evaluating those high-dimensional integrals. Only estimates from the simpler MAR analysis are needed to evaluate sensitivity. As a result, our method substantially reduces computational cost and provides a practical approach to delineating which MAR estimates in a panel regression model are susceptible to nonignorable non-monotone nonresponse, and which ones are not.

Prior literature on sensitivity analysis to nonignorable missingness is extensive (Vach and Blettner 1995, Rotnitzky *et al.* 1998, Scharfstein *et al.* 1999). A more recent approach is to study the local sensitivity to selection bias in the neighborhood of the ignorable model (Copas and Li 1997, Copas and Eguchi 2001 and Verbeke *et al.* 2001). Troxel *et al.* (2004) developed an *index of local sensitivity to non-ignorability* (ISNI) to evaluate the possible effect of nonignorability on MLEs. The method has been further applied to the potential outcome model (Xie and Heitjan 2004), censored data and the Bayesian analysis (Zhang and Heitjan 2006, 2007). Ma *et al.* (2005) and Xie (2008) extended the ISNI method to panel data with attrition. Their methods require that missingness be monotone. The requirement is restrictive because, in practice, panel data often suffer from both attrition and intermittent missingness, leading to non-monotone missingness. Moreover, it is often suspected that intermittent missingness is also nonignorable. In this paper, we extend the ISNI method to various types of panel outcomes with non-monotone missingness and derive formulas of the extended ISNI for a range of panel models. Our results include Ma *et al.* (2005) and Xie (2008) as special cases.

Our empirical application is concerned with studying seven-year smoking trends in U.S. young adults using a unique dataset. Smoking-related issues have been at the center of many health, economic and political debates. Although the tobacco industry contributes

significantly to the economy through employment and tax revenue, cigarette smoking has been found to be the leading cause of lung cancer and many other serious chronic diseases. The resulting morbidity and mortality costs impose a heavy economic and social burden on our society. In order to decrease smoking, legislative and economic regulations, such as prohibiting smoking in public and imposing cigarette taxation,[1] have been adopted over the past decades. On the other hand, the tobacco industry has adapted its marketing strategies, such as price promotion, market segmentation, and more targeted advertising, to maintain sales level. Therefore, it is of substantial interest to assess the changes in smoking behavior among different groups of the population under these conflicting forces. The Smoking Trend dataset is an individual-level panel dataset that enables the study of this question (Preisser *et al.* 2000). As in many other panel datasets, the missing-data issue occurs in this dataset. Previous analyses of this dataset (Preisser *et al.* 2000) relied on the MAR assumption when handling non-monotone missingness. We contribute to this literature by relaxing the MAR assumption and investigating how smoking-trend estimates would be affected by moderate nonignorable missingness. Because the type of missingness (dropout versus intermittent missingness) is unknown to us for some missing observations in the dataset, we use a bound approach to quantifying the uncertainty in the sensitivity analysis due to unobservability of the missingness type. To the best of our knowledge, this is the first attempt in the literature to systematically address the problem of not observing missingness type in missing data analysis.

The rest of the paper is organized as follows. Section 2 describes our nonignorable selection model for panel data with non-monotone missingness. Section 3 derives the extended ISNI with details provided in Appendix I, and specific formulas in Appendix II for several

[1] Cigarette taxes have been constantly increasing. The single largest federal tobacco tax increase from \$0.39 to \$1.01 per pack became effective recently, on April 1, 2009. This has provoked a hot debate. See the CNN i-Report (http://www.ireport.com/ir-topic-stories.jspa?topicId=173499) for a wide array of drastically different reactions to the tax hike.

popular ideal-data models. These models cover various types of outcomes, and are frequently used for panel data analysis. In Section 4 we apply the proposed methodology to the Smoking Trend data to quantify the impact of nonignorable missingness on the MAR smoking-trend estimates. In Section 5, we conduct a simulation study that demonstrates the validity and computational advantage of the extended ISNI. We end with a discussion in Section 6.

2. Model and Notation

We assume the following nonignorable selection model for a panel study subject to non-monotone missingness.

2.1 Ideal-data model: Let $Y_i = (Y_{i1}, ..., Y_{in_i})$ denote the vector of ideal panel outcomes for subject i, where Y_{ij} is the outcome for subject i at time j, $i = 1, ..., N, j = 1, ..., n_i$. We assume that the density function of Y_i is $f_\theta(Y_i|X_i)$, where θ is the vector of the parameters of interest with length p_θ, and X_i is the matrix for fully observed covariates. In this paper, we illustrate our method in the following three popular classes of panel models for Y_i. These models can be used to analyze a wide range of panel outcomes in smoking studies or similar studies in other fields.

Example 1: Marginal Multivariate Gaussian Model. For a continuous panel outcome, the following marginal model is often considered:

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

where θ_1 and θ_2 are the vector parameters of the population mean model and the variance-covariance matrix, respectively; the variance-covariance matrix $\Sigma_i(\theta_2)$ is unrestricted except for being symmetric and positive definite. The estimation of the model parameters is usually done through the generalized least-squares method.

Example 2: Generalized Linear Mixed Model (GLMM). This model is applicable to a wide variety of types of non-Gaussian panel outcomes, such as binary, binomial, Poisson, Gamma and Inverse-Gaussian outcomes. The model assumes that Y_{ij}, given a subject-specific random variable b_i, follows independently a distribution in the exponential family:

$$\begin{aligned} f_{\beta,\tau}(y_{ij}|b_i, X_{ij}) &= \exp\{[y_{ij}\lambda_{ij}(\beta, b_i) - \Psi(\lambda_{ij}(\beta, b_i))]/a(\tau) + c(y_{ij}, \tau)\}, \\ \text{with} \quad \mu_{ij}^Y|(b_i, X_{ij}) &= h_Y(\eta_{ij}) = h_Y(X_{ij}^T\beta + Z_{ij}^T b_i) \\ \text{and} \quad b_i|X_i &\overset{i.i.d}{\sim} MVN(0, \Sigma(\Omega)), \end{aligned}$$

where λ_{ij} is the canonical parameter; functions $\Psi(\cdot)$ and $c(\cdot)$ are determined by a particular distribution; $a(\tau) = \tau/w$, where τ is the dispersion parameter and w is a known weight; $\mu_{ij}^Y|(b_i, X_{ij})$ is the conditional mean of Y_{ij} given b_i and X_{ij}; $h_Y(\cdot)$ is the inverse function of a known differentiable monotonic link function; X_{ij} is the jth row of X_i; Z_{ij}, the covariates for the random effects b_i, is a subset of X_{ij}; β is the fixed-effects parameter; Ω is a vector that collects the unique parameters in the variance-covariance matrix $\Sigma(\Omega)$. When making an inference, the following marginal likelihood that integrates out random effects b_i is often used:

$$f_\theta(y_i|X_i) = \int \prod_{j=1}^{n_i} f_{\beta,\tau}(y_{ij}|b_i, X_{ij}) f_\Omega(b_i) db_i, \tag{1}$$

where $\theta = (\beta, \tau, \Omega)$.

Example 3: Panel Tobit Model. Corner solution outcomes, such as money spent on smoking and drug use or hours worked annually by married women, often arise in economics and public health studies. The outcome variable is zero with a positive probability and is a continuous variable on positive values. A standard technique for this type of outcome is the

Tobit model (Tobin 1956, Amemiya 1985). We assume the following standard form of the panel Tobit model (Wooldridge 2002) for Y_i:

$$\begin{aligned} Y_{ij}|b_i, X_{ij} &= \max(0, X_{ij}^T\beta + Z_{ij}^T b_i + \epsilon_{ij}) \\ b_i|X_i &\overset{i.i.d}{\sim} MVN(0, \Sigma(\Omega)); \quad \epsilon_{ij}|b_i, X_i \overset{i.i.d}{\sim} N(0, \sigma^2). \end{aligned}$$

The conditional density function of each observation is given as

$$f_{\beta,\sigma}(y_{ij}|b_i, X_{ij}) = \left\{1 - \Phi((x_{ij}^T\beta + z_{ij}^T b_i)/\sigma)\right\}^{I(y_{ij}=0)} \left\{\phi((y_{ij} - x_{ij}\beta - z_{ij}b_i)/\sigma)/\sigma\right\}^{I(y_{ij}>0)} \quad (2)$$

where Φ and ϕ are cumulative and density functions of the standard normal distribution, respectively; $I(\cdot)$ is one if the argument is true and zero if otherwise. The marginal likelihood for Y_i integrates out the random effects b_i from this conditional density function in a similar manner to Equation (1).

With non-monotone nonresponse, one can fit the above panel models with observed data only. When inference is likelihood-based, such analysis is valid if missingness is ignorable. That is, under ignorability, a valid likelihood-based inference does not require explicitly modeling how data is missing. This is appealing because a missing-data mechanism is often a nuisance, and investigators would like to avoid modeling it if possible. It would then be useful to have a simple and effective method to assess whether it is appropriate to do so. To evaluate the potential effect of nonignorable non-monotone nonresponse, we posit the following model for an alternative missing-data mechanism.

2.2 Missing-data model (MDM): Let G_{ij} denote the missingness status of subject i at time j, and

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

Let u index the status of missingness and $u = 0, 1$, or 2. Let $G_{iju} = 1$ if $G_{ij} = u$, and $G_{iju} = 0$ otherwise. We then assume the following first-order transitional multinomial logit model for the random variable G_{ij} where G_{ij} depends on the history of missingness status only through $G_{i,j-1}$:

$$\begin{aligned}
(G_{ij0}, G_{ij1}, G_{ij2}) | G_{i,j-1} = v, S_{ij} = s_{ij}, Y_{ij} = y_{ij} &\sim Multinomial(1, \left[P_{ij}^{0v}, P_{ij}^{1v}, P_{ij}^{2v}\right]); \\
G_{ij} &= \sum_{u=0}^{2} u * I(G_{iju} = 1) \\
P_{ij}^{uv} &= \frac{\phi_{ij}^{uv}}{\sum_{U=0}^{2} \phi_{ij}^{Uv}}, \quad \text{where } u = 0, 1, 2; v = 0, 1 \\
\text{and } \phi_{ij}^{uv} &= \exp(\gamma_0^{uv} s_{ij} + \gamma_1^{uv} y_{ij}) \qquad (3)
\end{aligned}$$

and by the definition of dropout, $G_{ij} = 2$ deterministically when $G_{i,j-1} = 2$ (the prior visit is a dropout); by the definition of intermittent missingness, $\phi_{ij}^{21} = 0$; and because the response probabilities must add up to unity, for identification purposes $\phi_{ij}^{00} = \phi_{ij}^{01} = 1$; S_{ij} is a vector of fully observed predictors for missingness at time j for subject i, and S_{ij} commonly includes the history of the predictors X in the ideal-data model up to and including time j as well as the history of the observed prior outcomes (*i.e.*, the observed elements in $(Y_{i1},...,Y_{i,j-1})$).

The above MDM can be extended in various ways. For example, the MDM as specified in Equation (3) is a first-order transition model where the probability of missingness depends on the history of missingness only through the immediately preceding missing-data variable, $G_{i,j-1}$. One can extend it to a Markov model of order q where G_{ij} depends on the history of missingness through $(G_{i,j-1}, ..., G_{i,j-q})$. In the Markov model of order q, the transition probability

$$P_{ij}^{uv_1 \ldots v_q} = \phi_{ij}^{uv_1 \ldots v_q} / \sum_{U=0}^{2} \phi_{ij}^{Uv_1 \ldots v_q}, \qquad (4)$$

where $\phi_{ij}^{uv_1,\ldots,v_q} = \exp(\gamma_0^{uv_1 \ldots v_q} s_{ij} + \gamma_1^{uv_1 \ldots v_q} y_{ij})$, v_l is a possible value of $G_{i,j-l}, l = 1, ..., q$. To complete the model, one would also need to posit a model for initial probabilities

$P(G_{i1}, ..., G_{iq})$. It is also possible to extend the model in other directions. For example, when S_{ij} contains continuous variables, it might be preferable to model their functional forms nonparametrically by using the generalized additive model (Qian and Xie 2009). Of course, such high-dimensional models would be considerably more difficult to fit and the results may be more difficult to interpret.

The model as specified in Equation (3) assumes that the probability of missingness at time j, given the missingness status at the previous visit, depends on the covariates, the prior observed outcomes, and the current outcome. Let $\gamma_1 = (\gamma_1^{10}, \gamma_1^{20}, \gamma_1^{11})$. In this model, the potentially unobserved outcome y_{ij} affects the probability of missingness through the parameter γ_1. When $\gamma_1 = 0$, the missing-data mechanism is then MAR and the missingness depends only on the fully observed values, $(G_{i,j-1}, S_{ij})$. Thus, the model nests the resulting parametric MAR model as a special case. If, additionally, the parameter of interest, θ, is distinct from the parameter $\gamma = (\gamma_0, \gamma_1)$ in the missing-data model (*i.e.*, parameter distinctness), then the missingness is ignorable. Note that in our MAR model, the missingness can depend on the past observed outcomes. For example, in our application we allow missingness to depend on the most recently observed past outcome. Hausman and Wise (1979) developed a selection model for panel data with attrition, and Albert (2000) developed a selection model for dynamic panel binary outcomes with intermittent missingness. Both prior models allow the missingness to depend only on contemporaneous outcomes but not on the past outcomes, so their models do not necessarily nest MAR as a special case. Our MDM is more general in this aspect of nesting MAR, and can be viewed as an extension of dropout models in Diggle and Kenward (1994), Hirano *et al.* (2001), Ma *et al.* (2005) and Xie (2008) to allow for intermittent missingness.

The model does require that the missingness not depend on the past *unobserved* outcomes, *given* the missingness status at the previous visits, the covariates, the prior observed

outcomes and the potentially unobserved current outcome. This is a much weaker assumption than that the missingness is independent of the past unobserved outcomes. In fact, our missing data model allows the missingness to depend on the past unobserved outcomes marginally, through affecting the missingness status at the prior visits, and through their correlations with the past observed outcomes as well as the potentially unobserved current outcome. This implies that the past unobserved outcomes can affect the missingness status at a current visit through affecting the missingness status at prior visits. Given the same history of past missingness status, the past unobserved outcomes can still affect the current missingness status through their correlations with the past observed outcomes (*e.g.* the most recently observed outcome). Our approach is to use the relatively parsimonious and reasonable model as specified in Equation (3) or (4) for an easily interpretable sensitivity analysis. Note that the assumption of the conditional independence between the missingness and the past unobserved outcomes is not required if the missingness is caused solely by dropouts. This is one manifestation of the fact that non-monotone nonignorable missingness is more challenging to analyze than monotone nonignorable missingness, and more assumptions regarding the missing-data mechanism may be required.

3. Methodology

Let $Y_i = (Y_i^{\text{obs}}, Y_i^{\text{mis}})$, where Y_i^{obs} refers to the components of Y_i that are observed, and Y_i^{mis} refers to the components of Y_i that are missing. Let K_i be the length of Y_i^{obs}. If subject i completed all the intended visits of the study, then $K_i = n_i$ and Y_i^{mis} vanishes; otherwise, $K_i < n_i$. Let L be the correct log-likelihood for $(\theta, \gamma_0, \gamma_1)$ under the above nonignorable

selection model. Then

$$
\begin{aligned}
L(\theta,\, \gamma_0,\, \gamma_1) &= \sum_{i=1}^{N} L_i\left(\theta,\, \gamma_0,\, \gamma_1; y_i^{\text{obs}}, g_i\right) \\
&= \sum_{i=1}^{N} \ln\Big(\int \prod_{j=1}^{n_i} f_\gamma\left(g_{ij}|s_{ij}, y_{ij}, g_{i,j-1}\right)\; f_\theta(y_i^{\text{obs}}, y_i^{\text{mis}}|x_i) dy_i^{\text{mis}}\Big) \\
&= \sum_{i=1}^{N} \ln f_\theta(y_i^{\text{obs}}|x_i) + \\
&\quad \sum_{i=1}^{N} \sum_{j:g_{ij}=0} \ln f_\gamma(g_{ij}|s_{ij}, y_{ij}, g_{i,j-1}) + \\
&\quad \sum_{i:K_i<n_i} \ln\left(\int \prod_{j:g_{ij}\neq 0} f_\gamma(g_{ij}|s_{ij}, y_{ij}, g_{i,j-1}) f_\theta(y_i^{mis}|y_i^{\text{obs}}, x_i) dy_i^{mis}\right), \qquad (5)
\end{aligned}
$$

where $g_i = (g_{i1}, ..., g_{in_i})$ is a vector of discrete variables for the missingness status of the ith subject; $f_\theta(y_i^{\text{obs}}, y_i^{\text{mis}}|x_i)$ is the density function of the ideal-data model defined above; $f_\gamma\left(g_{ij}|s_{ij}, y_{ij}, g_{i,j-1}\right)$ is the density function of the missing-data model defined in Equation (3), and if a higher-order ($q > 1$) Markov model is used for modeling G_{ij}, $f_\gamma\left(g_{ij}|s_{ij}, y_{ij}, g_{i,j-1}\right)$ is then replaced with $f_\gamma\left(g_{ij}|s_{ij}, y_{ij}, g_{i,j-1}, \cdots, g_{i,j-q}\right)$; and the integral will be replaced with summation for discrete outcomes. It is readily observable that the components of y_i^{mis} after dropout do not enter the integral in Equation (5) because these outcomes are deterministically missing. Thus, the dimensionality of the integration for the ith unit is $d_i = \sum_j I(g_{ij} = 1) + I(\text{any of } g_{ij} \text{ is } 2)$. Henceforth, the notation y_i^{mis} includes only the intermittent missing outcomes and the outcome at the time of dropout. With nonignorable missingness, the integral with respect to y_i^{mis} does not have a closed form and a numerical method is required for its evaluation. The computational workload for such numerical integration increases exponentially with the number of intermittent missing outcomes and makes the evaluation of the log-likelihood L difficult even with a moderate amount of intermittent

missingness.[2]

More importantly and beyond the aforementioned computational difficulty, numerous studies have shown that the direct-estimation approach, which maximizes the log-likelihood L, is often unreliable unless some instrumental variables for nonresponse are included in the model (Schluchter 1992; Diggle and Kenward 1994; Copas and Li 1997; Troxel *et al.* 1998; Nicoletti 2006)[3]. Without good instruments, the observed data provide little information for distinguishing between different missingness mechanisms. In some cases, the model is unidentifiable; even when the model is identified, the resulting inference is highly sensitive to the strong model assumptions that are untestable with the observed data (Little 1995; Kenward 1998).

Given the difficulties associated with relying on L for a definite answer, an alternative and more prudent approach is to perceive the above nonignorable model as provisional and to perform a sensitivity analysis over a plausible range of values for γ_1. For any given value of γ_1, we must obtain $(\hat{\theta}(\gamma_1), \hat{\gamma_0}(\gamma_1))$ by maximizing L over (θ, γ_0), and compare inferences for the range of γ_1 values. Computational cost is, however, still high because the problem of repeatedly evaluating high-dimensional integrals with respect to Y_i^{mis} still exists. A feasible approach that overcomes this difficulty uses a Taylor-series approximation as follows:

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

In this approximation, both $\hat{\theta}(0)$ and $\left.\frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_1^T}\right|_{\gamma_1=0}$ are easy to obtain: $\hat{\theta}(0)$ is obtained by maximizing only $\sum_{i=1}^{N} \ln f_\theta(y_i^{\text{obs}}|x_i)$; as shown below, $\left.\frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_1^T}\right|_{\gamma_1=0}$ also does not require evaluating the integrals with respect to Y_i^{mis}. This allows for a simple and fast evaluation of $\hat{\theta}(\gamma_1)$. Troxel *et al.* (2004) used a binary regression model to model the dichotomous missingness

[2] For example, Troxel *et al.* (1998) found that the computations are intractable for more than three time points, due to multiple integration caused by intermittent missingness.

[3] Instrumental variables also rely on some untestable assumptions for which a sensitivity analysis has been proposed (Ashley 2009)

status of an univariate outcome, and derived a general formula for $\frac{\partial\hat{\theta}(\gamma_1)}{\partial\gamma_1}\Big|_{\gamma_1=0}$ under their model. It can be shown that in our case where a multinomial logit model is employed for non-monotone nonresponse in panel data and the log-likelihood L is given in Equation (5), the general formula is also applicable, such that

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

where

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

and $\hat{\theta}(0)$ and $\hat{\gamma}_0(0)$ are MLEs of θ and γ_0 under the MAR assumption. The first term, $\nabla^2 L_{\theta,\theta}$, is the observed Hessian matrix of the ignorable model, which is usually readily available. The second term evaluates the orthogonality of θ and γ_1.

In our case where we employ a multinomial logit model for non-monotone missingness, we further derive the second term in the above general formula to be (see Appendix I for details)

$$\begin{aligned}\nabla^2 L_{\theta,\gamma_1} &= (\nabla^2 L_{\theta,\gamma_{10}}, \nabla^2 L_{\theta,\gamma_{20}}, \nabla^2 L_{\theta,\gamma_{11}}) \\ &= \sum_{i:K_i<n_i} \left.\frac{\partial E((y_i^{mis})^T|y_i^{obs},x_i)}{\partial\theta}\right|_{\gamma_1=0} \cdot A_i, \end{aligned} \tag{6}$$

where $\frac{\partial E((y_i^{mis})^T|y_i^{obs},x_i)}{\partial\theta}$ is a $p_\theta \times d_i$ matrix for derivatives of the conditional mean of missing outcomes given the observed data with respect to the parameters of the ideal-data model; $A_i = [A_i^{10}, A_i^{20}, A_i^{11}]$ is a $d_i \times 3$ matrix. Suppose the lth ($l = 1, ..., d_i$) component of y_i^{mis} corresponds to the jth element of y_i; the lth elements of $A_i^{10}, A_i^{20}, A_i^{11}$ are then shown in

Appendix I to be

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

It is interesting to note that A_{il}^{10} and A_{il}^{20} are of opposite signs when $g_{i,j-1}=0$. This can be explained by the fact that intermittent missingness and dropout are two competing reasons for being unobserved.

In Troxel *et al.* (2004), γ_1 was a scalar and the derivative $\left.\frac{\partial\hat{\theta}(\gamma_1)}{\partial\gamma_1}\right|_{\gamma_1=0}$ was labeled ISNI, the *index of local sensitivity to nonignorability*. That is, ISNI was defined as the change of an estimate when γ_1 is perturbed from 0 to 1. In our case, $\gamma_1=(\gamma_{10},\gamma_{20},\gamma_{11})$ is a vector. When each element of γ_1 is perturbed between -1 and 1, the possible values of γ_1 form a hypercube in space. We can consider the maximum sensitivity in this hypercube. That is, we define in our case

$$\begin{aligned} \text{ISNI}(\hat{\theta}) &= \sum_{i=1}^{q}\left|\frac{\partial\hat{\theta}(\gamma_1)}{\partial\gamma_{1i}}\right|_{\hat{\theta}(0),\hat{\gamma}_0(0),\gamma_1=0} \\ &\approx \max_{|\gamma_{1i}|=1,\ i=1\ to\ q}\left(\hat{\theta}(\gamma_1)-\hat{\theta}(0)\right), \end{aligned} \tag{7}$$

where q is the length of the γ_1 vector and $q=3$. The above defined ISNI allows the intermittent missingness and dropout to have different or even opposite nonignorable missingness mechanisms. When an investigator is willing to assume that the intermittent missingness and dropout have roughly the same nonignorability mechanism, one can fix $\gamma_1^{10},\gamma_1^{20}$ and γ_1^{11} at the same largest perturbation value. In this special case, γ_1 becomes a scalar, A_i becomes a vector of length d_i, and

$$\begin{aligned} A_i &= A_i^{10}+A_i^{20}+A_i^{11}, \\ \text{where}\ \ A_{il} &= \left[P_{ij}^{0,g_{i,j-1}}\right]_{\hat{\gamma}_0(0),\gamma_1=0}. \end{aligned} \tag{8}$$

Thus, the lth component of A_i becomes the propensity of being observed for the missing observation predicted under the ignorability assumption. We also note here that in the special case when there is no intermittent missingness and missingness is caused solely by dropout, the algorithm further simplifies: $A_i^{10} = A_i^{11} = 0$, and $\frac{\partial E((y_i^{mis})^T|y_i^{obs},x_i)}{\partial\theta}$ reduces to a $p_\theta \times 1$ vector. The calculation of $\nabla^2 L_{\theta,\gamma_1}$ then reduces to the results provided in Ma *et al.* (2005) and Xie (2008), where a binary regression model was employed to model the dropout mechanism.

There are two features in the computation of the extended ISNI that substantially reduce the computational burden, as compared to the global sensitivity analysis. Firstly, our algorithm is non-iterative as it avoids repeatedly optimizing the complicated likelihood of nonignorable models as needed in the global sensitivity analysis. Thus, it also avoids repeatedly evaluating the potentially high-dimensional integrals with respect to Y_i^{mis}. Because we employ a Taylor series expansion at the ignorable model, the optimization is only needed for fitting an MAR ideal-data model and an MAR missing-data model separately, neither of which involves such integrals. This feature of the ISNI statistics holds for any likelihood-based nonignorable model. Essentially, ISNI assesses sensitivity to nonignorability without the need to explicitly estimate alternative nonignorable models. In this respect, ISNI is very much like a score test statistic.[4]

Secondly, once the MAR model estimates are obtained, computing the ISNI value is relatively simple. In particular, $\frac{\partial E((y_i^{mis})^T|y_i^{obs},x_i)}{\partial\theta}$ in Equation (6) is much simpler to evaluate under MAR than under the more general nonignorable model. Note that $\frac{\partial E((y_i^{mis})^T|y_i^{obs},x_i)}{\partial\theta}$ quantifies the sensitivity of parameter estimates attributable to the missing data y_i^{mis}. It is larger in absolute value for the missing observations whose conditional means given the observed data are most affected by the values of the model parameters. This is intuitive since

[4] We thank Dr. Daniel Heitjan for pointing out this analogy.

such missing observations tend to cause sensitivity of parameter estimates. This quantity is difficult to calculate under the general nonignorable case. One can show that the conditional density function of y_i^{mis} given the observed data contains the density functions from both the ideal-data model and the missing-data model. Thus, $\frac{\partial E((y_i^{mis})^T|y_i^{obs},x_i)}{\partial\theta}$ has no closed form and requires numerical evaluations with respect to the missing data (and the random effects if there is any in the ideal-data model). Under MAR, the missing-data model does not depend on the missing data and its density function can be factored out of the expectation with respect to y_i^{mis}. Thus, the calculation of $\frac{\partial E((y_i^{mis})^T|y_i^{obs},x_i)}{\partial\theta}$ is considerably simplified. For example, when the ideal data (Y_i^{mis}, Y_i^{obs}) follows a multivariate normal distribution, $y_i^{mis}|(y_i^{obs}, x_i)$ under MAR is again a multivariate normal distribution with its mean having a closed form. Thus, ISNI in this case merely involves the evaluations of closed-form derivatives. For nonlinear models, as shown in Examples 2 and 3, the calculation of the above conditional mean and its derivative can be shown to involve integration with respect to the random effects but again no integration with respect to the missing data. Appendix II derives the explicit formula for it in our three examples of ideal-data models: the marginal multivariate Gaussian model, the GLMM and the panel Tobit model. These models cover a wide range of models used for panel data analysis. One might, however, encounter a model, other than the models presented in our paper, where such simplicity in calculating $\frac{\partial E((y_i^{mis})^T|y_i^{obs},x_i)}{\partial\theta}$ may not hold. In that case, one will need numerical evaluations with respect to the missing data. Note that such evaluation is only required once in the ISNI calculation, unlike in the global sensitivity analysis, where numerical integration is required repeatedly in the optimization process.

4. ISNI Analysis in the Smoking Trend Dataset

In this section, we apply our proposed method to the Smoking Trend dataset. As explained in the Introduction, the topic under examination here is of frequent interest in health economics and in public health studies. Preisser *et al.* (2000) analyzed a dataset to study seven-year trends in cigarette-smoking rates among the four race/sex groups in the United States. The study enrolled 5,115 young adults aged 18-30, and these participants were followed prospectively over time at years 0 (year 1986), 2, 5 and 7. The variable of most interest is the self-reported cigarette smoking status (yes/no) at the four exams from 1986 to 1993. The analysis sample consisted of 5,078 subjects whose smoking status was known at the baseline (year 0). The goal of the statistical analysis is to draw inferences on the time trend in smoking rates by race/sex. Table 1 summarizes the smoking rates by year and group. It shows that the smoking rate based on the observed data decreases in the seven-year period by 0.1, 1.8, 4.7 and 6.8 in percentage points for black males, black females, white males and white females, respectively. However, the dataset has a fair amount of missingness, as shown by the decreasing number of N in Table 1. Moreover, the missingness pattern is non-monotone, as shown in Table 2. Preisser *et al.* presented analyses of this dataset that are valid when data are missing at random. In their discussion section, they recommended conducting sensitivity analysis to evaluate the impact of the potential nonignorable missingness on MAR analyses. Here we apply our method to this dataset to measure the impact of nonignorable non-monotone missingness.

Let $Y_{ijk} = 1$ if the ith person in the kth race/sex group is a smoker at time j, and 0 if the participant is a nonsmoker. For the ideal complete binary outcome of smoking, we assume the following GLMM:

$$\begin{aligned} logit(P(Y_{ijk} = 1)) &= logit(\mu^Y_{ijk}) = \beta_{0k} + \beta_{1k} j + b_{0i} \\ b_{0i} &\sim N(0, \sigma^2_{b_0}), \end{aligned} \tag{9}$$

where $j = 0, 2, 5, 7$ for the years since baseline measurement and $k = 1, 2, 3, 4$ for the four race/sex groups. Preisser *et al.* (2000) mainly used a weighted generalized estimating equation (WGEE) approach for inference that is not likelihood-based. As shown below, the GLMM model gives qualitatively similar conclusions regarding the smoking trends under the MAR assumption. Unlike the WGEE approach, a valid likelihood-based GLMM inference does not require specifying a missing-data model when data is missing at random.

To evaluate the impact of nonignorable missingness, we augment the above ideal-data model with a Markov multinomial logit model of order 2 for missingness. [5] Given the missingness status at the previous two visits, $G_{i,j-1} = v_1, G_{i,j-2} = v_2$, the transitional probability to $G_{ij} = u$ is given as Equation (4), where

$$\begin{aligned} \phi_{ijk}^{uv_1v_2} &= \exp((Intercept, Time, Race, Gender, Race \times Gender, \\ &\quad Gender \times Time, Y_{ijk}^{MOP})^T \gamma_{0k}^{uv_1v_2} + Y_{ijk}\gamma_{1k}^{uv_1v_2}), \end{aligned} \tag{10}$$

$j = 5, 7$, $Time$ is a dummy for year 7, and Y_{ijk}^{MOP} is the most recently observed past outcome prior to time j. Missingness at $j = 2$ is modeled separately as a logistic regression model where the regressors include all the variables in the above except for the terms containing the variable $Time$. We do not need a model for $j = 0$ since all the observations were observed at baseline. Because in this study it was unlikely for intermittent missingness and dropout to have opposite signs in the nonignorability parameter, we fix $\gamma_{1k}^{uv_1v_2}$ all at the same largest perturbation value γ_{1k} in the following sensitivity analysis, where we discuss the choice of γ_{1k} below. The resulting model still allows the nonignorable parameters γ_{1k} to vary across

[5] Note that the order of 2 is the highest one that the Markov model can have for the dataset. Although the missingness at the last visit (the 4th visit) can nominally condition on the missingness status at the past three visits, this is equivalent to conditioning on the missingness status at the past two visits since all the observations were observed at the baseline visit. As a robustness check, we have also used a Markov model of order 1. The changes of ISNI values are all within 5% of those reported in Table 3 and do not qualitatively alter the conclusions of sensitivity analysis. We find that changing the order of the Markov model reduces the predicted probabilities of being observed for some groups of missing observations and increases the probabilities for some other groups of missing observations. In the end, the ISNI values are not sensitive to the order of the Markov model.

the four race/sex groups. It can be readily verified that the derived ISNI formula applies with A_i in Equation (8) being replaced by the fitted probabilities of being observed obtained from the above missing-data model under the ignorability assumption.

To apply the above missing-data model, we need to classify missingness to either dropout or intermittent missingness. For a well-conducted study that has good record keeping, there usually exists information on the reasons for missingness.[6] One can use such information to distinguish between dropout and intermittent missingness. When the true missingness types are discovered, the above missing-data model can be directly applied.

Because we do not have access to the detailed documentation on the reasons for the missingness in our empirical application, not all the missingness types are known to us. For the missing-data patterns in Table 2, the missing values in "XX.X", "X.XX", "X..X" as well as in the second visit of "X.X." were clearly intermittently missing. It is, however, unknown how to classify the missingness in "XX.." and "X..." as well as the last visit of "XXX." and "X.X.". These sequences of missing observations that are of unknown missingness types all occurred consecutively toward the end of the study with no follow-up visits. Practical analyses sometimes, if not frequently, assume that such a sequence of missingness was caused by attrition (*i.e.*, dropout). Because in this dataset the number of participants who returned to the study after two consecutive instances of missingness was small (less than 1%), it may seem reasonable to classify the missingness in "XX.." and "X..." as a dropout. There is, however, ambiguity with respect to the last visit of "XXX." and "X.X.". Alternatively, one can consider the missing values in "XX.." and "X..." as well as the last visit of "XXX." and "X.X." as potentially intermittent missingness, where if one participant missed the prior visit, he/she had some probability of returning to the study in the current visit.

[6] For example, the study participants who formally withdrew from the study provided formal notices. Whether or not a participant died during the study can be checked either through a public database for death certificates, or through relatives. With intermittent missingness, the reasons for missingness can be recorded during follow-up visits. See Albert (2000) for an application that use the reasons for missingness to distinguish between dropout and intermittent missingness.

Let $r = (r_1, \cdots, r_{n_s})$, where n_s is the total number of such sequences of unknown missingness types and r_i $(i = 1, \cdots, n_s)$ is a binary variable, with r_i being 1 if the missingness in the ith sequence was intermittent and r_i being 2 if the missingness in the ith sequence was attrition. We view this as a missing-data problem since we do not know the true value of r_i. To address the issue of unknown missingness types, we conduct a bound analysis, which considers the lower (upper) bound of the range of the ISNI values obtained from all plausible values of r. Formally the bounds are the solutions to

$$\underset{r=(r_1,\cdots,r_{n_s})\in R}{minimize(maximize)} : ISNI_r(Q(\theta)), \text{ subject to } r_i \in (1,2) \ \forall i,$$

where the objective function, $ISNI_r(Q(\theta))$, is the ISNI for $Q(\theta)$ calculated under a specific value of r, and R is the state space of r (the set of all possible combinations of r_i). $Q(\theta)$ is a scalar function of θ for which ISNI is calculated. For example, $Q(\theta)$ might be an element of θ if θ is a vector. Note that the objective function $ISNI_r(Q(\theta))$ is a nonlinear function of r. Given each configuration of r in the state space R, we can fit the above multinomial logit transitional missing-data model, and then calculate the corresponding ISNI value. In this process, r affects ISNI through A_i in a nonlinear and complicated fashion. Thus, the above optimization task is a nonlinear integer programming problem for a large dimension ($n_s = 1036$ for our dataset).

Although it is fast to calculate ISNI for any given configuration of r, it is still impossible to perform an exhaustive enumeration of ISNI values over the entire large state space R. To search for the upper and lower bounds, we apply a genetic algorithm (GA) which is particularly useful for the type of optimization problem encountered here. GA is a numerical search procedure in which a population of candidate solutions (also known as individuals or phenotypes in evolutionary biology) to an optimization problem evolves in order to find the best one(s). The genetic operators, such as crossover and mutation, are used in each

step to evolve toward better solutions. Judd (1998) has more detailed description of the genetic algorithms. GA has been successfully used in Horowitz and Manski (2006) to obtain solutions in a bound analysis. Note that our bound analysis is different from theirs in that the bound analysis conducted here addresses the issue of not observing the missingness status G, while theirs addresses the issue of not observing the outcome Y.[7] For our application, we have used the R package *rgenoud* that implements the genetic algorithms (Mebane and Sekhon 2007).[8] In our GA optimization, we have used a large population of 1000.

In Table 3, we summarize the MAR analysis and the associated ISNI sensitivity analysis based on the above selection model. The MAR analysis reveals that there was a borderline significant increase in the smoking rate in black males, no significant change in black females, and significant decreases in the white groups. The impact of nonignorable missingness on these slope estimates is quantified by the ISNI values. Because the nonignorability parameter γ_1 is a vector, the ISNI values are calculated in the same manner as in Equation (7). Thus, the ISNI values evaluate the maximal effects of nonignorability when each element γ_{1k} in the γ_1 vector is varied between -1 and 1. To gauge these effects better, it is useful to compare an ISNI value with the standard error of the corresponding estimate. This is equivalent to comparing the bias due to model misspecification with sampling variability. In our case, we define $c = |\frac{SE}{ISNI}|$, where SE is the standard error of an MAR estimate. The c statistic denotes the critical size of the hypercube formed by γ_1 above which the bias due to nonignorable missingness is larger than the sampling error and therefore causes concern. A small c statistic means large local sensitivity as a small magnitude of nonignorability is sufficient to change

[7] Our approach in the application uses ISNI analysis to quantify the uncertainty due to unobservability in Y in a GLMM with non-monotone missingness. We are not aware of a bound analysis on this type of problem. As noted in Horowitz and Manski (2006), their bound analysis can be computationally burdensome for some types of problems, and this is very likely to be the case for the problem considered here.

[8] The package combines genetic algorithms with a derivative-based (quasi-Newton) method to solve difficult optimization problems. It may also be used for optimization problems for which derivatives do not exist. We use the nonderivative version for our optimization problem.

the estimate appreciably. A large value of c statistic means that there is no local sensitivity and sensitivity can only occur for extreme nonignorability. Following Troxel *et al.* (2004), we suggest using $c = 1$ as a cutoff value for important sensitivity as this implies that the bias will be in the same size as the sampling error for a moderate nonignorability ($\max_{k=1}^{4} |\gamma_{1k}| = 1$).

For each MAR estimate in Column A of Table 3, the first (fourth) row of Column C reports the lower (upper) bound of ISNI values; the second and third rows report ISNI analyses based on two extreme classification schemes, where the second row assumes all r_i to be 2 and thus classifies all the unknown missingness types as dropouts (henceforth named as Classification II) and the third row assumes all r_i to be 1 and thus classifies all of them as intermittent missingness (henceforth named as Classification I). Note that under Classification I, all the missing observations were intermittently missing. The missing-data model for Classification I then reduces to a Markov binary logistic regression model, a special case of a Markov multinomial logit model.

Table 3 shows that ISNIs based on Classification I (II) are almost the same as the upper (lower) bounds. Furthermore, ISNIs assuming Classification I are on average about 40% larger than those assuming Classification II. This can be explained by inspecting the likelihood in Equation (5). In Classification II, a missing observation after the dropout time does not enter the likelihood and thus does not contribute sensitivity; in Classification I, the same missing observation, when considered as intermittent missingness, does enter the likelihood and contribute sensitivity. Thus, when information on distinguishing between dropout and intermittent missingness is available, it is preferrable to use the information to better reflect the true missingness behavior.

Using the c statistic, we find that the estimates in the black groups are more sensitive than the estimates in the white groups. Using the lower bound, the c statistics of the slope estimates in black males and black females are 0.6 and 0.9, respectively, indicating that a

moderate nonignorability ($\max_{k=1}^{4} |\gamma_{1k}| \leq 1$) can cause a change in the estimate that is larger than its sampling error. In the white groups, these estimates are comparatively less sensitive because the c statistics in both white groups are larger than one. Using the upper bound, the c statistic for the slope estimate in white females is also found to be substantially smaller than 1 (0.8), indicating that the numerical difference between the lower and upper bound is of a fair proportion of the sampling error.

To investigate whether the inference depends critically on the ignorability assumption, we can vary γ_1 in a plausible range and see how the inference is affected. One can solicit opinions from experts in the field on the likely range of γ_1. In this study, we can consider the plausible and sufficiently wide range of γ_1 to be a hypercube of size 1. This range is reasonable because the sizes of regression coefficient estimates for the most recently observed outcome Y_{ijk}^{MOP} in the MAR missing-data model are all less than 1. Thus, we consider the values of γ_1 within this range as empirically relevant, and then investigate if the conclusions from the MAR analysis still hold under such realistic and moderate nonignorability. This moderate nonignorability implies that the odds of missingness (either due to dropout or intermittent missingness) for a smoker is within 2.7 times larger or smaller than for a non-smoker, conditioning on the same values of race, sex and the missingness status at prior visits as well as the most recently observed previous smoking outcome.

Column D in Table 3 lists the range of estimates when γ_1 is varied within this hypercube. Figure 1 plots the area of the possible estimates in relation to the region of nonsignificance, which is defined as $\pm z_{1-\alpha/2} SE(\hat{\beta}_{1k})$. In the figure, the solid (dashed) frame reflects the range of estimates with the upper (lower) bound of ISNI values. The range of estimates using Classification I (II) is almost the same as the solid (dashed) frame. Despite the numerical differences between the solid and dashed frames, both the upper and lower bounds lead to the same qualitative conclusion as follows. We found that the trend estimate for black

males has substantially larger variability due to nonignorability than the trend estimates for the other three groups. As shown in Figure 1, the range of its estimates is wide, and encompasses both highly positive and statistically significant values as well as the close-to-zero and statistically insignificant values. Thus, we conclude the MAR inference for the smoking trend in black males is sensitive to moderate nonignorability. On the other hand, we find that the inferences in the other three groups are quite robust. Their ranges of estimates are considerably narrower. Moreover, the estimates in the two white groups all remain negative and statistically significant; the estimates in the black female group are all well contained in the region of nonsignificance.

The above GLMM estimates the subject-specific smoking trend. For a nonlinear model, it is well known that these estimates are not the same as the estimates of the population-averaged smoking trend. Let β^*_{1k} denote the population-averaged smoking trend. Zeger *et al.* (1988) showed that

$$\beta^*_{1k} \approx (k^2\sigma^2_{b_0}+1)^{-0.5}\beta_{1k},$$

where $k^2 \approx 0.346$. To obtain ISNI values for estimates $\hat{\beta}^*_{1k}$, simply note that:

$$\begin{aligned} ISNI(\hat{\beta}^*_{1k}) = \frac{\partial\hat{\beta}^*_{1k}}{\partial\gamma_1} &= \frac{\partial\hat{\beta}^*_{1k}}{\partial\hat{\beta}_{1k}}\frac{\partial\hat{\beta}_{1k}}{\partial\gamma_1} + \frac{\partial\hat{\beta}^*_{1k}}{\partial\hat{\sigma}_{b_0}}\frac{\partial\hat{\sigma}_{b_0}}{\partial\gamma_1} \\ &= \frac{\partial\hat{\beta}^*_{1k}}{\partial\hat{\beta}_{1k}} \times ISNI(\hat{\beta}_{1k}) + \frac{\partial\hat{\beta}^*_{1k}}{\partial\hat{\sigma}_{b_0}} \times ISNI(\hat{\sigma}_{b_0}). \end{aligned}$$

Table 3 lists the population-averaged smoking trend estimates, and the associated ISNI and c values. The SE for $\hat{\beta}^*_{1k}$ was obtained by the Delta method. We found that these population-averaged smoking trend estimates and the corresponding subject-specific estimates have approximately the same values of c statistics with numerical differences occurring at the second decimal place. Thus, the conclusion regarding the sensitivity of estimate $\hat{\beta}^*_{1k}$ to nonignorable missingness remains the same as that of $\hat{\beta}_{1k}$.

The preceding analyses demonstrate the usage of ISNI to assess the impact of nonignorability on the MAR estimates and to identify the subgroups where MAR estimates are sensitive. In the above analyses, the MAR model assumes that conditioning on race, sex, the missingness status at prior visits and the most recently observed previous smoking outcome, missingness is independent of unobserved outcome. One reason why nonignorable missingness could occur is that there are omitted variables that affect both smoking behavior and the missingness status. Therefore, to reduce nonignorability it can be important to control for such variables, if observed, in the analysis to eliminate selection on observables. Preliminary analysis shows that both the study participants' education and age are simultaneously related to the smoking outcome and the response behavior. Thus, we conduct further analysis by controlling for education and age[9]. *Education* is the attained education level, coded as "High School", "Some College" and "College Graduate". Based on their ages, participants are divided into three birth cohorts: "1963-1967", "1959-1962" and "1955-1958". The reason why the analysis uses the birth cohort instead of the continuous value of age is to obtain smoking-trend estimates adjusted by the readily available U.S. census education and birth-cohort distribution. The variable *Age* is reasonably homogeneous within each birth cohort so that *Age* is not found to be statistically related to the smoking outcome and nonresponse simultaneously in any group formed in the analysis below.

For each of the nine groups defined by crossing *Education* and birth cohort, we fit the model as specified in Equation (9) to obtain MAR estimates. Table 4 reports the population-averaged smoking trend MAR estimates for all resulting subgroups. Because the analysis additionally controls for education and age, the MAR assumption is more plausible than in the preceding analysis. The MAR trend estimates whose 95% confidence intervals exclude zero are highlighted with the "*" sign. We summarize some of the important findings in the

[9] This controls all the variables in the dataset analyzed in Preisser *et al.* (2000) and available to us. We then use the ISNI analysis to quantify the possible effects of selection on unobservables.

below: (1) Statistically significant increases in smoking are only found in two subgroups (*i.e.*, youngest high school and older college graduates) of the black-male group; (2) statistically significant decreases in smoking are only found in some subgroups of the two white groups; (3) the black-female group has statistically nonsignificant changes for all subgroups.

Table 4 also reports the ISNI values for these MAR estimates. In calculating these ISNIs, we apply the missing-data model as specified in Equation (10) to each group formed by crossing *Education* and the birth cohort. It is time-consuming to calculate the lower and upper bound of the ISNI values using GA as there are many subgroups here. As we observed and discussed in the preceding analysis, ISNIs using Classification I (II) are almost the same as the upper (lower) bound. Thus, we calculate the ISNIs using the two classification schemes that can be considered as the approximate bounds of the ISNI values. Again we see that the conclusions using Classification I and II for those ambiguous missingness types agree with each other qualitatively, although there exist considerable numerical differences for some subgroups.

The ISNI analysis identifies five subgroups whose statistical-significance results can be changed with moderate nonignorability, highlighted by the "+" sign in Table 4. Among these five subgroups, the smoking-trend estimates for "High School; 1963-1967" and "High School; 1959-1962" in black males as well as the one for "High School; 1963-1967" in black females have the smallest three c values among all of the subgroups, indicating that they are most sensitive to nonignorability. It is important to note that these three subgroups all come from the high-school category. The other two subgroups whose statistical-significance results could be changed are "High School; 1959-1962" and "College Degree; 1959-1962" of the white-female group. The range of possible estimates for "College Degree; 1959-1962" is not particularly wide. However, because its estimate is only borderline significant, its statistical-significance result can be affected by moderate nonignorability. The only quali-

tative difference between the two classification schemes occurs in "High School; 1955-1958" in the white-female group: using Classification II the range of estimates all remain statistically significant, while using Classification I the range of estimates could become borderline nonsignificant.

As noted in a previous analysis (Preisser *et al.* 2000), the study sample had higher education levels on average than young adults nationwide. In order to make some inference about the U.S. population of young adults, we calculate the overall smoking-trend estimates for race/gender groups adjusted by the U.S. census data for education and birth-cohort distribution. We follow the approach given in Preisser *et al.* (2000) to calculate these overall U.S.-census-adjusted trend estimates[10]and report them in the "U.S. Est." rows of Table 4. The overall trend estimates for black males, black females, white males, and white females are 0.013, 0.0018, -0.021, and -0.042, respectively, which are very close to the U.S.-census-adjusted trend estimates presented in Preisser *et al.* (2000). Our calculations show that the corresponding estimated changes in the smoking rate over the seven-year period were 2.1, 0.3, -2.6, and -5.3 in percentage points for the four race/sex groups. These analyses based on the MAR assumption show that the black-male group had a positive but statistically nonsignificant smoking trend. The smoking in the black-female group was quite stable over this period while the white groups had negative and statistically significant smoking trends.

The ISNI value for each of the overall trend estimates can be conveniently calculated since the ISNI for a function of model estimates can be obtained by using the chain rule

[10]The approach first collapses *Education* into two levels -"Less than College Degree" and "College Degree"- and the birth cohort into two levels -"1955-1962" and "1963-1967"- resulting in four subgroups within each race/gender group. These cutoffs were chosen because U.S. census data for the four subgroups within each race/gender group were readily available. The smoking-trend estimate for each of the four subgroups within each race/gender group is obtained by averaging the trend estimates in the comprising uncollapsed subgroups weighted by their sample sizes. The U.S.-census-adjusted smoking trend estimate in each race/gender group was then calculated as the weighted average of the smoking-trend estimates in the four subgroups, where the weights for these four subgroups -"Less than College Degree; 1963-1967", "College Degree; 1963-1967", "Less than College Degree; 1955-1962" and "College Degree; 1955-1962"- reflect the distribution of the four subgroups in the U.S. population and is calculated in Preisser *et al.* (2000) to be (.34, .05, .54, .07), (.35, .04, .54, .07), (.27, .09,.49, .15), and (.27, .09, .50, .14) in black males, black females, white males and white females, respectively.

of differentiation operator. Qualitatively, both classification schemes for those ambiguous missingness types result in the same conclusion. The ISNI analysis shows that the estimate in the black-male group has the largest sensitivity to nonignorability. With moderate nonignorability, $\gamma_1 = \pm 1$, the overall trend estimate can be a large and statistically significant value of 0.029 or it could be a very small value of -0.003 based on the more conservative classification scheme: Classification I. The corresponding range of the estimated change in the smoking rate over the study period was -0.5 to 4.8 in percentage points. This range of values is quite wide for a seven-year period. In contrast, the ranges of the estimated changes in the smoking rate in the other three groups are significantly narrower, which are (-1.3, 1.8), (-3.7, -1.7), and (-6.5, -4.4) in percentage points for black females, white males and white females, respectively. The statistical-significance result in white males could be turned to nonsignificance despite its narrow range of estimates. The statistical-significance results in black females and white females remain unaffected.

For comparison purposes, the "Study Est." rows in Table 4 report the overall trend estimates where the weights reflect the empirical distribution of education and the birth cohort in the Smoking Trend dataset instead of the U.S. census distribution. As shown in Table 4, the U.S.-census-adjusted estimates are generally larger because the study sample has higher average education levels than U.S young adults. Comparing the estimates reported in the "Study Est." rows to the population-averaged trend estimates reported in Table 3, we find that they correspond well except in the black-male group. It seems that in the Smoking Trend dataset, the random effects control for education and age reasonably well when these variables are not included in the analysis,[11] although there may still be some noticeable bias in the black-male group.

The results from the above sensitivity analysis can also be useful for designing further

[11] We obtained the empirical Bayes estimates of the random effects for the analysis reported in Table 3, and regressed them on education and age. The regression coefficients on education and age are highly significant with the expected signs.

data collection. When resources are available, one can obtain a refreshment sample to relax some of the untestable assumptions in the selection model (Hirano *et al.* 2001) or obtain other useful external data (Qian 2007) to help understand the missing-data mechanism. Our sensitivity analysis can guide the allocation of resources among the different groups. For example, we might like to allocate more resources (*e.g.*, more refreshment sampling units) into the black-male group because the conclusion in this group is most susceptible to nonignorable missingness. It is also useful to allocate more resources to the high-school category since the subgroups whose c values are smallest concentrate on the high-school category.

5. A Simulation Study

In this section, we conduct a simulation study to evaluate the accuracy of local approximation by the extended ISNI. We first simulate complete panel binary outcomes from the following ideal-data model: $Y_{ij}|\beta_{0i}$ is distributed independently as Bernoulli(μ_{ij}^Y), where $\text{logit}(\mu_{ij}^Y) = \beta_{0i} + \beta_1 j$, $\beta_{0i} \sim N(\beta_0, \sigma^2)$, $i = 1, ..., N$, and $j = 0, 2, 5, 7$. In the simulations, we set $\beta_0 = -1.5$ and $\beta_1 = 0.05$, and $\sigma = 5$. The parameter setting corresponds roughly to the MAR estimates in the black-male group. We then simulate the occurrence of missingness from the nonignorable MDM as specified in Equation (3), where $\phi_{ij}^{uv} = \gamma_{00}^{uv} + \gamma_{01}^{uv} I(j = 5) + \gamma_{02}^{uv} I(j = 7) + \gamma_{03}^{uv} y_{ij}^{MOP} + \gamma_1 y_{ij}$, $j = 2, 5, 7$, and y_{ij}^{MOP} is the most recently observed past outcome before time j. Note that y_{ij}^{MOP} is not the same as $y_{i,j-1}$ in the presence of intermittent missingness. In the simulations, we set γ_1 as -0.5, 0 or 0.5, and we set the value of γ_0^{uv} as follows: $(\gamma_{00}^{10}, \gamma_{01}^{10}, \gamma_{02}^{10}, \gamma_{03}^{10}) = (-3, -0.4, -0.5, 0.45)$, $(\gamma_{00}^{20}, \gamma_{01}^{20}, \gamma_{02}^{20}, \gamma_{03}^{20}) = (\gamma_{01}^{20}, 0.4, 0.8, 0.45)$ and $(\gamma_{00}^{11}, \gamma_{01}^{11}, \gamma_{02}^{11}, \gamma_{03}^{11}) = (-3, 3, -7, 0.04)$, where γ_{01}^{20} is -2.5, -3 or -3.5. The value of γ_0^{uv} when $\gamma_{01}^{20} = -3$ roughly corresponds to the MAR estimates when fitting the above MDM to the Smoking Trend dataset. There are nine parameter settings in total by varying the values of γ_1 and γ_{01}^{20}, and we simulate one dataset under each parameter setting. By varying

the values of γ_1 and γ_{01}^{20}, we generate datasets that have various degrees of missingness proportions and magnitudes of nonignorability.

Figure 2 reports the simulation results for a sample size $N = 300$. The generated datasets have a wide range of missingness proportions. The proportions of units with a dropout (given by Prop_D) and the proportions of units with intermittent missingness (given by Prop_I) cover the likely scenarios occurring in the Smoking Trend dataset. Figure 2 compares the ISNI analysis with the corresponding global sensitivity analysis in each of the nine simulated datasets. The comparison is focused on the slope parameter β_1. To conduct the global sensitivity analysis, we construct the correct log-likelihood for the simulated dataset that is in the form of Equation (5) and obtain $\hat{\beta}_1(\gamma_1)$ for a series of γ_1 values between -3 and 3. The resulting sequence of estimates $\hat{\beta}_1(\gamma_1)$ as a function of γ_1 is plotted as the solid curve in Figure 2. The entire process of obtaining this global sensitivity curve was computationally costly and took a very long time. We used the adaptive quadrature method to evaluate integrals involved in the likelihood evaluation, and the Quasi-Newton method to optimize the likelihood in the R language. The computation bottleneck was the evaluation of the likelihood function at the subject level. Thus, we vectorized its evaluation to maximize the speed. In the simulation study, we found that the average time to estimate the model for a posited value of γ_1 was about three hours. Much time was spent on evaluating the likelihood for subjects who had multiple occurrences of missingness. The computational cost is much higher as the numbers of intermittent missing outcomes per subject increase.

In contrast, the calculation of ISNIs in these datasets took an average of less than ten seconds, because the ISNI calculation does not require optimization of the nonignorable likelihood or numerical evaluation of integration with respect to the missing outcomes. ISNIs calculate the first derivative of the $\hat{\beta}_1(\gamma_1)$ curve at $\gamma_1 = 0$. One can then approximate $\hat{\beta}_1(\gamma_1)$ by $\hat{\beta}_1(\gamma_1) \approx \hat{\beta}_1(0) + ISNI * \gamma_1$, which is the tangent line of the $\hat{\beta}_1(\gamma_1)$ curve at $\gamma_1 = 0$. As

shown in Figure 2, the tangent line (dashed line) closely approximates the $\hat{\beta}_1(\gamma_1)$ curve when γ_1 falls between -2 and 2, across all the datasets. In particular, the ISNI approximation line and the global sensitivity curve are almost identical in the range of -1 to 1, the range of the nonignorability parameter considered relevant in our Smoking Trend dataset. Figure 3 shows the simulation result for the same setting except that the sample size is reduced to 100. As we can see, the message is very much the same. We also conducted the simulation using the parameter settings for white females (the results are not shown here and are available from the authors upon request). The conclusion based on that simulation is the same as the one presented here in that the ISNI approximation is adequate in the range of -2 to 2 and almost identical with the global sensitivity curve when it falls between -1 and 1. These results demonstrate that the local approximation by ISNI is accurate for a wide range of empirically relevant γ_1 values.

Overall, we feel that the speed factor and sufficient accuracy make the extended ISNI appealing for quickly assessing the potential impact of nonignorable non-monotone missingness on an MAR analysis. The gain in efficiency would be more appreciable if the panel were longer with more intermittent missingness or if one were to compare sensitivity with several different model specifications.

6. Discussion

The analysis of panel data with non-monotone missingness is challenging. Because the missing data are themselves unobserved, there is generally insufficient information to reliably estimate a joint selection model without the aid of external data or untestable assumptions. In practice, a significant amount of analyses (at least initially) have assumed that missingness is ignorable, and have performed a likelihood-based inference without explicitly modeling how data is missing. In many cases, there may still be some suspicion that missingness is

moderately related to the unobserved outcomes, even after conditioning on a set of powerful observed predictors for missingness. In these cases, it can be important to assess whether the MAR analysis is credible by measuring the sensitivity of such inferences to alternative missing-data assumptions. In this paper, we have extended the ISNI method to accomplish this goal in the panel data with an arbitrary pattern of missingness. We believe that the method is applicable to a wide range of economic studies with nonresponse. In such cases, ISNI is useful to quickly evaluate whether the impact of nonignorable selection is important before deciding to invest a great deal of valuable resources and effort to collect additional data and to perform arduous nonignorable modeling. Once one finds important sensitivity and decides to collect more data to gain a better understanding of the missing-data mechanism, the ISNI analysis result can also serve as a useful direction on how to optimally allocate new sampling units as illustrated in the Smoking Trend dataset.

A common finding that emerges from our analysis of the Smoking Trend dataset is that there is a great deal of uncertainty regarding the smoking-trend estimates in the black-male group. Moderately nonignorable missingness can produce a wide range of trend estimates that lead to qualitatively different conclusions. In particular, if a black-male smoker is moderately more likely to miss a visit than a nonsmoker, the trend estimate could be changed to become highly statistically significant. Thus, one should be cautious with conclusions based solely on the MAR estimate in this group. Further studies or complementary data would be helpful to confirm the findings. If further studies find that the smoking-trend estimate is positive and significant in this group, perhaps more smoking-cessation programs should be targeted at this group of U.S. young adults. Such programs combined with adolescence education in high school can be very beneficial in light of the growing smoking trend for the samples of youngest black citizens with high-school education levels (rows 1 and 23 in Table 4 for black male and female high-schoolers, respectively).

The conclusions regarding the black-female and white-female groups are robust to moderate nonignorability, although the U.S.-census-adjusted trend estimate in the white-male group is also subject to some uncertainty due to its borderline statistical significance. The negative time trend for smoking among white U.S. young adults sets a stark contrast to the trend among black U.S. young adults. Such racial inequality in smoking trends potentially reveals other dimensions of inequality and points to severe societal consequences. Notably, government reports indicate that blacks are under served by cancer-prevention research and health education programs, and that they have limited access to health care (MMWR 1993). Another potentially useful finding is that much sensitivity occurs in groups of young adults with the lowest level of education (*i.e.*, the high-school category). It may be worthwhile to allocate more new sampling units to these groups if one decides to obtain additional data and conduct closer analyses. The results compiled in Table 4, therefore, shed lights on a series of policies and guide future research priorities.

Our modeling framework allows dropout and intermittent missingness to enter the likelihood for the nonignorable model in a different way. In applications where information on distinguishing dropout and intermittent missingness exists, it is preferable to use the information when constructing the likelihood. The resulting sensitivity analysis more accurately reflects the true missingness behavior and the fact that any missing observation after the dropout visit does not enter the likelihood and thus contributes no sensitivity. When such information is absent, one can view the unknown missingness types as missing data, and then obtain the bounds of ISNI to reflect the uncertainty in defining the missingness types, as demonstrated in our empirical application. In our view, the bound analysis provides a general approach to quantifying the uncertainty whenever researchers are unsure of the effect of those unknown missingness types on their analysis. The approach is rigorous in that it avoids ad-hoc specification of unknown missingness types. Such ad-hoc specification

commonly has little support from data and thus the corresponding results can be questionable. As one referee pointed out, a potential drawback of the bound approach is that the bounds have to be calculated numerically. We use numerical method because no closed-form solutions of bounds are available. A numerical search procedure, such as GA used in our application, requires evaluating the objective function many times. Fortunately our objective function, ISNI, is very fast to calculate as explained in Section 3. In particular, ISNI avoids fitting any nonignorable model. Therefore the simplicity of ISNI evaluation renders the numerical search for the bounds practical. Finally we note that the bound analysis does not preclude the usefulness of collecting detailed information on the reasons for missingness and using them when available. Besides other benefits, using such information would narrow or eliminate the bounds due to the uncertain missingness types in a sensitivity analysis.

We have derived in this paper explicit formulas to calculate the extended ISNI for three popular classes of panel data models: the marginal multivariate Gaussian model, the GLMM, and the panel Tobit model. These formulas could be useful for analyzing other types of outcomes in smoking studies or similar panel studies in other fields. The method is general and can be extended to any other likelihood-based estimation for panel outcomes. In addition to its wide range of applicability, the proposed procedure is relatively easy and efficient to carry out due to its similarity to score statistics. This feature is particularly desirable in the presence of non-monotone missingness. which would otherwise greatly complicate the analysis. The advantage is more valuable when the number of intermittent missingness is larger or when one would like to compare sensitivity with several different specifications of nonignorable models. After ISNIs are obtained for the model parameters, the ISNI for any well-behaved function of these model-parameter estimates can be conveniently obtained.

Our method offers a parametric approach to relax the MAR assumption. Like all model-based approaches, the analysis results depend on the underlying distributional assumptions

in the selection model, as captured in the likelihood function given in Equation (5). In the presence of nonignorable missingness, these model assumptions are generally untestable. Our view is that in the missing-data analysis, untestable assumptions are often unavoidable. Our approach is to view the MAR analysis as the standard analysis and investigate whether the conclusions obtained from the standard analysis still hold under a broader set of assumptions. If so, the standard analysis is deemed as credible; otherwise, it is considered suspicious. There is also some empirical evidence demonstrating that the conclusions of local sensitivity to nonignorability are less sensitive to model assumptions, at least locally, than conclusions drawn from the direct estimation of the nonignorable model (Copas and Li 1997; Zhang and Heitjan 2008). This is intuitive because the sensitivity analysis fixes the nonignorable parameter γ_1, which is the component of the model for which data carry little information, thereby removing the major source of sensitivity to model assumptions.

Throughout the paper we have employed the selection modeling framework to account for nonignorable missingness. In selection models, nonignorability is induced by allowing the probability of missingness to depend directly on the current outcome, which is a time-varying variable. An alternative modeling approach is to use the shared-parameter model (see, *e.g.*, Follman and Wu 1995), where nonignorable missingness is modeled by allowing the probability of missingness to depend on the random effects in the ideal-data model, thereby depending indirectly on the outcome. An important difference is that random effects are commonly time-constant, and thus do not account for idiosyncratic variations in the outcome as does the selection model. Diggle *et al.* (2002) provide more-detailed comparisons between the selection model and the shared-parameter model. Another reason we use the selection model in our development is that the nonignorability parameter is easier to interpret in a selection model while the nonignorability parameter in shared-parameter models depends on random effects, which are entirely latent.

Other available methods for measuring sensitivity to nonignorable non-monotone nonresponse in panel data include Jansen *et al.* (2003) and Minini and Chavance (2004). Both papers restrict their attention to panel binary data and employ a log-linear model. Although the log-linear model has certain advantages and may be preferred under some circumstances, the model is difficult to apply with unequal numbers of scheduled visits across subjects[12] and time-varying covariates. Moreover, Minini and Chavance mention that their method is not applicable due to heavy computation when the number of scheduled visits is greater than 10. Our method does not have this computational restriction. These considerations suggest that our method is applicable to a wider range of applications and more appropriate for certain important situations.

Our method is well-suited for the cases where missingness is plausibly ignorable, *i.e.*, nonignorability is moderate. The first-order Taylor-series approximation to the nonignorable likelihood is accurate in such cases. Although in many panel datasets the missingness can be considered plausibly ignorable, there may exist cases where compelling prior information suggests that the nonignorability is very extreme, and, thus, the evaluation of the change of estimates in this remote area from the MAR model becomes relevant. The approximation might not be accurate enough for such extreme nonignorability. In such cases, one may need to conduct a global sensitivity analysis that is exact but arduous. The ISNI method can facilitate this global sensitivity analysis in two ways. First, the approximation of $\hat{\theta}(\gamma_1)$ using ISNI can provide a good starting value for a faster and a more reliable fitting of the nonignorable model for a given value of γ_1. Second, when γ_1 is a vector and, thus, the size of the plausible values of γ_1 is large, the ISNI method can help find a smaller relevant set of values of γ_1, which can reduce the workload of the global search for maximal sensitivity.

A limitation of our method is that we require that the covariates in X_i are fully observed.

[12] One example of such an unbalanced panel is the rotating panel where the survey design rotates participants or firms out of the sample based on pre-specified rules. Wooldridge (2002) gives an example of a rotating panel.

For subject-specific time-varying covariates, X_{ij} will often be missing if Y_{ij} is missing. Moreover, the missingness of X_{ij} is also nonignorable since its missingness depends on unobserved data Y_{ij}. In order to calculate ISNI in this general case, we need to posit a model for the missing covariates. This substantially increases the modeling and computation task, and is beyond the scope of this paper. We leave this for future research.

Acknowledgements

We are grateful to Dr. Daniel Heitjan, Dr. Donald Hedeker and Dr. Hua Yun Chen for their constructive comments that improved the draft. We also thank the Co-Editor, Dr. Manuel Arellano, and two anonymous referees for many constructive comments that led to substantial improvements in the manuscript.

References

Albert, P. (2000) "A Transitional Model for Longitudinal Binary Data Subject to Nonignorable Missing Data," *Biometrics,* **56**, 602–608.

Amemiya, T. (1985) "Advanced Econometrics," Cambridge, MA: Harvard University Press.

Ashley, R. (2009) "Assessing the Credibility of Instrumental Variables Inferences with Imperfect Instruments Via Sensitivity Analysis," *Journal of Applied Econometrics,* **24**, 325–337.

Copas, J. B., and Eguchi, S. (2001), "Local Sensitivity Approximations for Selectivity Bias," *Journal of the Royal Statistical Society B,* **63**, 871–895.

Copas, J. B., and Li, H. G. (1997), "Inference for Non-random Samples (with discussion)," *Journal of the Royal Statistical Society B,* **59**, 55–95.

Diggle, J. P., Heagerty, P., Liang, K.Y. and Zeger, S.L. (2002), "Analysis of Longitudinal Data", 2nd Edition, Oxford University Press: New York.

Diggle, J. P., and Kenward, M. G. (1994), "Informative Drop-out in Longitudinal Data Analysis" (with discussion), *Journal of the Royal Statistical Society C,* **43**, 49–73.

Follmann, D. and Wu, M. (1995), "An Approximate Generalized Linear Model with Random effects for Informative Missing Data", *Biometrics,* **51**, 151–168.

Hausman, J.A. and Wise, D.A., (1979), "Attrition bias in experimental and panel data: the Gary income maintenance experiment," *Econometrica,* **47**, 455–473.

Heckman, J. and Hotz, V.J., (1989), "Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower trainings," *Journal of the American Statistical Association,* **84**, 862–874.

Heitjan, D. F., and Rubin, D. B. (1991), "Ignorability and Coarse Data," *Annals of Statistics,* **19**, 2244–2253.

Hirano, K., Imbens, G.W., Ridder, G. and Rubin, D.R., (2001), "Combining panels with attrition and refreshment samples," *Econometrica,* **69**, 1645–1659.

Horowitz, J.L., Manski, C.F., (2006), "Identification and Estimation of Statistical Functionals using Incomplete Data," *Journal of Econometrics,* **132**, 445–459.

Jansen, I., Molenberghs, G., Aerts, M., Thijs, H., van Steen, K. (2003), "A Local Influence Approach Applied to Binary Data From a Psychiatric Study," *Biometrics,* **59**, 410–419

Judd, K.L. (1998), "Numerical Methods in Economics". The MIT Press, Cambridge, MA.

Kenward, M. G. (1998), "Selection Models for Repeated Measurements with Non-random Dropout: An Illustration of Sensitivity," *Statistics in Medicine,* **17**, 2723–2732.

Little, R. J. A. (1995), "Modeling the Drop-out Mechanism in Longitudinal Studies," *Journal of the American Statistical Association,* **90**, 1112–1121.

Ma, G., Troxel, A. B., and Heitjan, D. F. (2005), "An Index of Local Sensitivity to Nonignorable Dropout in Longitudinal Modeling," *Statistics in Medicine,* **24**, 2129–2150.

Mebane, W. R. Jr. and Sekhon, J.S. (2007). "Genetic Optimization Using Derivatives: The rgenoud package for R." http://sekhon.berkeley.edu/papers/rgenoudJSS.pdf

Minini, P. and Chavance, M. (2004), "Sensitivity Analysis of Longitudinal Binary Data with Non-monotone Missing Values," *Biostatistics,* 54: 531–544.

MMWR (Morbidity and Mortality Weekly Report). (1993), "Smoking Cessation During Previous Year Among Adults – United States, 1990 and 1991.," Issue 42: 504–507.

Nicoletti, C. (2006), "Nonresponse in Dynamic Panel Data Models," *Journal of Econometrics,* **132**, 461–489.

Preisser, J.S., Galecki, A.T., Lohman, K.K., Wagenknecht, L.E. (2000), "Analysis of Smoking Trends with Incomplete Longitudinal Binary Responses," *Journal of the American Statistical Association,* **95**, 1021–1031.

Qian, Y (2007), "Do National Patent Laws Stimulate Domestic Innovation in a Global Patenting Environment? A Cross-Country Analysis of Pharmaceutical Patent Protection, 1978-2002," *The Review of Economics and Statistics,* **89(3)**, 436–453

Qian, Y and Xie, H. (2009), "Analyzing Nonignorably Missing Data with a Generalized Additive Model for Missing Data Mechanism," Working paper.

Rotnitzky, A., Robins, J. M., and Scharfstein, D. O. (1998), "Semiparametric Regression for Repeated Outcomes with Nonignorable Nonresponse," *Journal of the American Statistical Association,* **93**, 1321–1339.

Rubin, D. B. (1976), "Inference and Missing Data (with discussion)," *Biometrika,* **63**, 581–592.

Scharfstein, D., Rotnitzky, A., and Robins, J. M. (1999), "Adjusting for Nonignorable Dropout Using Semiparametric Models (with discussion)," *Journal of the American Statistical Association,* **94**, 1096–1146.

Schluchter, M. D. (1992), "Methods for the Analysis of Informatively Censored Longitudinal Data," *Statistics in Medicine,* **11**, 1861–1870.

Tobin, J. (1956) "Estimation of Relationships for Limited Dependent Variables," *Econometrica,* **26**, 24–36.

Troxel, A. B., Harrington, D. P., and Lipsitz, S. R. (1998), "Analysis of Longitudinal Measurements with Non-ignorable Non-monotone Missing Values," *Applied Statistics,* **47**, 425–438.

Troxel, A. B., Ma, G., and Heitjan, D. F. (2004), "An Index of Local Sensitivity to Nonignorability," *Statistica Sinica,* **14**, 1221–1237.

Vach, W., and Blettner, M. (1995), "Logistic Regression with Incompletely Observed Categorical Covariates — Investigating the Sensitivity Against Violation of the Missing at Random Assumption," *Statistics in Medicine,* **14**, 1315–1329.

Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E., and Kenward, M. G. (2001), "Sensitivity Analysis for Nonrandom Dropout: a Local Influence Approach," *Biometrics,* **57**, 7–14.

Wooldridge, J.M. (2002) "Econometric Analysis of Cross Section and Panel Data," Cambridge, MA: the MIT Press.

Xie, H., and Heitjan, D. F. (2004), "Sensitivity Analysis of Causal Inference in a Clinical Trial Subject to Crossover," *Clinical Trials,* **1**, 21–30.

Xie, H.(2008), "A Local Sensitivity Analysis Approach to Longitudinal Non-Gaussian Data with Non-ignorable Dropout," *Statistics in Medicine,* **27**, 3155–3177.

Zeger, S.L., Liang, K.Y., and Albert, P.S. (1988). "Models for Longitudinal Data: a Generalized Estimating Equation Approach," *Biometrics,* **44**, 1049–60.

Zhang, J., Heitjan, D.F. (2006), "A Simple Local Sensitivity Analysis Tool for Nonignorable Coarsening: Application to Dependent Censoring," *Biometrics,* **62**,1260–1268.

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

Appendix I: Derivation of $\nabla^2 L\theta, \gamma_1$

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

$$\begin{aligned}
\nabla^2 L_{\theta,\gamma_1^{10}} &= \sum_{i:K_i<n_i} \frac{\partial}{\partial\theta}\frac{\partial}{\partial\gamma_1^{10}} \ln\left(\int \prod_{j:g_{ij}\neq 0} f_\gamma(g_{ij}|s_{ij}, y_{ij}, g_{i,j-1}) f_\theta(y_i^{mis}|y_i^{\text{obs}}, x_i) dy_i^{mis}\right)\Bigg|_{\gamma_1=0} \\
&= \sum_{i:K_i<ni} \frac{\partial}{\partial\theta}\left(\frac{\int \frac{\partial \prod_{j:g_{ij}\neq 0} f_\gamma(g_{ij}|s_{ij},y_{ij},g_{i,j-1})}{\partial\gamma_1^{10}} f_\theta(y_i^{mis}|y_i^{\text{obs}}, x_i) dy_i^{mis}}{\int \prod_{j:g_{ij}\neq 0} f_\gamma(g_{ij}|s_{ij}, y_{ij}, g_{i,j-1}) f_\theta(y_i^{mis}|y_i^{\text{obs}}, x_i) dy_i^{mis}}\Bigg|_{\gamma_1=0}\right) \\
&= \sum_{i:K_i<ni} \frac{\partial}{\partial\theta}\left(\int \sum_{j:g_{ij}\neq 0} \frac{I(g_{i,j-1}=0)\frac{\partial P_{ij}^{g_{ij},g_{i,j-1}}}{\partial\phi_{ij}^{1,g_{i,j-1}}}\phi_{ij}^{1,g_{i,j-1}}}{P_{ij}^{g_{ij},g_{i,j-1}}} y_{ij} f_\theta(y_i^{mis}|y_i^{obs}, x_i) dy_i^{mis}\Bigg|_{\gamma_1=0}\right) \\
&= \sum_{i:K_i<n_i} \frac{\partial}{\partial\theta} E((y_i^{mis})^T A_i^{10}|y_i^{obs}, x_i)\Big|_{\gamma_1=0} \\
&= \sum_{i:K_i<n_i} \frac{\partial E((y_i^{mis})^T|y_i^{obs}, x_i)}{\partial\theta}\Bigg|_{\gamma_1=0} \cdot A_i^{10} \qquad (11)
\end{aligned}$$

Suppose the lth component of y_i^{mis} corresponds to the jth element of y_i; then the lth element of A_i^{10} is shown to be

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

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

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

Appendix II: Derivation of $\left.\frac{\partial E((y_i^{mis})^T|y_i^{obs},x_i)}{\partial\theta}\right|_{\gamma_1=0}$

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

(1) Marginal Multivariate Gaussian Model

Since

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

then, by vector differentiation, we have

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

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

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

(2) Generalized Linear Mixed Model

We apply the Cholesky decomposition to $\Sigma(\Omega)$ such that $\Lambda_\Omega\Lambda_\Omega^T = \Sigma(\Omega)$ and $b_i = \Lambda_\Omega B_i$, where the random vector B_i is drawn from the multivariate standard normal $MVN(0, I_{q\times q})$

with the density function $\phi(B_i)$. As a result, $\mu_{ij}^Y = \mu_{ij}^Y(\theta, B_i) = h_Y(X_{ij}^T\beta + Z_{ij}^T\Lambda_\Omega B_i)$. Then

$$
\begin{aligned}
&\frac{\partial E((y_i^{mis})^T|y_i^{obs}, x_i)}{\partial\theta}\bigg|_{\gamma_1=0} \\
&= \sum_{i:K_i<n_i}\left[\frac{\partial}{\partial\theta}\int (y_i^{mis})^T f_\theta(y_i^{mis}|y_i^{obs}, x_i)dy_i^{mis}\right]_{\hat\theta(0),\hat\gamma_0(0),\gamma_1=0} \\
&= \sum_{i:K_i<n_i}\left[\frac{\partial}{\partial\theta}\int \mu_{i,mis}^T(\theta, B_i) f_\theta(B_i|y_i^{obs})dB_i\right]_{\hat\theta(0),\hat\gamma_0(0),\gamma_1=0} \\
&= \sum_{i:K_i<n_i}\left[\frac{\partial}{\partial\theta}\int \mu_{i,mis}^T(\theta, B_i)\frac{f_\theta(y_i^{obs}|B_i, x_i)\phi(B_i)}{f_\theta(y_i^{obs}|x_i)}dB_i\right]_{\hat\theta(0),\hat\gamma_0(0),\gamma_1=0} \\
&= \sum_{i:K_i<n_i}\Bigg[f_\theta(y_i^{obs}|x_i)^{-1}\int \frac{\partial\mu_{i,mis}^T(\theta, B_i)}{\partial\theta} f_\theta(y_i^{obs}|B_i, x_i)\phi(B_i)dB_i \\
&\quad + f_\theta(y_i^{obs}|x_i)^{-1}\int \frac{\partial \ln f_\theta(y_i^{obs}|B_i, x_i)}{\partial\theta}\mu_{i,mis}^T(\theta, B_i) f_\theta(y_i^{obs}|B_i, x_i)\phi(B_i)dB_i \\
&\quad - f_\theta(y_i^{obs}|x_i)^{-2}\frac{\partial f_\theta(y_i^{obs}|x_i)}{\partial\theta}\int \mu_{i,k_m is}^T(\theta, B_i) f_\theta(y_i^{obs}|B_i, x_i)\phi(B_i)dB_i\Bigg]_{\hat\theta(0),\hat\gamma_0(0),\gamma_1=0} \\
&= \sum_{i:K_i<n_i}\Bigg\{E\left[\frac{\partial\mu_{i,mis}^T(\theta, B_i)}{\partial\theta}|y_i^{obs}\right] + E\left[S_i(\theta; B_i)\mu_{i,mis}^T(\theta, B_i)|y_i^{obs}\right] \\
&\qquad\qquad - E\left[S_i(\theta; B_i)|y_i^{obs}\right]\ E\left[\mu_{i,mis}^T(\theta, B_i)|y_i^{obs}\right]\Big\}_{\hat\theta(0),\hat\gamma_0(0),0}, \qquad (12)
\end{aligned}
$$

where the interchange of the differentiation and integration signs is justified by the dominate convergence theorem for the exponential family distributions and common link functions. Expectation E is taken with respect to the posterior distribution of B_i given y_i^{obs}. $S_i(\theta; B_i) = \frac{\partial \ln f_\theta(y_i^{obs}|B_i, x_i)}{\partial\theta}$. $S_i(\theta; B_i)$ and $\frac{\partial\mu_{ij}^Y}{\partial\theta}$ can be given in closed forms and readily evaluated at the MAR estimates. In GLMM, these two terms are non-linear functions of the random effects B_i. Therefore, there are generally no closed forms for calculating the conditional expectations of them or their functions. We apply the method of Gaussian quadrature to approximate

these conditional expectations as follows. For function $g(B_i)$,

$$\begin{aligned} E\left[g(B_i)|y_i^{obs}\right] &= \int g(B_i)f(B_i|y_i^{obs})dB_i = \frac{1}{f_\theta(y_i^{obs}|x_i)}\int g(B_i)f_\theta(y_i^{obs}|B_i,x_i)\phi(B_i)dB_i \\ &\approx \frac{1}{f_\theta(y_i^{obs})}\sum_{k=1}^{K} g(Q_k)f_\theta(y_i^{obs}|Q_k,x_i)w_k, \end{aligned}$$

where Q_k is the kth $q-$dimensional quadrature point, w_k is the associated quadrature weight and K is the number of quadrature points.

(3) Panel Tobit Model.

Equation (12) also applies to the Panel Tobit model, where $S_i(\theta;B_i) = \sum_{j:g_{ij}=0} \frac{\partial \ln f_\theta(y_{ij}|B_i,x_i)}{\partial\theta}$ and $f_\theta(y_{ij}|B_i,x_i)$ can be obtained from Equation (2) by appropriate Cholesky decomposition. In the Tobit model, μ_{ij}^Y is given in the following form:

$$\mu_{ij}^Y(\theta,B_i) = \Phi((x_{ij}^T\beta + z_{ij}^T\Lambda(\Omega)B_i)/\sigma)(x_{ij}^T\beta + z_{ij}^T\Lambda(\Omega)B_i) + \sigma\phi(x_{ij}^T\beta + z_{ij}^T\Lambda(\Omega)B_i).$$

[hP]

Table 1

Summary Statistics in the Smoking Trends Dataset. % denotes smoking rate. N denotes the number of observed subjects.

	Black Males		Black Females		White Males		White Females	
Time in Years	%	N	%	N	%	N	%	N
0	36.9%	1145	31.3%	1473	26.5%	1160	27.3%	1300
2	38.4%	972	30.6%	1280	25.7%	1091	24.0%	1221
5	37.2%	896	31.9%	1205	24.3%	1043	21.0%	1167
7	36.8%	820	29.5%	1128	21.8%	998	20.5%	1096

[hP]

Table 2

Missing Data Patterns in the Smoking Trend Dataset.
X indicates presence during the visit and . indicates absence during the visit.

	Black Males		Black Females		White Males		White Females	
Pattern	Freq	Percent	Freq	Percent	Freq	Percent	Freq	Percent
XXXX	715	62.5	1006	68.3	939	81.0	1033	79.5
XXX.	110	9.6	117	7.9	69	6.0	93	7.2
XX..	103	9.0	108	7.3	53	4.6	66	5.1
X...	89	7.8	99	6.7	29	2.5	37	2.9
XX.X	44	3.8	49	3.3	30	2.6	29	2.2
X.XX	48	4.2	61	4.1	24	2.1	33	2.5
X.X.	23	2.0	21	1.4	11	1.0	8	0.6
X..X	13	1.1	12	0.8	5	0.4	1	0.1

Table 3

ISNIs of the Smoking Trend Estimates in a GLMM Model. The MAR estimates are the annual changes of the log odds of smoking. For each MAR estimate in Column A of the table, ISNI in the first (fourth) row of Column C is the lower (upper) bound, and ISNI in the second (third) row of Column C is obtained using Classification II (I).

Parameter	(A) MAR Estimate	(B) S.E.	(C) ISNI	(D) MAR Estimate ± ISNI	(E) c
Subject-Specific Slopes					
Black Males	0.055	0.026	0.0400	(0.015, 0.095)	0.6
			0.0401	(0.015, 0.095)	0.6
			0.0570	(-0.002, 0.112)	0.4
			0.0572	(-0.002, 0.112)	0.4
Black Females	-0.0036	0.018	0.0181	(-0.022, 0.015)	0.9
			0.0182	(-0.022, 0.015)	0.9
			0.0277	(-0.031, 0.024)	0.6
			0.0278	(-0.031, 0.024)	0.6
White Males	-0.099	0.028	0.0216	(-0.121, -0.077)	1.3
			0.0219	(-0.122, -0.077)	1.3
			0.0288	(-0.128, -0.070)	1.0
			0.0292	(-0.128, -0.070)	1.0
White Females	-0.174	0.027	0.0237	(-0.198, -0.150)	1.1
			0.0239	(-0.198, -0.150)	1.1
			0.0325	(-0.207, -0.142)	0.8
			0.0329	(-0.207, -0.141)	0.8
Population-Average Slopes					
Black Males	0.018	0.008	0.014	(0.005, 0.032)	0.6
			0.014	(0.005, 0.032)	0.6
			0.019	(-0.001, 0.037)	0.4
			0.019	(-0.001, 0.037)	0.4
Black Females	-0.0012	0.006	0.006	(-0.007, 0.005)	0.9
			0.006	(-0.007, 0.005)	0.9
			0.008	(-0.009, 0.007)	0.6
			0.008	(-0.009, 0.007)	0.6
White Males	-0.033	0.01	0.008	(-0.041, -0.025)	1.3
			0.008	(-0.041, -0.025)	1.3
			0.010	(-0.043, -0.023)	1.0
			0.010	(-0.043, -0.023)	1.0
White Females	-0.058	0.009	0.008	(-0.066, -0.050)	1.1
			0.008	(-0.066, -0.050)	1.1
			0.011	(-0.069,-0.047)	0.8
			0.011	(-0.069,-0.047)	0.8

Table 4

*Population-Average Smoking Trend Estimates adjusted by the Level of Education and Birth Cohort. *: Statistically significant at the 0.05 level; +: Statistical significance result is changed when γ_1 is ± 1. For each MAR estimate in the table, ISNI in the first (second) row is obtained using Classification II (I).*

Group	Education	Birth Cohort	MAR Est.	S.E.	ISNI	MAR Est. ± ISNI	c
Black Males	High School	1963-1967	0.052*	0.021	0.014	(0.038, 0.067)$^+$	1.4
					0.020	(0.032, 0.073)$^+$	1.0
		1959-1962	-0.037	0.028	0.018	(-0.055,-0.019)$^+$	1.5
					0.024	(-0.062,-0.013)$^+$	1.1
		1955-1958	0.004	0.027	0.015	(-0.011,0.019)	1.8
					0.019	(-0.015,0.023)	1.4
	Some College	1963-1967	0.009	0.027	0.003	(0.006, 0.013)	8.6
					0.005	(0.005, 0.014)	5.7
		1959-1962	0.005	0.022	0.010	(-0.005,0.015)	2.1
					0.016	(-0.010, 0.021)	1.4
		1955-1958	0.020	0.022	0.011	(0.009,0.031)	2.1
					0.014	(0.006, 0.034)	1.5
	College Degree	1963-1967	0.016	0.061	0.017	(-0.000, 0.033)	3.7
					0.019	(-0.003, 0.036)	3.3
		1959-1962	0.114*	0.048	0.013	(0.101, 0.127)	3.7
					0.018	(0.096, 0.132)	2.6
		1955-1958	-0.048	0.036	0.008	(-0.055,-0.040)	4.7
					0.011	(-0.058, -0.037)	3.4
	Study Est.		0.013	0.009	0.012	(0.001,0.025)$^+$	0.7
					0.016	(-0.003,0.029)$^+$	0.6
	U.S. Est.		0.013	0.009	0.012	(0.001,0.026)$^+$	0.7
					0.016	(-0.003,0.029)$^+$	0.6
Black Females	High School	1963-1967	0.043	0.026	0.019	(0.024, 0.063)$^+$	1.4
					0.027	(0.017, 0.070)$^+$	1.0
		1959-1962	0.001	0.028	0.012	(-0.011, 0.013)	2.3
					0.016	(-0.015, 0.017)	1.8
		1955-1958	0.012	0.022	0.011	(0.001,0.023)	2.0
					0.015	(-0.003, 0.027)	1.5
	Some College	1963-1967	0.002	0.006	0.003	(-0.001, 0.004)	2.4
					0.004	(-0.002, 0.006)	1.7
		1959-1962	-0.019	0.018	0.006	(-0.025,-0.013)	2.8
					0.011	(-0.030, -0.008)	1.7
		1955-1958	-0.014	0.016	0.006	(-0.020,-0.008)	2.6
					0.010	(-0.024, -0.004)	1.7
	College Degree	1963-1967	0.029	0.054	0.008	(0.020, 0.037)	6.5
					0.010	(0.019, 0.038)	5.6
		1959-1962	-0.011	0.044	0.005	(-0.016,-0.005)	8.4
					0.011	(-0.021, 0.000)	4.1
		1955-1958	-0.017	0.028	0.004	(-0.022, -0.013)	6.3
					0.006	(-0.024, -0.011)	4.3
	Study Est.		0.0004	0.008	0.008	(-0.008,0.008)	1.0
					0.011	(-0.011, 0.011)	0.7
	U.S. Est.		0.0018	0.007	0.008	(-0.006, 0.009)	0.9
					0.011	(-0.009, 0.012)	0.6

continued on next page

Table 4: *continued*

Group	Education	Birth Cohort	MAR Est.	S.E.	ISNI	MAR Est. ± ISNI	c
White Males	High School	1963-1967	0.001	0.016	0.003	(-0.002, 0.004)	5.6
					0.004	(-0.003, 0.005)	4.1
		1959-1962	0.010	0.036	0.010	(0.000, 0.020)	3.7
					0.012	(-0.001,0.022)	3.0
		1955-1958	-0.023	0.030	0.011	(-0.034,-0.011)	2.7
					0.013	(-0.036,-0.009)	2.2
	Some College	1963-1967	0.006	0.018	0.002	(0.004, 0.008)	11.0
					0.002	(0.004, 0.008)	8.0
		1959-1962	-0.022	0.022	0.009	(-0.031,-0.013)	2.5
					0.010	(-0.033, -0.012)	2.2
		1955-1958	-0.025	0.021	0.006	(-0.030,-0.019)	3.7
					0.007	(-0.032, -0.018)	3.1
	College Degree	1963-1967	-0.038	0.033	0.003	(-0.042, -0.035)	10.2
					0.005	(-0.043, -0.033)	6.8
		1959-1962	-0.029	0.025	0.005	(-0.034,-0.023)	4.5
					0.007	(-0.036, -0.022)	3.4
		1955-1958	-0.090^*	0.021	0.009	(-0.099,-0.081)	2.3
					0.012	(-0.102, -0.077)	1.7
	Study Est.		-0.037^*	0.009	0.007	(-0.044,-0.030)	1.2
					0.009	(-0.046, -0.028)	1.0
	U.S. Est.		-0.021^*	0.008	0.006	$(-0.027, -0.015)^+$	1.2
					0.008	$(-0.029, -0.013)^+$	1.0
White Females	High School	1963-1967	-0.058	0.039	0.014	(-0.072, -0.044)	2.8
					0.020	(-0.077, -0.039)	2.0
		1959-1962	-0.070	0.036	0.016	$(-0.086, -0.054)^+$	2.2
					0.023	$(-0.092, -0.047)^+$	1.6
		1955-1958	-0.060^*	0.027	0.005	(-0.066,-0.055)	5.2
					0.007	$(-0.067, -0.053)^+$	4.0
	Some College	1963-1967	-0.006	0.017	0.001	(-0.007, -0.005)	14.5
					0.002	(-0.008, -0.004)	9.7
		1959-1962	-0.009	0.026	0.006	(-0.015,-0.003)	4.4
					0.009	(-0.018,-0.001)	3.0
		1955-1958	-0.039	0.022	0.004	(-0.043,-0.035)	5.3
					0.005	(-0.044,-0.034)	4.3
	College Degree	1963-1967	-0.068^*	0.031	0.005	(-0.073, -0.063)	6.3
					0.005	(-0.073, -0.062)	5.7
		1959-1962	-0.046^*	0.023	0.009	$(-0.056, -0.037)^+$	2.5
					0.011	$(-0.058, -0.035)^+$	2.0
		1955-1958	-0.096^*	0.020	0.010	(-0.105,-0.086)	2.0
					0.012	(-0.108, -0.084)	1.6
	Study Est.		-0.057^*	0.009	0.008	(-0.065,-0.049)	1.2
					0.010	(-0.067, -0.047)	0.9
	U.S. Est.		-0.042^*	0.009	0.007	(-0.050, -0.036)	1.3
					0.009	(-0.052, -0.034)	1.0

[hP]

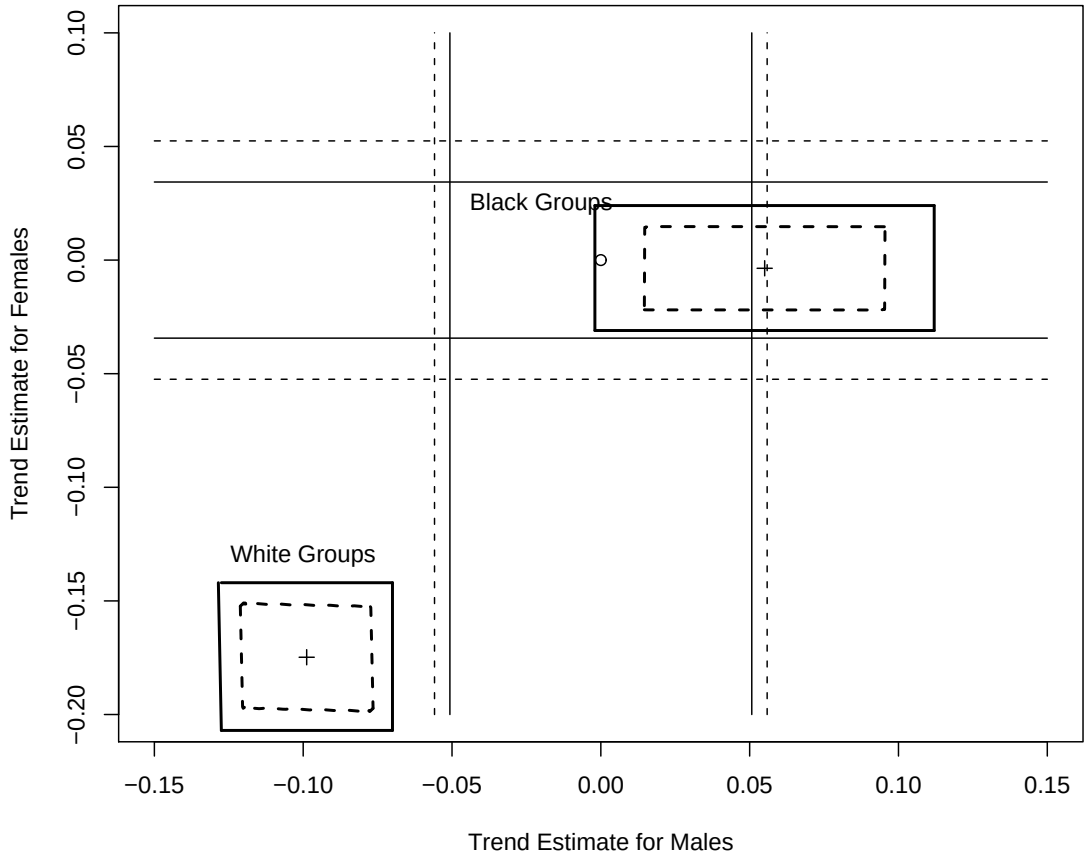

Figure 1. Graphical Representation of the Sensitivity of the Group-Specific Smoking Trend Estimates to Nonignorable Non-Monotone Nonresponse. The two points marked by the "+" sign are the MAR estimates of the smoking trend for the black and white groups, respectively. The dashed (solid) frames surrounding the two points denote the range of estimates when the nonignorability parameter γ_{1k} is varied between -1 and 1, using the lower (upper) bounds of ISNI values. Between the two vertical solid (dashed) lines is the region of non-significance where the smoking-trend estimate for the black (white) males is statistically insignificant from zero. Between the two horizontal solid (dashed) lines is the region of non-significance where the smoking-trend estimate for black (white) females is statistically insignificant from zero.

[hP]

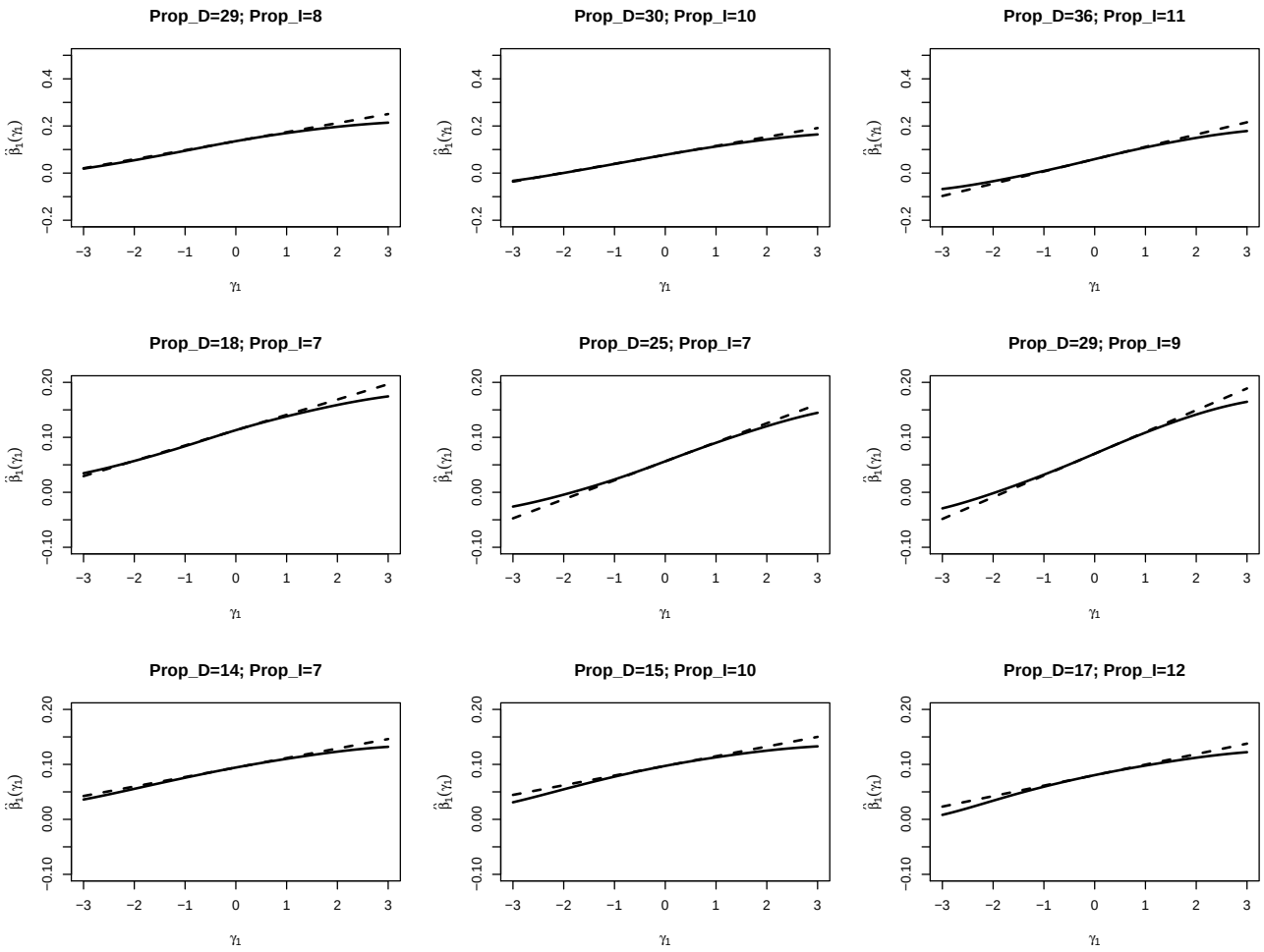

Figure 2. ISNI Approximation to $\hat{\beta}_1(\gamma_1)$ in Simulated Panel Datasets of Sample Size 300 with Non-Monotone Missingness. The three columns from left to right correspond to $\gamma_1 = (-0.5, 0, 0.5)$, and the three rows from top to bottom correspond to $\gamma_{01}^{20} = (-2.5, -3, -3.5)$. Prop_D (%) is the proportion of units that drop out of study; Prop_I (%) is the proportion of units that have intermittent missingness. The solid curves are $\hat{\beta}_1(\gamma_1)$; the straight lines are the ISNI tangent lines $\hat{\beta}_1(0) + ISNI * \gamma_1$. The average time to obtain $\hat{\beta}(\gamma_1)$ for a posited value of γ_1 was about three hours. In contrast, the calculation of ISNI took on average less than ten seconds.

[hP]

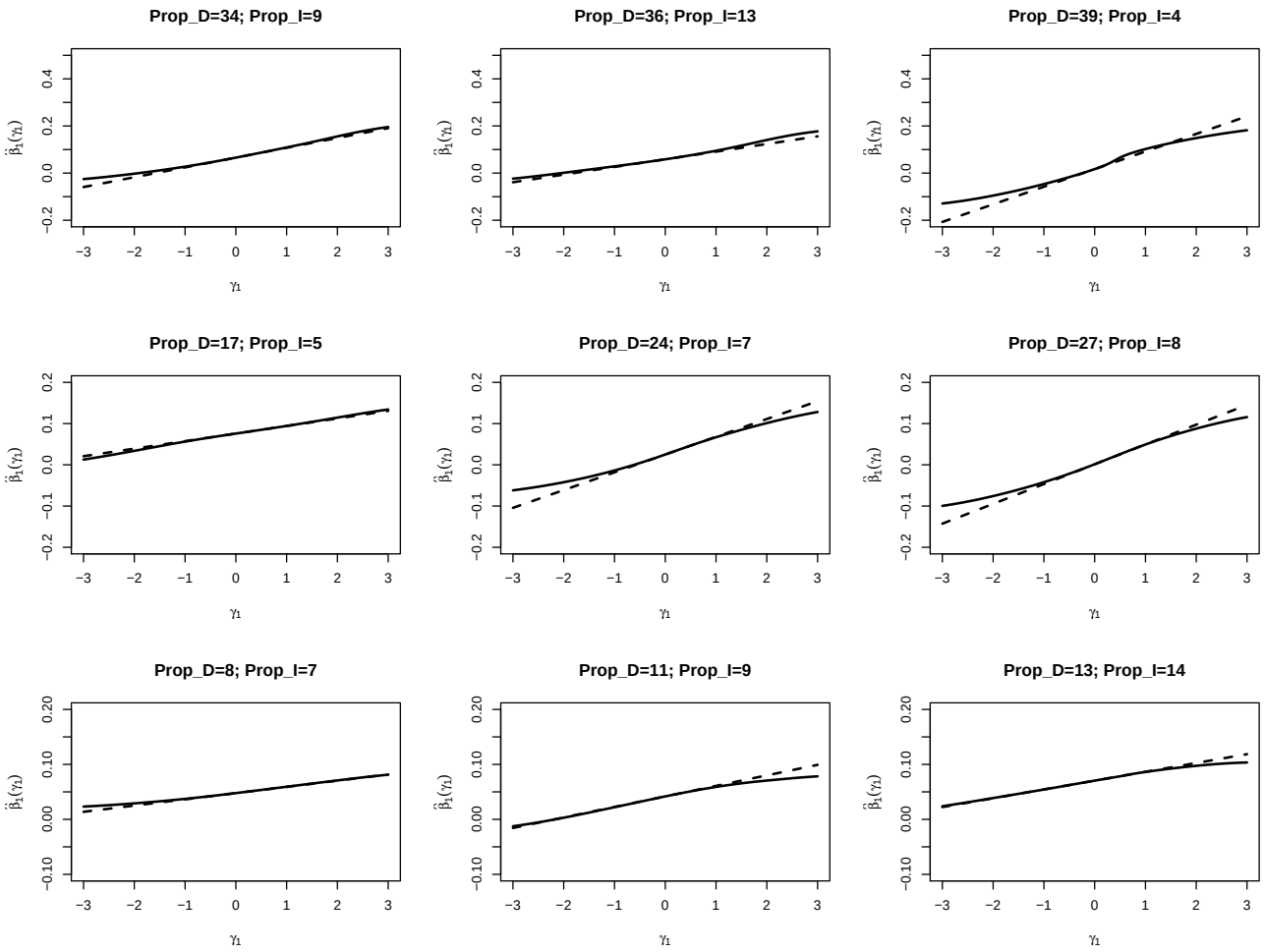

Figure 3. ISNI Approximation to $\hat{\beta}_1(\gamma_1)$ in Simulated Panel Datasets of Sample Size 100 with Non-Monotone Missingness. The three columns from left to right correspond to $\gamma_1 = (-0.5, 0, 0.5)$, and the three rows from top to bottom correspond to $\gamma_{01}^{20} = (-2.5, -3, -3.5)$. Prop_D (%) is the proportion of units that drop out of study; Prop_I (%) is the proportion of units that have intermittent missingness. The solid curves are $\hat{\beta}_1(\gamma_1)$; the straight lines are the ISNI tangent lines $\hat{\beta}_1(0) + ISNI * \gamma_1$. The average time to obtain $\hat{\beta}(\gamma_1)$ for a posited value of γ_1 was about one hour. In contrast, the calculation of ISNI took on average less than ten seconds.