Statistica Sinica **14**(2004), 1221-1237

AN INDEX OF LOCAL SENSITIVITY TO NONIGNORABILITY

Andrea B. Troxel[1], Guoguang Ma[2] and Daniel F. Heitjan[1]

[1] *University of Pennsylvania School of Medicine and*
[2] *Clinical Biostatistics, Merck & Co., Inc.*

Abstract: Standard methods of analysis can give misleading results when some observations are nonignorably missing. Analysts currently assess nonignorability by performing sensitivity analyses using models with and without a nonignorable component. Because this approach can involve complicated modeling and arduous computation, and can yield results that are highly sensitive to untestable model assumptions, there is a need for a simple screening tool that measures the potential impact of nonignorability on an analysis. We propose a measure based on a Taylor-series approximation to the nonignorable likelihood, evaluated at the parameter estimates under the assumption of ignorability. From this approximate likelihood, we derive an *index of sensitivity to nonignorability*, or *ISNI*. One can compute ISNI without estimating a nonignorable model or positing specific values of a nonignorability parameter. We interpret ISNI in terms of an intuitive parameter that captures the extent of sensitivity. We derive a general expression for ISNI in the generalized linear model with fully observed predictors and potentially missing outcomes. We illustrate the method with two regression examples.

Key words and phrases: Ignorability, missing at random, missing data, sensitivity analysis.

1. Introduction

A common model for data that are subject to incompleteness is the *selection model*, which factors the joint distribution of the variables of interest and missingness indicators into two parts: a *complete-data model* that describes the distribution of the underlying complete data, and a *missing-data mechanism* that describes the distribution of the missing-data indicators given the underlying complete data. In most cases we hope to ignore the missing observations by treating them as though we had no intention of collecting them in the first place; that is, we assume that our achieved sample size is the intended sample size and proceed with estimation and testing. We also want to ignore the missing-data mechanism in the sense of avoiding estimating its parameters. We say that the missing data mechanism is *ignorable* if the data and model are such that we can proceed in this way without imperiling the validity of our inferences.

Rubin (1976) defined a typology of incomplete-data models and related it to sufficient conditions for ignorability in the major modes of inference. Data are *missing completely at random* (*MCAR*) if the probability of the observed missingness indicator, conditional on the notional (i.e., intended) complete data, does not depend on the values of either the observed or missing observations. MCAR data are ignorable in the sense that frequentist inferences conditional on the observed pattern of missing data need not include a model for the missing-data mechanism. A model that generates only MCAR data is called an MCAR mechanism (Little and Rubin (2002)). Data are *missing at random (MAR)* if the probability of the observed missingness indicator, given the notional complete data, does not depend on the values of the missing observations. If the data are MAR and the parameters of the complete-data model and the missing-data mechanism are distinct, then Bayesian and direct-likelihood inferences that ignore the missing data mechanism are correct — that is, one can ignore any "nuisance" parameters in the missing data model and simply maximize a likelihood that is a function of the parameters of the complete-data distribution. A model that gives rise only to MAR data is said to be an MAR mechanism. The missingness is said to be nonignorable if inferences that ignore the missing data and its mechanism, in this sense, are invalid. See Heitjan and Basu (1996) for illustrative examples.

Departures from ignorability can substantially affect inferences, and in applications it is seldom obvious from the context whether an assumption of ignorability is justified. Currently, data analysts who face missing data can investigate nonignorability directly by positing and estimating nonignorable selection models (e.g., Schluchter (1992), Diggle and Kenward (1994), Troxel, Harrington and Lipsitz (1998)). Unfortunately, such models can be difficult to identify and estimate from real data, because one cannot directly assess the dependence of the missingness probability on observations that are themselves missing. Any hypothesis tests of ignorability must be carried out within the context of the model's assumptions, which are typically unverifiable from the observed data alone. Alternatively, one can fit pattern-mixture models (e.g., Little and Wang (1996); Hogan and Laird (1997)) or frailty models (e.g., Pulkstenis, Ten Have and Landis (1998)). To accommodate nonignorability, however, these models also require important structural assumptions that can be difficult to justify.

Another alternative is sensitivity analysis, in which one estimates models under a range of assumptions about the nonignorability parameters, and then assesses the impact of these parameters on key inferences. If there is little effect, the model is robust to specification of the missing data component, and the simpler, ignorable analysis is valid. Troxel (1998) and Scharfstein, Rotnitzky and Robins (1999) present analyses of this kind, in the context of selection models for likelihood-based and semiparametric inference, respectively. Little and

Wang (1996) and Daniels and Hogan (2000) have developed analogous methods for pattern-mixture model analysis. Sensitivity analysis avoids some of the uncertainties of direct nonignorable modeling, but it can still involve strong and unverifiable assumptions (e.g., assumptions about the complete data distribution). Moreover, computations can be difficult under some models.

Other authors have considered studying sensitivity only in the neighborhood of the MAR model. Copas and Li (1997) proposed an approach in which the complete-data component is a normal linear model whose error term is correlated with the underlying normal error term in a probit selection model for the missingness probability. Their nonignorability parameter is the correlation of the two errors, and they examined sensitivity of various inferences to small departures from MAR. Copas and Eguchi (2001) generalized this approach, defining a sensitivity parameter as the variance, with respect to the outcome, of the log odds of selection. Verbeke, et al. (2001) have applied the local influence approach of Cook (1986) to the problem of potentially nonignorable dropout in normally distributed longitudinal data. Their approach posits a separate nonignorability parameter for each subject, and measures sensitivity by the curvature of the graph of the nonignorability parameters and the *likelihood displacement*, which is the effect of a small perturbation from ignorability on the likelihood. With their method, one can identify the effects of nonignorability in individual observations, or collections of observations, on parameter estimates.

In this article, we propose a local sensitivity index that generalizes the approach of Copas and Li (1997) to encompass a general parametric complete-data model together with a general parametric selection model. Like theirs, our method focuses on the behavior of the MLE in the neighborhood of the MAR model, which we believe is the area of greatest interest in most applications. Because it is easy to compute — requiring one to fit only the MAR model and a single binary regression — we envision our index as a tool for screening data, allowing the analyst to evaluate the potential effects of nonignorability without actually fitting a nonignorable model. Advantages of our generalization compared to that of Copas and Eguchi (2001) include greater simplicity and intuitive appeal, as well as the computational ease noted above. Although we focus here on the important and pervasive problem of missing outcomes in generalized linear models, our general equations are valid in any parametric selection model.

We describe the model and index in Section 2. In Section 3, we derive a general expression for our index in the generalized linear model, and show explicit formulas for some key distributions; formulas for other distributions are given in Appendix 2. In Section 4, we describe the use and interpretation of the index. In Section 5 we illustrate the method with a complete data set from which we delete observations in a variety of patterns. In Section 6 we apply the

method to an incomplete data set that others have previously analyzed using nonignorable models. Section 7 concludes with further discussion of the model.

2. Model and Index Development

Consider estimation of a parameter θ of the conditional distribution of an outcome Y given predictors Z, whose density is $f_\theta^{Y_i|Z_i}(y_i|z_i)$, for independent subjects $i = 1, \ldots, n$. Following the notation of Heitjan (1994), let G_i be the completeness indicator that takes the value 1 for subjects who are observed and 0 for subjects who are missing, and assume that the probability of being observed depends on y_i and a set of predictors x_i (possibly, but not necessarily, overlapping with z_i). We assume that Y_i is independent of X_i given Z_i, i.e., that X_i serves only to inform the probability of missingness and contains no additional predictive information about Y_i.

We allow the probability of Y_i being observed to depend on the value of Y_i through a parameter γ_1, as follows:

$$\Pr_\gamma[G_i = 1|Y_i = y_i, X_i = x_i] = h(\gamma_0' x_i + \gamma_1 y_i), \tag{1}$$

where $h(\cdot)$ is a specified monotonic link function (for example, the logit or probit). Thus a value of $\gamma_1 = 0$ implies that the missing-data mechanism is MAR. This is a straightforward way to describe dependence on Y and — in the absence of specific knowledge about the form of that dependence — the most direct. The log likelihood under this model, $L(\theta, \gamma; y, g, z, x)$, is

$$\begin{aligned} L = \sum_{i=1}^{n} \Big\{ & g_i \left[\ln f_\theta^{Y_i|Z_i}(y_i|z_i) + \ln h(\gamma_0' x_i + \gamma_1 y_i) \right] \\ & + (1 - g_i) \ln \int f_\theta^{Y_i|Z_i}(u|z_i) \left[1 - h(\gamma_0' x_i + \gamma_1 u) \right] du \Big\}. \end{aligned}$$

In a sensitivity analysis, one evaluates the extent to which an estimate of θ for fixed γ_1, say $\hat\theta(\gamma_1)$, depends on the value of γ_1. Our notion is to execute this analysis in a neighborhood of the MAR model by determining the rate at which $\hat\theta(\gamma_1)$ departs from $\hat\theta_0 = \hat\theta(0)$ as γ_1 changes from zero. Consequently, we base our index on the derivative of $\hat\theta(\gamma_1)$ with respect to γ_1, evaluated at the ignorable model $\gamma_1 = 0$. We show these steps in detail below.

Note here that we do not assume that the model is correctly specified. Rather, we assume only that a data analyst judges the complete-data model to be a convenient approximation to reality, but is concerned that conclusions based on the model (for example, in the form of the MLE) may be substantially misleading if the missingness mechanism is nonignorable. The purpose of the sensitivity analysis is to determine whether inferences based on the model in

question are sensitive, not to determine whether all possible inferences based on the data are sensitive.

To express the MLE as a function of γ_1, we first expand L around $\theta = \hat{\theta}_0$ (the MLE of θ assuming ignorability), $\gamma_0 = \hat{\gamma}_{00}$ (the MLE of γ_0 assuming ignorability) and $\gamma_1 = 0$ as follows:

$$L(\theta, \gamma_0, \gamma_1) \approx L(\hat{\theta}_0, \hat{\gamma}_{00}, 0) + \left[(\theta - \hat{\theta}_0)', (\gamma_0 - \hat{\gamma}_{00})', \gamma_1\right] \nabla L + \frac{1}{2}\left[(\theta - \hat{\theta}_0)', (\gamma_0 - \hat{\gamma}_{00})', \gamma_1\right] \nabla^2 L \left[(\theta - \hat{\theta}_0)', (\gamma_0 - \hat{\gamma}_{00})', \gamma_1\right]',$$

where

$$\nabla L = \left.\begin{pmatrix} \frac{\partial L}{\partial \theta} \\ \frac{\partial L}{\partial \gamma_0} \\ \frac{\partial L}{\partial \gamma_1} \end{pmatrix}\right|_{\theta=\hat{\theta}_0, \gamma_0=\hat{\gamma}_{00}, \gamma_1=0},$$

$$\nabla^2 L = \left.\begin{pmatrix} \frac{\partial^2 L}{\partial \theta \partial \theta'} & \frac{\partial^2 L}{\partial \theta \partial \gamma_0'} & \frac{\partial^2 L}{\partial \theta \partial \gamma_1} \\ \frac{\partial^2 L}{\partial \gamma_0 \partial \theta'} & \frac{\partial^2 L}{\partial \gamma_0 \partial \gamma_0'} & \frac{\partial^2 L}{\partial \gamma_0 \partial \gamma_1} \\ \frac{\partial^2 L}{\partial \gamma_1 \partial \theta'} & \frac{\partial^2 L}{\partial \gamma_1 \partial \gamma_0'} & \frac{\partial^2 L}{\partial \gamma_1^2} \end{pmatrix}\right|_{\theta=\hat{\theta}_0, \gamma_0=\hat{\gamma}_{00}, \gamma_1=0}.$$

We rewrite ∇L as $\{\nabla L_i\}_{i=1,2,3}$ and $\nabla^2 L$ as $\{\nabla^2 L_{ij}\}_{i,j=1,2,3}$ in an obvious notation. Appendix 1 presents general formulas for the elements of ∇L and $\nabla^2 L$ in our selection model.

Next we write the log likelihood as a function of θ and γ_0 for fixed γ_1 and differentiate, finding the maximum likelihood estimate $(\hat{\theta}, \hat{\gamma}_0)$ as a function of $\hat{\theta}_0, \hat{\gamma}_{00}$, and γ_1:

$$\begin{pmatrix} \hat{\theta}(\gamma_1) \\ \hat{\gamma}_0(\gamma_1) \end{pmatrix} \approx \begin{pmatrix} \hat{\theta}_0 \\ \hat{\gamma}_{00} \end{pmatrix} - \gamma_1 \begin{pmatrix} \nabla^2 L_{11} & \nabla^2 L_{12} \\ \nabla^2 L_{21} & \nabla^2 L_{22} \end{pmatrix}^{-1} \begin{pmatrix} \nabla^2 L_{13} \\ \nabla^2 L_{23} \end{pmatrix}.$$

The index of sensitivity to nonignorability (ISNI) is the derivative of $\hat{\theta}$ with respect to γ_1, evaluated at $\gamma_1 = 0$:

$$\text{ISNI} = \left.\frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_1}\right|_{\gamma_1=0} = -\left(\nabla^2 L_{11}\right)^{-1} \nabla^2 L_{13}.$$

This quantity describes the rate of change of $\hat{\theta}$ with respect to γ_1 — that is, the amount by which a unit change in the nonignorability parameter displaces the MLE of θ from its value $\hat{\theta}_0$ under the MAR model. Note that the first factor in ISNI is just the estimated variance-covariance matrix of $\hat{\theta}$ under MAR, and the second factor is a measure of the orthogonality of θ and γ_1.

One can apply this general development to any parametric model by specifying $f_\theta^{Y_i|Z_i}(\cdot|\cdot)$ and $h(\cdot)$. In the next section we derive some special cases within the class of generalized linear models.

3. ISNI in Generalized Linear Models

The generalized linear model (McCullagh and Nelder (1989)) assumes that the components Y_i, $i = 1, \ldots, n$ of a random vector Y are independent with densities

$$f_\theta^{Y_i}(y_i) = \exp\Big\{\Big[y_i\lambda_i(\theta_1) - b(\lambda_i(\theta_1))\Big]a(\theta_2) + c(y_i, \theta_2)\Big\},$$

where λ is a vector of canonical parameters; functions $b(\cdot)$ and $c(\cdot)$ determine a particular family; θ_1 is a vector of regression coefficients, denoted $\theta_1 = (\theta_{11}, \ldots, \theta_{1p})'$, attached to predictors $z = (z_1, \ldots, z_p)$ through a linear predictor $\eta = \sum_{j=1}^p \theta_{1j} z_j$; θ_2 is a dispersion parameter; and $a(\theta_2)$ is commonly of the form $a(\theta_2) = \theta_2/w$ where w is a known weight. A link function relates the linear predictor η to the expected value μ of Y, i.e., $g(\mu) = \eta = \theta_1' z$.

We assume that the complete-data model is of this type and that the probability of being observed (the selection model) is of the form given above with a logistic link for h:

$$\Pr_\gamma[G_i = 1 | Y_i = y_i, X_i = x_i] = h(\gamma_0' x_i + \gamma_1 y_i) = \frac{\exp(\gamma_0' x_i + \gamma_1 y_i)}{1 + \exp(\gamma_0' x_i + \gamma_1 y_i)}. \qquad (2)$$

Note that for the logistic model, $h' = h(1-h)$ and $h'' = h(1-h)(1-2h)$. Again, the selection-model predictors x_i need not overlap with the complete-data model predictors z_i.

Following the development in the previous section, a general expression for ISNI in the generalized linear model with a logistic selection model is

$$\text{ISNI} = \begin{pmatrix} \Big[\sum g_i(y_i \frac{\partial^2 \lambda_i}{\partial \theta_1^2} - \frac{\partial^2 b}{\partial \theta_1^2})a(\theta_2)\Big]^{-1} \sum (1-g_i) h_i \frac{\partial^2 b}{\partial \lambda_i \partial \theta_1} \\ 0 \end{pmatrix}_{\hat\theta_0 = (\hat\theta_{10}, \hat\theta_{20})}, \qquad (3)$$

where $\hat\theta_{10}$ and $\hat\theta_{20}$ are ML estimates of θ_1 and θ_2 assuming $\gamma_1 = 0$, and h_i equals $h(\hat\gamma_{00} x_i)$ as before. Below and in Appendix 2, we present formulas for ISNI in a range of specific generalized linear models. Our assumption that $h(\cdot)$ is linear in Y also implies that $\text{Var}\,(Y|G) = \text{Var}\,(Y)$ (see Little and Wang (1996)) and the estimated dispersion parameter therefore requires no adjustment; thus we will present only the ISNI component for the regression parameter θ_1.

3.1. Univariate normal data

We consider first the case of independent, normally distributed observations with common mean μ and variance τ. Applying (3) with an identity link, we obtain $\text{ISNI} = -\hat\tau_0 \left(\sum g_i\right)^{-1} \sum (1-g_i) h_i$, where $\hat\mu_0 = \sum_{i=1}^n g_i y_i / \sum_{i=1}^n g_i$ and

$\hat{\tau}_0 = \sum_{i=1}^n g_i(y_i - \hat{\mu}_0)^2 / \sum_{i=1}^n g_i$ are the usual observed-data MLEs of the mean and variance. If there are no predictors in the logistic regression for the missingness mechanism, then $\sum(1-g_i)h_i = n_o n_m/n$, where n is the total number of potential subjects and n_o and n_m are the numbers of subjects with y_i observed and missing, respectively. ISNI then simplifies to ISNI $= -(n_m/n)\hat{\tau}_0$. That is, the index for the mean is minus the variance estimate times the proportion of missing observations, and consequently the sensitivity in this case is proportional to the fraction of data missing.

3.2. Normal linear model

Assume now that $Y_i \sim \mathcal{N}(z_i\beta, \tau)$. Then the index for β is ISNI$= -\hat{\tau}_0(\sum g_i z_i z_i')^{-1} \sum(1-g_i)h_i z_i$ or, equivalently, ISNI $= -\hat{\tau}_0 (Z_o' Z_o)^{-1} Z_m' h_m$, where Z_o and Z_m are the matrices of predictors for subjects with $g_i = 1$ and $g_i = 0$, respectively; h_m is the vector of h_i values for subjects with $g_i = 0$, that is, the vector of propensity scores for the missing subjects; and $\hat{\tau}_0$ is the variance MLE.

Formulas for ISNI in other GLMs appear in Appendix 2.

3.3. Computing

We have written code in S-Plus (Insightful Corporation; Seattle, WA) for computing ISNI in a generalized linear model with missing outcome data. Code and data are available at http://www.cceb.upenn.edu/heitjan/isni.

4. Use and Interpretation of ISNI

Recall that γ_1 is the parameter linking the outcome y_i to the observation indicator in the nonignorable model. For ease of interpretation, assume henceforth that we are using the logistic selection model, in which case γ_1 is the log odds ratio in the observation probability associated with a one-unit change in y. If $\gamma_1 = 1$, then a unit change in y_i implies that the odds of being observed increases by a factor of 2.7. This linkage of interpretation to a unit change in y makes the scale of y relevant, as noted below. Because ISNI is the derivative of $\hat{\theta}$ with respect to γ_1, we can approximate the value of $\hat{\theta}$ for fixed γ_1 by $\hat{\theta}(\gamma_1) \approx \hat{\theta}_0 + \text{ISNI}\gamma_1$. Thus, if the nonignorability is such that $\gamma_1 = 1$, we should adjust each MAR parameter estimate by the corresponding element of the ISNI vector. When doing preliminary screening for sensitivity to nonignorability, an appealing parameter to consider is the ratio of ISNI to the standard error (SE) of a coefficient of interest. If this ratio exceeds 1, then the model is highly susceptible to nonignorability. In general, both the sample size and the fraction of missing information will be important, as can be seen directly from the derivation of ISNI in Section 3.1 for the univariate normal model.

In models where there is only a single natural scale for the outcome, such as the Poisson and binomial, one can interpret ISNI directly in this way. For continuous outcomes, however, where the outcome may be denominated in various units, the linkage of the log odds interpretation of γ_1 to the units of y makes the scale of the outcome relevant.

Consider transforming Y to aY, and denote the ISNI on the transformed scale by ISNI_{aY} and the SE for a parameter of interest by SE_{aY}. For a linear regression coefficient, it is easy to show that

$$\text{ISNI}_{aY}/\text{SE}_{aY} = a\text{ISNI}_Y/\text{SE}_Y.$$

Now suppose we transform from Y to aY/σ_Y, where σ_Y is the standard deviation (SD) of Y. The ISNI calculated on this scale, $\text{ISNI}_{aY/\sigma_Y}$, is the sensitivity when a change of $1/a$ standard deviations in Y is associated with an odds ratio of 2.7. One way to quantify the sensitivity is to determine for which value of a in this formulation the ISNI is exactly equal to the SE of the regression parameter. Therefore we propose to calculate a transformation that we call c, or the *sensitivity transformation*, defined to be that transformation for which $\text{ISNI}_{cY/\sigma_Y}/\text{SE}_{cY/\sigma_Y} = 1$. That is, if a change of $1/c$ standard deviations of Y is associated with an odds ratio of 2.7 in the observation probability, then the analysis is sensitive to nonignorability in the sense that ISNI equals the SE of a regression coefficient of interest. Solving, we have

$$c = |\sigma_Y \text{SE}_Y/\text{ISNI}_Y|;$$

we take absolute value because the sign of the transformation is irrelevant. Note that c is scale-independent.

An alternative interpretation is that c is the scale on which the sensitivity is extreme enough that an odds ratio of 2.7 corresponds to an effect on $\hat{\theta}$ of one standard error. This can be viewed as the minimal nonignorability that would lead to sensitivity, in units of 1/SD. With this interpretation, c is similar to the parameters of sensitivity to confounding bias in observational studies, as described by Rosenbaum (1987, 1995). Our sensitivity transformation is measured in terms of inverse standard deviations, but has a similar interpretation as a scale on which nonignorability is seen to substantially affect inference, in our case by causing a substantial change in $\hat{\theta}$.

If c is large, then there is sensitivity only if the nonignorability is extreme, whereas if c is small, say less than 1, there is potential sensitivity even for modest nonignorability. For example, $c = 10$ indicates that the data are sensitive to nonignorability if a 0.1 SD change in Y substantially changes the odds of being observed. This is severe nonignorability indeed, and stronger than one would expect in most practical situations. Compare this to $c = 0.2$, for which sensitivity

results when a 5-SD change in Y substantially affects the odds of being observed. Because such a degree of nonignorability seems plausible in many studies, such a low value of c would cause concern, and could trigger more detailed sensitivity analysis or nonignorable modeling.

5. Example 1: Smoking and Mortality Data

In this first example, we show how various configurations of missing observations, all derived from a single complete data set, can affect ISNI and therefore our evaluation of sensitivity. We apply ISNI to a data set with actual missing data and all its attendant complications in Section 6.

Our first example concerns data on smoking and mortality in England and Wales in the early 1970s (Moore and McCabe (1989), citing *Occupational Mortality: The Registrar Generals' Decennial Supplement for England and Wales, 1970–72*, Her Majesty's Stationery Office, London (1978)). The data summarize a study of men in 25 occupational groups in England using two indices: the smoking index is the ratio of the average number of cigarettes smoked per day by men in the occupational group to the average number of cigarettes smoked per day by all men; the mortality index is the ratio of the rate of deaths from lung cancer among men in the occupational group to the rate of deaths from lung cancer among all men. There are no missing values. Figure 1 shows a scatter plot of the complete data set along with the regression line relating smoking to mortality. From the complete data, the intercept is -2.885 (SE of 23.034) and the slope is 1.088 (SE of 0.221).

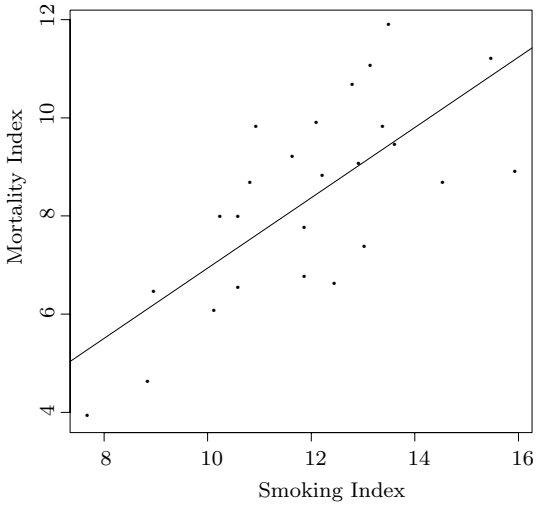

Figure 1. Smoking and mortality data with complete-data regression line.

For demonstration purposes, we ordered the data according to the values of the smoking index (the predictor). We sequentially selected outcome data to be "missing," first singly and then in groups of five. We then calculated the parameter estimates and their SEs from a linear regression of mortality on smoking, using the "observed" data only (the ignorable analysis). Then, we calculated ISNI as $-\hat{\tau}_0(Z_o'Z_o)^{-1}Z_m'h_m$. Here h_m is the vector of propensity scores, or predicted probabilities of being observed, for the units who have missing data on mortality, the outcome variable. Table 1 gives an ISNI analysis of the slope for the data with one value missing at selected places (point 1, point 2 and point 12) and with five consecutive values missing, also at selected places (points 1-5, points 3-7, points 10-14 and points 20-24). Because the logistic regression of the observation probability on x gives an infinite slope when all the missing values are grouped at one end, the propensity scores for the missing subjects in this configuration are all 0, and ISNI also equals 0. This makes sense, because when x is a perfect predictor of observation status there is no scope for outcome to further modify these predictions.

With single points deleted, the ISNI for the slope is largest when points 2 and 24 (not shown) are missing and is smallest when point 12 is missing. On the whole, the c values are large and suggest only modest sensitivity. For example when we omit point 2, $c = 4.77$, suggesting that the slope would be insensitive unless a change of 1/4.77 SDs was associated with an odds ratio of 2.7 in the observation probability, which is very strong nonignorability. Note that missing values in the middle of the range generally have a smaller potential for sensitivity, because such points also generally have less influence on the slope. In any case, missing a single point in this dataset can have little effect on slope estimation.

With five adjacent points deleted, ISNI is largest for points 3−7. The sensitivity transformation in this case is 1.73, which says that we begin to have sensitivity when a change of 1/1.73 SDs is associated with an odds ratio of 2.7 in the observation probability. This is clearly a stronger sensitivity than we saw with one point omitted, and it could be enough to give an investigator pause about trusting an MAR analysis.

We observe that in general ISNI is positive when values from the left side of the graph are missing and negative when values from the right side of the graph are missing. The ISNI in this case estimates the change in the slope that would be associated with a γ_1 value of 1, which suggests that the observation probability *increases* with increasing Y (see (1)). Therefore, the presence of a missing value suggests that the actual value is smaller than the MAR prediction for that observation. Missing values at the left of the graph thus suggest that the slope is larger than its MAR estimate, and consequently are consistent with a positive ISNI. Missing values at the right of the graph have the opposite effect, and consequently give negative ISNI.

Table 1. Smoking and mortality data with missing points.

Missing Point(s)	Slope Estimate	SE	$\hat{\tau}_0$	ISNI	Sensitivity Transformation
1	0.967	0.235	315	0.0000	∞
2	1.000	0.222	313	1.1435	4.77
12	1.085	0.210	315	0.0390	141.94
1−5	0.655	0.317	315	0.0000	∞
3−7	1.088	0.265	374	4.1161	1.73
10−14	1.086	0.222	347	0.0685	91.73
20−24	1.036	0.261	324	−3.3479	1.92

6. Example 2: A Survey of Sexual Behavior

Table 2 presents data from a survey of students at the University of Edinburgh conducted during school registration in October 1993. The goal of the study was to estimate the prevalence of HIV infection and its association with HIV risk factors such as sexual practices, which some participants declined to divulge (Raab et al. (1995)). Raab and Donnelly (1999) considered the estimation of sexual activity for the whole population, focusing on the answer to the question "Have you ever had sexual intercourse?". A simplified data set consisted of the answer to this question, together with the student's gender and faculty as predictors. To enable comparison with Raab and Donnelly (1999), the variable faculty was categorized as medical (medical, dental or veterinary) *versus* all others. The last column of Table 2 shows the percent of respondents in each category. Overall, 3,828 (62.4%) of 6,136 students responded, with the percent responding about 16% higher among medical students than among non-medical students.

Assuming ignorable missingness, we fit a logistic model (using responders only) to predict the outcome by gender, faculty and their interaction. Estimated regression coefficients appear in Table 3, together with ML estimates from a nonignorable model fitted by Raab and Donnelly (1999). The estimates show that for both the ignorable and nonignorable models, students in a medical faculty were less likely to report having had sexual intercourse. In a model that assumes that the missingness probability depends on gender, faculty, and outcome, their ML estimate of the coefficient for outcome is $\hat{\gamma}_1 = -\infty$; the estimated percent who have had intercourse is 83.3%, compared to only 73.3% among the responders. We used ISNI to assess the extent to which the logistic regression coefficients are sensitive to nonignorability, calculating ISNI by the formula for grouped binomial data (Appendix 2). Results appear in Table 3.

A negative value of $\hat{\gamma}_1$ implies that a student whose correct answer was "yes" was more likely to avoid giving an answer. Raab and Donnelly's estimates are

consistent with this prediction, as their estimate $\hat{\gamma}_1$ is negative, and the observed fraction responding yes underestimates the MLE under the nonignorable model. Our values of ISNI for each parameter are of opposite sign to the differences in parameter estimates under the nonignorable models, confirming that ISNI accurately detects the direction of the sensitivity. For example, the ISNI for the Faculty covariate is 0.17, indicating that if $\gamma_1 = 1$, the MLE for the estimate should change from -0.73 to -0.56. On the other hand, if $\gamma_1 = -1$, as is more plausible here, the estimate would change from -0.73 to -0.9, which is similar to the change actually observed. The ratio of ISNI to the coefficient standard error tells us that the intercept and faculty coefficients are most sensitive to nonignorability, as indeed Raab and Donnelly's estimates show that they are. Raab and Donnelly's estimates also confirm ISNI's suggestion that neither the gender coefficient nor the interaction is sensitive.

Raab and Donnelly's model, like other such nonignorable models, requires lengthy calculations that can be highly unstable. One could achieve the same estimates via sensitivity analysis by setting $\gamma_1 = -\infty$, avoiding these difficult numerical problems. Our index is also simple to compute, and it ably detects the sign and the rough magnitude of the potential sensitivity to nonignorable missingness while avoiding some of the pitfalls of the full nonignorable model.

Table 2. Sexual behavior data set.

		Response			
Faculty	Gender	No	Yes	Nonrespondents	Percent Responding
Non-medical	Male	433	1277	1189	59.0%
	Female	410	1247	978	62.9%
Medical	Male	89	126	68	76.0%
	Female	94	152	73	77.1%

Table 3. ML estimates of ignorable and nonignorable models and ISNI analysis.

	Estimates of Regression Coefficients			
	Intercept	Gender	Faculty	Faculty*Gender
Ignorable Model	1.08	0.03	−0.73	0.10
SE of Coefficients	0.06	0.08	0.15	0.21
Nonignorable Model	1.74	−0.05	−0.95	0.14
ISNI	−0.41	0.04	0.17	−0.03
|ISNI/SE|	7.39	0.50	1.16	0.16

7. Discussion

We have developed an index of local sensitivity of the MLE to nonignorable missingness. Our index is simple to compute, involving only the calculation of the MLE of the complete-data model parameters under MAR, together with a model for predicting completeness from available predictors. We envision our index as a screening tool that will allow the data analyst to assess the need for more elaborate sensitivity analysis or nonignorable modeling.

The distinction between an MAR dataset and an MAR mechanism, noted in the Introduction, is not academic. For example, suppose that a selection model in which Y and G are correlated (and θ and γ are distinct) gives rise by chance to a dataset that has no missing observations. Such a dataset is MAR, and the likelihood factors into two parts — one depending on θ alone and the other on γ alone. For such a dataset ISNI $= 0$, and Bayesian and direct-likelihood inferences about θ are completely insensitive to nonignorability, in that $\hat{\theta}(\gamma_1)$ does not depend on γ_1. As the fraction of missing data increases from zero, we expect the sensitivity to nonignorability, together with ISNI, to increase as well. Note that under this same model the data are not MCAR and moreover the missingness is nonignorable for frequentist inferences even if no data are missing (Rubin (1976) and Heitjan and Basu (1996)). Thus in this sense ISNI is really a Bayesian diagnostic, in that it measures sensitivity of summaries of the observed likelihood function, and not of other possible likelihoods that might have been observed but were not.

In applied statistics there is always a strong presumption in favor of MAR models, which are — even when nontrivial — by far easier to fit, understand and explain than the simplest nonignorable model. Moreover, fully observed predictor variables can often go far in explaining the probability of dropout, in which case small values of the nonignorability parameter are most relevant (see David, Little, Samuhel and Triest (1996)). Thus although a local sensitivity analysis is admittedly less informative than a global analysis, we believe that the neighborhood of the MAR model is the area of greatest interest, and should be the starting point for any analysis of nonignorability.

In principle, the form of the selection model can have considerable influence on the assessment of sensitivity, and in most cases the selection model that one uses in an ISNI analysis should be considered only provisional. We conjecture, however, that selection models that are monotone in the outcome are likely to give similar predictions in the vicinity of the MAR model; consequently, common models such as the logit and probit should yield similar local analyses. When the selection model is non-monotone in the outcome (as in reported income data, where nonresponse is more common in both tails), a model that, like ours, assumes monotonicity may give misleading results. Yet there are many

other examples where a monotone mechanism may be appropriate, as in medical studies where the probability of a missing observation may be correlated with the health state being measured (Troxel (1998)).

In this article we have derived detailed equations for likelihood inference on generalized linear models with missing data on the outcome variable. The general ISNI equation can easily accommodate essentially any model, and the idea should be straightforward to carry through, numerically if not analytically, for more general patterns of missing data and for other forms of incompleteness such as censoring and rounding (Heitjan and Rubin (1991) and Heitjan (1994, 1997)). The exploration of sensitivity in frequentist inferences involves distinct concepts (Heitjan and Basu (1996)), but similar mathematical tools will apply.

Acknowledgements

The US Public Health Service supported this research under grants CA-13696, CA-97470 and HL-68074. We obtained the mortality data from the Statlib Data and Story Library. We thank Professor Gillian Raab for kindly providing the sexual behavior data, and Professors Daniel Rabinowitz and Donald Rubin and two anonymous referees for valuable suggestions.

Appendix 1. Additional Details

Let $h'(u) = dh(u)/du$, $h''(u) = d^2h(u)/du^2$, $h_i = h(\hat{\gamma}_{00}'x_i)$, $h_i' = h'(\hat{\gamma}_{00}'x_i)$ and $h_i'' = h''(\hat{\gamma}_{00}'x_i)$. Letting $\mathrm{E}_{\hat{\theta}_0}^{Y_i|Z_i}$ be the conditional mean of Y_i given Z_i, the terms in ∇L and $\nabla^2 L$ take the following forms:

$$\nabla L_1 = \left.\frac{\partial L}{\partial \theta}\right|_{\hat{\theta}_0,\hat{\gamma}_{00},0} = \sum g_i \left.\frac{\partial \ln f_\theta^{Y_i|Z_i}(y_i|z_i)}{\partial \theta}\right|_{\hat{\theta}_0} = 0$$

$$\nabla L_2 = \left.\frac{\partial L}{\partial \gamma_0}\right|_{\hat{\theta}_0,\hat{\gamma}_{00},0} = \sum z_i g_i \frac{h_i'}{h_i} - \sum z_i \frac{(1-g_i)h_i'}{1-h_i}$$

$$\nabla L_3 = \left.\frac{\partial L}{\partial \gamma_1}\right|_{\hat{\theta}_0,\hat{\gamma}_{00},0} = \sum g_i y_i \frac{h_i'}{h_i} - \sum \frac{(1-g_i)h_i'}{1-h_i} \mathrm{E}_{\hat{\theta}_0}^{Y_i|Z_i}$$

$$\nabla^2 L_{11} = \left.\frac{\partial^2 L}{\partial \theta \partial \theta'}\right|_{\hat{\theta}_0,\hat{\gamma}_{00},0} = \sum g_i \left.\left(\frac{\partial^2 \ln f_\theta^{Y_i|Z_i}(y_i|z_i)}{\partial \theta \partial \theta'}\right)\right|_{\hat{\theta}_0}$$

$$\nabla^2 L_{12} = \left.\frac{\partial^2 L}{\partial \theta \partial \gamma_0'}\right|_{\hat{\theta}_0,\hat{\gamma}_{00},0} = 0$$

$$\nabla^2 L_{13} = \left.\frac{\partial^2 L}{\partial \theta \partial \gamma_1}\right|_{\hat{\theta}_0,\hat{\gamma}_{00},0} = -\sum \frac{(1-g_i)h_i'}{1-h_i} \left.\frac{\partial \mathrm{E}_\theta^{Y_i|Z_i}}{\partial \theta}\right|_{\hat{\theta}_0}$$

$$\nabla^2 L_{22} = \left.\frac{\partial^2 L}{\partial\gamma_0\partial\gamma_0'}\right|_{\hat\theta_0,\hat\gamma_{00},0} = \sum g_i z_i z_i' \frac{h_i'' h_i - h_i'^2}{h_i^2} - \sum (1-g_i) z_i z_i' \frac{h_i''(1-h_i) + h_i'^2}{(1-h_i)^2}$$

$$\nabla^2 L_{23} = \left.\frac{\partial^2 L}{\partial\gamma_0\partial\gamma_1}\right|_{\hat\theta_0,\hat\gamma_{00},0} = \sum z_i g_i y_i \frac{h_i'' h_i - h_i'^2}{h_i^2} - \sum z_i (1-g_i) \mathrm{E}_{\hat\theta_0}^{Y_i|Z_i} \frac{h_i''(1-h_i) + h_i'^2}{(1-h_i)^2}$$

$$\nabla^2 L_{33} = \left.\frac{\partial^2 L}{\partial\gamma_1^2}\right|_{\hat\theta_0,\hat\gamma_{00},0} = \sum g_i y_i^2 \frac{h_i'' h_i - h_i'^2}{h_i^2} - \sum (1-g_i) \left[\left(\mathrm{E}_{\hat\theta_0}^{Y_i|Z_i}\right)^2 \frac{h_i'^2}{(1-h_i)^2} + \mathrm{E}_{\hat\theta_0}^{Y_i^2|Z_i} \frac{h_i''}{1-h_i} \right].$$

Appendix 2. ISNI for Additional Distributions

Binomial Distribution

Suppose that the Y_i's are n independent Bernoulli trials with $\Pr(Y_i = 1) = \pi_i$ and $\Pr(Y_i = 0) = 1 - \pi_i$, $i = 1, \ldots, n$. Assuming a dispersion parameter of 1 and a logit link, $g(\pi_i) = \ln \pi_i/(1-\pi_i) = \eta_i = \sum_{j=1}^p \theta_{1j} z_{ij}$, the ISNI for the regression parameter θ_1 is

$$\mathrm{ISNI} = -\left[\sum g_i \frac{\exp(\hat\theta_{10}' z_i)}{[1+\exp(\hat\theta_{10}' z_i)]^2} z_i z_i' \right]^{-1} \sum (1-g_i) h_i \frac{\exp(\hat\theta_{10}' z_i)}{[1+\exp(\hat\theta_{10}' z_i)]^2} z_i,$$

where θ_{10} is the ML estimate of θ_1 assuming $\gamma_1 = 0$.

Individuals may share the same covariate vectors and form covariate classes, what McCullagh and Nelder (1989) call *grouped data*. Suppose the data are grouped into K covariate classes, with values z_k for the complete-data model predictors and x_k for the selection model predictors. Let n_{ok} be the number of respondents in class k and n_{mk} be the number of nonrespondents in class k, where $\sum_{k=1}^K (n_{ok} + n_{mk}) = N$. Let $h_k = h(\hat\gamma_{00} x_k)$, for $k = 1, \ldots, K$. With the logistic link in the selection model, the index reduces to

$$\mathrm{ISNI} = -\left[\sum_k n_{ok} \frac{\exp(\hat\theta_{10}' z_k)}{[1+\exp(\hat\theta_{10}' z_k)]^2} z_k z_k' \right]^{-1} \sum_k n_{mk} h_k \frac{\exp(\hat\theta_{10}' z_k)}{[1+\exp(\hat\theta_{10}' z_k)]^2} z_k.$$

Note that with this model the values of x_k and z_k need not be equal, but the classes must be the same.

Poisson Distribution

The canonical Poisson model assumes a log link for the mean, $g(\mu_i) = \ln(\mu_i) = \eta_i = \sum_{j=1}^p \theta_{1j} z_{ij}$, with the variance equal to the mean. Assuming $a(\theta_2) = 1$, the index for θ_1 is

$$\mathrm{ISNI} = -\Big[\sum g_i \exp(\hat\theta_{10}' z_i) z_i z_i' \Big]^{-1} \sum (1-g_i) h_i \exp(\hat\theta_{10}' z_i) z_i.$$

Gamma Distribution

Assume that Y_i follows the gamma distribution with a constant coefficient of variation, or a constant index ν, and parameter μ_i. Let $\theta_2 = \nu$, and $a(\theta_2) = \theta_2$. The canonical link is the reciprocal transformation $g(\mu_i) = \mu_i^{-1} = \eta_i = \sum_{j=1}^{p} \theta_{1j} z_{ij}$. The index is then

$$\text{ISNI} = \frac{1}{\hat{\theta}_{20}} \Big[\sum g_i (\hat{\theta}'_{10} z_i)^{-2} z_i z_i{}' \Big]^{-1} \sum (1 - g_i) h_i (\hat{\theta}'_{10} z_i)^{-2} z_i.$$

An approximate MLE for θ_2 is $\hat{\theta}_{20} \approx (6 + 2\overline{D})/\overline{D}(6 + \overline{D})$, where $D(y; \hat{\mu})$ is the deviance of the model for sensitivity under MAR and $\overline{D} = D(y; \hat{\mu})/n_o$, where n_o is the number of observed values (McCullagh and Nelder (1989)).

Inverse Gaussian Distribution

The inverse Gaussian density represents a wide class of distributions for continuous observations (Chhikara and Folks (1989)). Denote a variable Y_i from the inverse Gaussian distribution as $Y_i \sim \text{IG}(\mu_i, \tau)$, where τ is the variance. Let $\theta_2 = \tau$, and $a(\theta_2) = 1/\theta_2$. Under its canonical link, $g(\mu_i) = \mu_i^{-2} = \eta_i = \sum_{j=1}^{p} \theta_{1j} z_{ij}$, we obtain

$$\text{ISNI} = 2\hat{\theta}_{20} \Big[\sum g_i \Big(\hat{\theta}'_{10} z_i \Big)^{-3/2} z_i z_i{}' \Big]^{-1} \sum (1 - g_i) h_i \Big(\hat{\theta}'_{10} z_i \Big)^{-3/2} z_i.$$

We can estimate θ_2 by its MLE,

$$\hat{\theta}_{20} = \frac{1}{n_o} \sum \frac{(y_i - \hat{y}_i)^2}{y_i \, \hat{y}_i^2},$$

where n_o is the number of observed subjects, and $\hat{y}_i$ the estimated expected outcome for subject i.

References

Chhikara, R. S. and Folks, J. L. (1989). *The Inverse Gaussian Distribution: Theory, Methodology and Application*. Marcel Dekker, New York.

Cook, R. D. (1986). Assessment of local influence. *J. Roy. Statist. Soc. Ser. B* **48**, 133-169.

Copas, J. B. and Eguchi, S. (2001). Local sensitivity approximations for selectivity bias, *J. Roy. Statist. Soc. Ser. B* **63**, 871-895.

Copas, J. B. and Li, H. G. (1997). Inference for non-random samples (with discussion). *J. Roy. Statist. Soc. Ser. B* **59**, 55-95.

Daniels, M. J. and Hogan, J. W. (2000). Reparameterizing the pattern mixture model fo sensitivity analyses under informative dropout. *Biometrics* **65**, 1241-1248.

David, M., Little, R. J. A., Samuhel, M. E. and Triest, R. K. (1986). Alternative methods for CPS income imputation. *J. Amer. Statist. Assoc.* **81**, 29-41.

Diggle, P. and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis (with discussion). *Appl. Statist.* **43**, 49-73.

Heitjan, D. F. (1994). Ignorability in general incomplete-data models. *Biometrika* **81**, 701-708.
Heitjan, D. F. (1997). Ignorability, sufficiency and ancillarity. *J. Roy. Statist. Soc. Ser. B* **59**, 375-381.
Heitjan, D. F. and Basu, S. (1996). Distinguishing 'missing at random' and 'missing completely at random'. *Amer. Statist.* **50**, 207-213.
Heitjan, D. F. and Rubin, D. B. (1991). Ignorability and coarse data. *Ann. Statist.* **19**, 2244-2253.
Hogan, J. W. and Laird, N. M. (1997). Mixture models for the joint distribution of repeated measures and event times. *Statist. Medicine* **16**, 239-257.
Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. Second edition. Wiley, New York.
Little, R. J. A. and Wang, Y. (1996). Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics* **52**, 98-111.
McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall, New York.
Moore, D. S. and McCabe, G. P. (1989). *Introduction to the Practice of Statistics*. Freeman, New York.
Pulkstenis, E. P., Ten Have, T. R. and Landis, J. R. (1998). Model for the analysis of binary longitudinal pain data subject to informative dropout through remedication. *J. Amer. Statist. Assoc.* **93**, 438-450.
Raab, G. M., Burns, S. M., Scott, G., Cudmore, S., Ross, A., Gore, S. M., O'Brien, F. and Shaw, T. (1995). HIV prevalence and risk factors in university students. *AIDS* **9**, 191-197.
Raab, G. M. and Donnelly, C. A. (1999). Information on sexual behaviour when some data are missing. *Appl. Statist.* **48**, 117-133.
Rosenbaum, P. R. (1987). Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika* **74**, 13-26.
Rosenbaum, P. R. (1995). *Observational Studies*. Springer-Verlag, New York.
Rubin, D. B. (1976). Inference and Missing Data, *Biometrika* **63**, 581-592.
Scharfstein, D. O., Rotnitzky, A. and Robins, J. M. (1999). Adjusting for nonignorable drop-outs using semiparametric nonresponse models. *J. Amer. Statist. Assoc.* **94**, 1096-1120.
Schluchter, M. D. (1992). Methods for the analysis of informatively censored longitudinal data. *Statist. Medicine* **11**, 1861–1870.
Troxel, A. B. (1998). A comparative analysis of quality of life data from a Southwest Oncology Group randomized trial of advanced colorectal cancer. *Statist. Medicine* **17**, 767-779.
Troxel, A. B., Harrington, D. P. and Lipsitz, S. R. (1998). Analysis of longitudinal measurements with non-ignorable non-monotone missing values. *Appl. Statist.* **47**, 425-438.
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.

Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6021, U.S.A.

E-mail: atroxel@cceb.upenn.edu

Clinical Biostatistics, Merck & Co., Inc., BL3-2, Blue Bell, PA 19422, U.S.A.

E-mail: guoguang_ma@merck.com

Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6021, U.S.A.

E-mail: dheitjan@cceb.upenn.edu

(Received February 2002; accepted March 2003)