Impact of nonignorable coarsening on Bayesian inference

JIAMENG ZHANG*
Genentech, Inc., South San Francisco, CA 94080, USA
zhang.jiameng@gene.com

DANIEL F. HEITJAN
Department of Biostatistics and Epidemiology, University of Pennsylvania,
School of Medicine, Philadelphia, PA 19104, USA

SUMMARY
The coarse data model of Heitjan and Rubin (1991) generalizes the missing data model of Rubin (1976) to cover other forms of incompleteness such as censoring and grouping. The model has 2 components: an ideal data model describing the distribution of the quantity of interest and a coarsening mechanism that describes a distribution over degrees of coarsening given the ideal data. The coarsening mechanism is said to be nonignorable when the degree of coarsening depends on an incompletely observed ideal outcome, in which case failure to properly account for it can spoil inferences. A theme in recent research is to measure sensitivity to nonignorability by evaluating the effect of a small departure from ignorability on the maximum likelihood estimate (MLE) of a parameter of the ideal data model. One such construct is the “index of local sensitivity to nonignorability” (ISNI) (Troxel and others, 2004), which is the derivative of the MLE with respect to a nonignorability parameter evaluated at the ignorable model. In this paper, we adapt ISNI to Bayesian modeling by instead defining it as the derivative of the posterior expectation. We propose the application of ISNI as a first step in judging the robustness of a Bayesian analysis to nonignorable coarsening. We derive formulas for a range of models and apply the method to evaluate sensitivity to nonignorable coarsening in 2 real data examples, one involving missing CD4 counts in an HIV trial and the other involving potentially informatively censored relapse times in a leukemia trial.

Keywords: Censoring; Coarse data; Ignorability; Missingness; Sensitivity analysis.

1. INTRODUCTION
Heitjan and Rubin (1991) defined the data to be “coarse” if it is an imperfectly observed version of an underlying ideal data set. If the degree of coarseness depends on the true value of the ideal data, a condition known as “nonignorability,” standard models that ignore the coarsening mechanism can lead to incorrect inferences. Although estimating a correctly specified nonignorable model can yield valid...
inferences (Heitjan, 1993, 1994, 1997; Diggle and Kenward, 1994; Robins, 1997), the assumptions of such models are typically not directly verifiable from the data (Fisher and Kanarek, 1974; Tsiatis, 1975; Lagakos, 1979). Moreover, models that are sufficiently rich may be unidentifiable in the sense that there is no one-to-one mapping from the parameter space to the set of distributions of observables. A practical alternative approach is to perform a sensitivity analysis that evaluates the impact of nonignorability on key inferences. If the sensitivity is modest, we can trust the inferences from the simpler analysis that assumes ignorability.

An elementary sensitivity analysis that avoids the use of explicit nonignorable models would impute the coarsened observations by “extreme” values. For instance, a simple sensitivity analysis for censored data is to impute the censored survival time by either the maximum survival time or the censoring time (Allison, 1995). Large discrepancy between the inferences from different imputed data sets indicates high sensitivity to nonignorability—a common finding with this method, as it in effect assumes a high degree of nonignorability. Park and others (2006) introduced a nonparametric approach that enumerates the collection of all attainable values for the survival function using information about the causes of censoring. Their sensitivity analysis examines the change of the collection when one varies the classification of censoring as either dependent or independent. One may also use the difference of the likelihood or the posterior density between the ignorable and nonignorable models to represent the difference in key inferences. This difference can be quantified, for example, by the Kullback–Leibler divergence or any other relevant measure of distance. Cook (1986) used geometric normal curvatures to characterize the behavior of this likelihood displacement. Verbeke and Molenberghs (2001) and Jansen and others (2003) extended the method to continuous outcomes and multivariate and longitudinal binary data with possibly nonignorable missingness.

An alternative approach is to explicitly specify a coarsening model in which a single parameter indexes the magnitude of nonignorability and to quantify how the inference of interest varies as a function of this parameter. One variant is to evaluate the range of an estimate as the nonignorability parameter varies over its parameter space (Moeschberger and Klein, 1984; Zheng and Klein, 1995; Robins, 1997). Such a “global” sensitivity analysis can be computationally burdensome because it may require fitting the nonignorable model many times. Moreover, the bounds from such an analysis may be too wide to be of practical value if the limits of the nonignorability parameter represent values of nonignorability that are unlikely to be realized in practice. By contrast, “local” sensitivity analysis methods examine the sensitivity of parameter estimates to small perturbations from the ignorable model. Copas and Li (1997) first proposed this approach by measuring sensitivity of inferences on a normal mean to a probit selection mechanism. Siannis (2004) and Siannis and others (2005) applied a similar idea to measure sensitivity of the maximum likelihood estimate (MLE) of parameters in a survival model to small departures from noninformative censoring. Troxel and others (2004) derived a simple, general measure that they denoted the index of local sensitivity to nonignorability (ISNI), which they applied in the univariate generalized linear model. Other recent papers have applied ISNI analysis to causal modeling in a clinical trial with noncompliance (Xie and Heitjan, 2004), to longitudinal modeling (Ma and others, 2005), and to the coarse data model (Zhang, 2004; Zhang and Heitjan, 2005, 2006).

Past approaches to local sensitivity analysis have in most cases considered sensitivity of the MLE. As the MLE is equivalent to the posterior mode with a flat (and therefore potentially improper) prior, this approach may be unsatisfactory if the ultimate goal is to make Bayesian inferences, particularly when the sample size is small. Moreover, there is no straightforward non-Bayesian analogue of quantities like the probability that a parameter falls in some subset of the parameter space. Thus, in this paper, we propose a local sensitivity analysis for Bayesian inferences in the coarse data model.

The rest of the paper is organized as follows: Section 2 reviews the coarsening model. In Sections 3 and 4, we derive the general form of ISNI for Bayesian inference and present methods to evaluate the sensitivity. In Sections 5 and 6, we apply our method to real data sets involving missing and censored data.
To illustrate the range of applicability of the method, in the appendices we present ISNI formulas for a function of a parameter of interest (Appendix A), for a range of models with missing data (Appendix B), and for censored data with parametric models (Appendix C).

2. Nonignorable Coarsening

The coarsening model (Heitjan and Rubin, 1991) assumes that a random variable $X$, taking values in a sample space $\Omega_X$, is distributed according to the density $f_{\theta}^X(x)$, where $\theta$ is the parameter of interest. A second random variable $G$, taking values in a sample space $\Omega_G$, controls the degree of coarsening of $X$ in the sense that $G$ determines a partition of the sample space of $X$. The resulting data element is $Y = Y(X, G)$, a subset of $\Omega_X$ in which $X$ is known to lie. If we also do not observe $G$ precisely, then we denote $H = H(X, G)$ to be the range of plausible values of $G$, a subset of $\Omega_G$. We moreover assume that the distribution of $G$ given $X$ is $f_{\gamma|X}^G(g|x)$ and that $\theta$ and $\gamma$ are “distinct” in the sense that they are a priori independent. The correct likelihood is

$$L_C(\theta, \gamma; y, h) = \int_y f_{\theta}^X(x) \left\{ \int_h f_{\gamma|X}^G(g|x) f_{Y,H|X,G}(y, h|x, g) dg \right\} dx,$$

where $f_{Y,H|X,G}(y, h|x, g)$ is a known, degenerate function that equals 1 when $Y(x, g) = y$ and $H(x, g) = h$ and 0 otherwise.

Because the correct likelihood may be challenging to work with both analytically and numerically, in applications we typically favor the simplified likelihood

$$L_I(\theta; y) = \int_y f_{\theta}^X(x) dx.$$

With the same prior for $\theta$, this likelihood yields the same inferences as the correct likelihood if, for the fixed observed $(y, h)$ and for every $\gamma$, the quantity

$$\int_h f_{\gamma|X}^G(g|x) f_{Y,H|X,G}(y, h|x, g) dg$$

(2.1)

takes the same value for all $x \in y$. In such a case, we say that the data are “coarsened at random”; if, moreover, the parameters $\theta$ and $\gamma$ are distinct, in the sense of Rubin (1976), then the coarsening mechanism is “ignorable.”

Although the model stated above covers a broad range of data structures, it is helpful to have a notation that allows for $n$ independent units, possibly not identically distributed. Thus, we assume that $X$ and $G$ are $n$-vector random variables and that the components $X_i$ of $X$, $i = 1, \ldots, n$, are independent with density $p_{\theta}^{X_i}(x_i)$. The density of $X$ is then

$$f_{\theta}^X(x) = \prod_{i=1}^n p_{\theta}^{X_i}(x_i).$$

(2.2)

We also assume that the elements $G_i$ given $X_i = x_i$ are independent and distributed according to densities $p_{\gamma|X_i}^{G_i}(g_i|x_i)$, leading to the conditional density of $G$ given $X$ taking the form

$$f_{\gamma|X}^G(g|x) = \prod_{i=1}^n p_{\gamma|X_i}^{G_i}(g_i|x_i).$$

(2.3)
3. ISNI for Bayesian Inference

Bayesian data analysis reduces to the computation of posterior expectations of functions of key parameters of interest. Thus, our local sensitivity analysis is based on the derivative of a posterior mean with respect to a nonignorability parameter, evaluated at the ignorable model. We begin by further defining $\gamma = (\gamma_0, \gamma_1)'$, where $\gamma_1$ indicates the degree of nonignorability in the sense that if $\gamma_1 = 0$, $f_{\gamma_0,\gamma_1}(g|x) = f_{\gamma_0}(g)$ and the coarsening mechanism is ignorable. We assume that primary interest is in the posterior distribution of $\theta$. Let $\eta = (\theta, \gamma_0)'$, $E(\cdot)$ be the posterior expectation operation, and $\hat{\eta}(\gamma_1) = E(\eta|\gamma_1, y, h)$ be the posterior expectation of $(\theta, \gamma_0)$ when the nonignorability parameter is fixed at $\gamma_1$.

To assess the sensitivity of $\hat{\eta}(\gamma_1)$ to the nonignorable coarsening, we calculate the change of $\hat{\eta}(\gamma_1)$ when $\gamma_1$ is perturbed around the ignorable model. Define $\text{ISNI}[\hat{\eta}(\gamma_1)]$ as the rate of change of the posterior mean of $\eta$ evaluated at $\gamma_1 = 0$. We then have (see Appendix A)

$$
\text{ISNI}[\hat{\eta}(\gamma_1)] = \frac{\partial E(\eta|\gamma_1, y, h)}{\partial \gamma_1}_{\gamma_1=0} = \begin{bmatrix}
\text{COV}_I\left(\theta, \frac{\partial \ell_{C}(\theta, \gamma_0, y, h)}{\partial \gamma_1}|_{\gamma_1=0} \right)
\text{COV}_I\left(\gamma_0, \frac{\partial \ell_{C}(\theta, \gamma_0, y, h)}{\partial \gamma_1}|_{\gamma_1=0} \right)
\end{bmatrix},
$$

where $\ell_{C}(\theta, \gamma_0, y, h)$ is the log-likelihood and $\text{COV}_I(\cdot)$ is the posterior covariance under the ignorable model. Note that $\frac{\partial \ell_{C}(\theta, \gamma_0, y, h)}{\partial \gamma_1}|_{\gamma_1=0}$, the score statistic, is a function of $\theta$ and $\gamma_0$. ISNI describes the amount by which a unit change in $\gamma_1$ displaces the posterior expectation of the parameter of interest from its value under the ignorable model. Large absolute values of ISNI thus generally suggest high sensitivity to nonignorability. If no data are coarsened, then the coarsening mechanism is ignorable in the Bayesian sense, in that varying $\gamma_1$ can have no effect on inferences regarding $\theta$. In such situations, ISNI in general equals 0.

Under the ignorable model with distinct parameters, we can factor the posterior into 2 parts—one involving only $\theta$ and the other involving only $\gamma$. Therefore, $\text{COV}_I(\theta, \gamma_0) = 0$, and by the delta method,

$$
\text{ISNI}[\hat{\theta}(\gamma_1)] \approx \text{VAR}_I(\theta) \left( \frac{\partial^2 \ell_{C}(\theta, \gamma_1, y, h)}{\partial \gamma_1 \partial \theta} \right)_{\gamma_1=0, \hat{\theta}(0), \gamma_0(0)},
$$

where $\hat{\theta}(0)$ and $\gamma_0(0)$ are the posterior means of $\theta$ and $\gamma_0$ under the ignorable model and $\text{VAR}_I(\cdot)$ is the posterior variance under the ignorable model.

Note that this approximation is similar to ISNI for the MLE $\hat{\theta}$ (Troxel and others, 2004), whose formula replaces the posterior variance of $\theta$ with the inverse information under the ignorable model

$$
\text{ISNI}[\hat{\theta}(\gamma_1)] = \frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_1}_{\gamma_1=0} = \left[ \frac{\partial^2 \ell_{C}}{\partial \theta \partial \theta'} \right]^{-1} \left| \frac{\partial^2 \ell_{C}}{\partial \gamma_1 \partial \theta} \right|_{\gamma_1=0, \hat{\theta}(0), \gamma_0(0)},
$$

where $\gamma_0(0)$ and $\hat{\theta}(0)$ are the MLEs of $\gamma_0$ and $\theta$ from the ignorable model. In both cases, the first factor measures the precision of estimation of $\theta$ under the ignorable model and the second factor measures the nonorthogonality of $\theta$ and $\gamma_1$.

If the ignorable model actually holds and $\theta(0)$ is the true value of $\theta$, then under suitable regularity conditions, both $|\hat{\theta}(0) - \theta(0)|$ and $|\hat{\theta}(0) - \theta(0)|$ decrease to 0 at the rate $1/\sqrt{n}$ (Serfling, 2002; Walker, 1969; Bernardo and Smith, 1994). Because these 2 estimates converge to the truth at the same rate, as $n \to \infty$, $\frac{\text{ISNI}[\hat{\theta}(\gamma_1)]}{\text{ISNI}[\hat{\theta}(\gamma_1)]} \to 1$. 
4. Sensitivity evaluation by ISNI

We cannot in all cases use ISNI alone to index the sensitivity because its interpretation generally depends on the units of measurement. To address this problem (in the context of maximum likelihood estimation), Troxel \textit{and others} (2004) proposed to normalize ISNI by the standard error of the parameter of interest. If the higher derivatives of \( \theta(\gamma_1) \) with respect to \( \gamma_1 \) evaluated at \( \gamma_1 = 0 \) are small, we can approximate \( \hat{\theta}(\gamma_1) \) by \( \hat{\theta}(\gamma_1) \approx \hat{\theta}(0) + \gamma_1 \times \text{ISNI}(\hat{\theta}) \). We define the posterior mean to be sensitive to nonignorability if, for a plausible value of \( \gamma_1 \), the difference between the posterior means of the parameters under the ignorable and nonignorable models exceeds one posterior standard deviation (SD) under the ignorable model.

Here, we adapt the graphical sensitivity evaluation method of Zhang (2004) to the Bayesian case. The idea is to set \( \gamma_1^* = \frac{\text{SD}_I(\theta)}{\text{ISNI}(\theta)} \), where \( \text{SD}_I(\theta) \) is the posterior SD of \( \theta \) under the ignorable model; \( \gamma_1^* \) is then approximately the value of \( \gamma_1 \) that can cause a one-SD change in \( \hat{\theta} \). If \( \gamma_1^* \) is plausible for the data, in the sense that it is in line with prior notions of how bad the nonignorability can be, then we say that \( \hat{\theta} \) is sensitive to the nonignorable coarsening. One can check the plausibility of \( \gamma_1^* \) by plotting \( q(\gamma_1^*, x^*) \) versus \( x^* \), where \( x^* \) ranges over the set of all plausible values of \( x \) and \( q(\cdot, \cdot) \) is a measure whose plausibility is easy to interpret, such as the conditional expectation or hazard function of \( G \) given \( X = x \). Note that because \( \gamma_1 = 0 \) implies a nonignorable coarsening mechanism, \( q(\gamma_1^*, x^*) \) should be constant in the \( x^* \) argument. The greater the departure of \( \gamma_1^* \) from 0, the larger will be the variation of \( q(\gamma_1^*, x^*) \). If the range of \( q(\gamma_1^*, x^*) \) is small enough to be considered plausible for the data at hand, then \( \gamma_1^* \) is plausible for the data, and we conclude that inferences are sensitive. If on the other hand the range of \( q(\gamma_1^*, x^*) \) is larger than seems plausible, then inferences are insensitive and the ignorable model can be considered reliable. Thus, \( \gamma_1^* \) is similar to the minimum value of the odds ratio for hidden bias \( \Gamma \) that appears in the work of Rosenbaum (1995) on observational studies.

Likely choices of \( q(\cdot, \cdot) \) may involve \( \gamma_0 \) as well as \( \gamma_1 \). For example, in the missing data model, one may set \( q(\gamma_1^*, x^*) = \Pr_{\gamma_0}(X_j \text{ is observed}|X_j = x^*) \), in which case \( q(\cdot, \cdot) \) implicitly depends on \( \gamma_0 \). In such cases, we may propagate uncertainty properly by evaluating not the limits of \( q(\gamma_1^*, x^*) \) itself but rather the posterior probability that \( q(\gamma_1^*, x^*) \) exceeds some benchmark value. We might then evaluate the posterior probabilities

\[
P_1 = \Pr \left[ \Pr_{\gamma_0}(X_j \text{ is observed}|\gamma_1^*, X_i = \text{max}(x)) < 0.1 \right]
\]

and

\[
P_2 = \Pr \left[ \Pr_{\gamma_0}(X_j \text{ is observed}|\gamma_1^*, X_i = \text{min}(x)) > 0.9 \right],
\]

where the outer probability operator refers to the posterior distribution of \( \gamma_0 \) assuming that \( \gamma_1 = 0 \). If both \( P_1 \) and \( P_2 \) are close to 1, the suggestion is that under \( \gamma_1^* \) there is a wide range of probabilities of missingness, and therefore, we conclude that \( \gamma_1^* \) is implausible and that the inference is insensitive to nonignorability. We will use this approach in what follows.

The values of ISNI depend on the specification of the joint distribution of \( X \) and \( G \), both the distribution of \( X \) and the conditional distribution of \( G|X \) (the coarsening mechanism). In our experience, the conclusions of ISNI-based local sensitivity analyses have been robust to the choice of parametric model for the coarsening mechanism, as long as it allows a sufficient range of nonignorability.

5. ISNI for Missing Data

5.1 General formulation

Let \( X_i, i = 1, \ldots, n \), be independent outcomes with joint density given by (2.2). Let \( G_i \) be an accompanying indicator of whether \( X_i \) is observed, and assume that the distribution of \( G \) given \( X = x \) is given by
(2.3). Assume specifically that the conditional distribution of $G_i$ given $X_i = x_i$ and $\gamma$ is Bernoulli:

$$P_{G_i|X_i}(g_i|x_i) = [h(\gamma_0 + \gamma_1 x_i)]^{g_i} [1 - h(\gamma_0 + \gamma_1 x_i)]^{(1-g_i)},$$

where $h(\cdot)$ is a monotone increasing link function. If $\gamma_1 < 0$ ($>0$), there is a risk of overestimating (underestimating) the mean outcome.

Research to date suggests that MLEs from nonignorable models can be sensitive to the complete data model but are robust to the selection model (Kenward, 1998; Xie, 2003). Thus, the choice of link may have a modest effect on the sensitivity analysis, and we henceforth assume $h(\cdot)$ to be the commonly used logistic link,

$$h(u) = \expit(u) = \exp(u)/(1 + \exp(u)).$$

It is convenient to express the prior of $\gamma_0$ as a beta $(v, w)$ density on the $h(\gamma_0)$ scale,

$$\pi(h(\gamma_0)) \propto (h(\gamma_0))^{v-1} (1 - h(\gamma_0))^{w-1},$$

which allows us to think of $v$ and $w$ as prior numbers of observed and missing outcomes, respectively. Then, we have the general result that

$$\text{ISNI}(\tilde{\theta}) = -\left(\frac{\sum_{i=1}^{n} g_i}{n + v + w}\right) + \frac{v}{\sum_{i=1}^{n} (1 - g_i) \text{COV}_i(\theta, E_0[X_i])}.$$

Note that, as with ISNI for the MLE, this expression equals 0, and consequently, sensitivity is 0 if all units are observed.

### 5.2 Application to simulated normal data

We illustrate the analysis in the special case of estimation of a normal mean $\mu$; formulas for some other distributions appear in Appendix B. Let $\theta = (\theta_1, \theta_2)' = (\mu, \sigma^2)'$. A closed-form expression for ISNI for the posterior mean $\tilde{\mu}$ depends on the prior distribution of $(\mu, \sigma^2)$. We consider three cases:

#### Case 1. A noninformative prior. We assume that $\mu$ and log $\sigma$ are uniform and independent: $\pi(\mu, \sigma^2) \propto 1/\sigma^2$. Then, $\bar{\mu}_0 = \bar{y} = \frac{1}{n_0} \sum_{i=1}^{n} g_i y_i$ and

$$\text{ISNI}(\tilde{\mu}) = -\frac{(n_0 + v)(n - n_0)(n_0 - 1)s^2}{(n + v + w)(n_0 - 3)n_0},$$

where $n_0 = \sum_{i=1}^{n} g_i$ and $s^2 = \frac{1}{n_0 - 1} \sum_{i=1}^{n} g_i (y_i - \bar{y})^2$.

#### Case 2. A semi-conjugate prior. Assuming that $\sigma^2$ is known and $\mu \sim N(\mu_0, \tau_0^2)$, we have that $\bar{\mu}_0 = \frac{\mu_0 \tau_0^2 + n_0 \tilde{\mu}/\sigma^2}{1/\tau_0^2 + n_0/\sigma^2}$ and

$$\text{ISNI}(\tilde{\mu}) = -\frac{(n_0 + v)(n - n_0)}{(n + v + w)(1/\tau_0^2 + n_0/\sigma^2)}.$$

#### Case 3. A conjugate prior. We assume that $\mu|\sigma^2 \sim N(\mu_0, \sigma^2/k_0)$ and $\sigma^2 \sim \text{Inv-}\chi^2(v_0, \sigma_0^2)$. Then, we have $\bar{\mu}_0 = \frac{k_0}{k_0 + n_0} \mu_0 + \frac{n_0}{k_0 + n_0} \tilde{\mu}$ with

$$\text{ISNI}(\tilde{\mu}) = -\frac{(n_0 + v)(n - n_0)n_0(v_0\sigma_0^2 + (n_0 - 1)s^2 + \frac{k_0n_0}{k_0 + n_0} (\tilde{\mu} - \mu_0))}{(n + v + w)(n_0 - 2)(k_0 + n_0)(v_0 + n_o)}.$$
Fig. 1. Simulated normal data with nonignorable missingness. Missing subjects are labeled with a +.

To illustrate the graphical sensitivity analysis of Section 4, we generated 8 samples containing 4000 \((X_i, G_i)\) pairs, where the \(X_i\) values for each sample were the same and were generated from a standard normal. Here, \(G_i\) given \(X_i = x_i\) are independent binomials with 
\[
\Pr[G_i = 1 | X_i = x_i] = \expit(\gamma_0 + \gamma_1 x_i).
\]
We set the values of \(\gamma_0 = 3\) and \(\gamma_1 = \pm 3, \pm 5, \pm 8, \pm 10, \pm 20\) for these simulated data. In Figure 1, we randomly chose and plotted 100 pairs from 2 samples with \(\gamma_1 = 3\) and \(-3\), respectively. For \(\gamma_1 = 3\) (\(-3\)), the missingness is heavier for subjects with smaller (larger) outcomes.

We analyzed the missing data by the ignorable and nonignorable models assuming a noninformative prior, with results summarized in Table 1. We used importance sampling to calculate the posterior means. The posterior mean from the ignorable model (\(\tilde{\mu}_0\)) was an overestimate (underestimate) when \(\gamma_1 > 0\) (<0). The posterior mean from the nonignorable model (\(\tilde{\mu}_{nong}\)) was close to the true value 0. We calculated \(\gamma_1^*\) as defined in Section 4. Taking the range of plausible values of \(q(\cdot, \cdot)\) to be (0.1, 0.9), we defined \(P_1\) and \(P_2\) to be the posterior probabilities, averaging over the distribution of the \(\eta\) parameters, that the probability of being observed is either less than 0.1 or greater than 0.9, for fixed \(\gamma_1^*\) and for the maximum or minimum values of \(x\). If both \(P_1\) and \(P_2\) are near 1, indicating that subjects with extreme values of \(x\) either all give complete data or all are missing, then \(\gamma_1^*\) would be implausible for the data.

In Figure 2, we plot the posterior means assuming different values for the ignorability parameter \(\gamma_1\) for the data with true \(\gamma_1 = \pm 3\). ISNI(\(\tilde{\mu}\)) is the slope of \(\tilde{\mu}\) at \(\gamma_1 = 0\), which reflects the behavior of \(\tilde{\mu}\) near the ignorable model. Evidently, the linear approximation is reasonable in the range \(\gamma_1 \in (-3, 3)\) for these data, and consequently, the value of \(\gamma_1^*\) accurately estimates the degree of nonignorability that is sufficient to cause one-SD change in \(\mu\).
Table 1. Posterior means and ISNIs from the simulated missing data ($n = 4000$)

<table>
<thead>
<tr>
<th>$\mu = 0$</th>
<th>Missing %</th>
<th>$\bar{\mu}_{\text{nonig}}$ (SD)</th>
<th>$\bar{\mu}_0$ (SD)</th>
<th>ISNI($\bar{\mu}$)</th>
<th>$\gamma_1^*$</th>
<th>$P_1$</th>
<th>$P_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_1 &gt; 0$</td>
<td>38.5</td>
<td>0.012 (0.032)</td>
<td>0.527 (0.015)</td>
<td>-0.219</td>
<td>0.069</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>19.3</td>
<td>0.018 (0.020)</td>
<td>0.297 (0.014)</td>
<td>-0.128</td>
<td>0.110</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>8.3</td>
<td>0.021 (0.019)</td>
<td>0.143 (0.015)</td>
<td>-0.065</td>
<td>0.231</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>0.020 (0.017)</td>
<td>0.053 (0.015)</td>
<td>-0.022</td>
<td>0.682</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.013 (0.018)</td>
<td>0.018 (0.016)</td>
<td>-0.007</td>
<td>2.286</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\gamma_1 &lt; 0$</td>
<td>38.9</td>
<td>0.034 (0.029)</td>
<td>-0.537 (0.015)</td>
<td>-0.226</td>
<td>-0.066</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>19.4</td>
<td>0.001 (0.020)</td>
<td>-0.285 (0.015)</td>
<td>-0.137</td>
<td>-0.110</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>6.9</td>
<td>0.000 (0.017)</td>
<td>-0.115 (0.015)</td>
<td>-0.058</td>
<td>-0.259</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2.2</td>
<td>-0.006 (0.017)</td>
<td>-0.043 (0.015)</td>
<td>-0.021</td>
<td>-0.714</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.003 (0.0014)</td>
<td>-0.012 (0.016)</td>
<td>-0.005</td>
<td>-3.200</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Fig. 2. The posterior expectation of $\mu$ assuming different $\gamma_1$.

The probability of being observed given the outcome value $x$ and fixed values of $\gamma_0$ and $\gamma_1$ is \text{expit}(\gamma_0 + \gamma_1 x). One can check whether $\gamma_1^*$ is plausible for the data by plotting \text{expit}(\gamma_0^* + \gamma_1^* x^*) versus $x^*$, where $x^*$ ranges over the set of all plausible values of the outcome and $\gamma_0^*$ is the posterior mean of $\gamma_0$ from the ignorable model. If the range of probabilities is plausible for the data, $\gamma_1^*$ is also plausible for the data and $\mu$ is sensitive to nonignorable missingness. We present these plots in Figure 3. The ranges of the probability in the last plot in each row are much larger than the ranges in other plots, showing relatively...
Fig. 3. The probability of being observed given $x$ assuming $\gamma_\star = \gamma_1^\star$ that can cause one-SD change in the posterior expectation of the mean of $X$ (simulated data). The solid lines represent the probability of being observed assuming an ignorable censoring mechanism. The dotted lines assume $\gamma_1^\star > 0$. The dashed lines assume $\gamma_1^\star < 0$.

low sensitivity to the dependent censoring. As shown in the 5th and 10th rows of Table 1, the posterior probabilities $P_1$ and $P_2$ were both 1, indicating low sensitivity; the differences of posterior expectations from the ignorable and nonignorable models for these data sets were close to or less than one SD, whereas for the other data sets, the differences were larger. The plots and posterior probabilities show that the sensitivity to nonignorability increases with the percentage of censoring.

5.3 Sensitivity of mean CD4 count in an HIV trial

Adult AIDS Clinical Trials Group (AACTG) study 5125 sought to compare the impact of protease inhibitors-containing lopinavir/ritonavir (LPV/r) and nucleoside reverse transcriptase inhibitors (NRTI) containing regimens on HIV lipoatrophy. Subjects received either LPV/r or NRTI and were followed for 72 weeks beyond the enrollment of the last subject. CD4 count at week 72 was a key safety end point but was missing for 10 of the 61 subjects (6 on NRTI and 4 on LPV/r). There was a concern that the data were missing nonignorably because poor health may be associated with both low CD4 counts and a propensity to miss scheduled visits. We analyzed the data to determine whether the difference between the arms in mean 72-week CD4 count was sensitive to nonignorability.
Impact of nonignorable coarsening on Bayesian inference

Fig. 4. The curved lines are the probability of having a CD4 count measured at week 72 for subjects with outcome \( x \), assuming a value \( \gamma_1 = \gamma_1^* \) that can cause a one-SD change in the posterior estimates (AACTG 5125 data). The solid line is the probability of having a CD4 count available at week 72 assuming an ignorable missingness mechanism.

We assume that CD4 count \( (X_i) \) is independent and normally distributed with mean \( z_i \beta \) and variance \( \sigma^2 \), where \( z_i = (1, 1) \) for subject \( i \) in the NRTI arm, \( z_i = (1, 0) \) for subject \( i \) in the LPV/r arm, and \( \beta = (\beta_0, \beta_1) \). With the noninformative prior \( \pi(\beta, \sigma^2 | Z) \propto \frac{1}{\sigma^2} \), we have the ignorable model posterior

\[
\beta | y \sim t_{n_o-p}(\tilde{\beta}, s^2(Z_o'Z_o)^{-1}),
\]

where \( Z_o \) is the subset of \( Z \) corresponding to subjects for which the outcome is observed, \( \tilde{\beta} \) is the least-squares estimate of \( \beta \), and \( s^2 \) is the residual variance. We measure the sensitivity with

\[
\text{ISNI}(\tilde{\beta}) = -\frac{(n_o + v)(n_o - 2)}{(n + v + w)(n_o - 4)} s^2(Z_o'Z_o)^{-1}\sum_{i=1}^{n} (1 - g_i)z_i'.
\]

To check the sensitivity of \( \tilde{\beta}_1 \) without fitting the nonignorable model, we set \( \gamma_1^* = 0.0122 \) as defined in Section 4 and \( \gamma_0^* \) to be the posterior mean of \( \gamma_0 \) under the ignorable model. Figure 4 shows plots of the probability of observing a CD4 outcome given \( \gamma_0^* \) and \( \gamma_1^* \) versus the set of all plausible values of the outcome. The range of probabilities is large (from 0 to 1), which suggests that we would not have sensitivity unless the missingness mechanism is such that subjects with CD4 less than 100 would always miss visits and subjects with CD4 greater than 1000 would always attend visits. This is not plausible, in that subjects with evidently low CD4 do sometimes appear for visits and subjects with evidently high CD4 do sometimes skip visits. We also calculated the posterior probability, under the ignorable model, that the probability of staying in the trial is less than 0.1 (greater than 0.9) at max(\( x \)) (min(\( x \))). Both probabilities are equal to 1, in agreement with Figure 4. Therefore, we conclude that the posterior mean of the treatment effect on day-72 CD4 is robust to all but extreme nonignorability.

We are also interested in estimating the probability that the mean day-72 CD4 is less than 500 (the lower bound of the normal range) for the 2 treatment arms separately. Table 2 shows the inferences from the ignorable model along with the ISNIs and corresponding \( \gamma_1^* \). We check the plausibility of the \( \gamma_1^* \) values in Figure 4. The ranges of the probability of not dropping out given \( \gamma_1^* \) seem plausible for the trial. Therefore, we conclude that the probabilities that the mean CD4 is below normal for the 2 arms are
Table 2. Posterior probabilities and ISNIs from the ignorable model (AACTG 5125 data)

<table>
<thead>
<tr>
<th>Arm</th>
<th>Pr_I (zi β &lt; 500) (SD)</th>
<th>Missed/observed</th>
<th>ISNI(Pr_I (zi β &lt; 500))</th>
<th>γ^*</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRTI</td>
<td>0.24 (0.09)</td>
<td>7/31</td>
<td>56.94</td>
<td>0.0015</td>
</tr>
<tr>
<td>LPV/r</td>
<td>0.61 (0.09)</td>
<td>3/31</td>
<td>27.34</td>
<td>0.0034</td>
</tr>
</tbody>
</table>

sensitive to possible non ignorable dropout. The probability that the mean CD4 is below normal for the NRTI arm is more sensitive than that for the LPV/r arm, probably because there are more missing data in the NRTI arm.

This analysis demonstrates how the ISNI approach can produce a separate sensitivity analysis for each model parameter of interest—indeed, for each function of interest of the parameters. This property is important because it may happen that the 2 parameters representing the 2 study arms separately are both substantially sensitive to nonignorability, whereas their difference (i.e. the treatment effect) is not.

6. ISNI FOR CENSORED DATA

Let \( X \) be an \( n \)-vector of times to the onset of a particular event and \( G \) be an accompanying \( n \)-vector of censoring times from another possible event. Assume that the components of \( X \) are independent with joint density given by (2.2) and that the components of \( G \) given \( X = x \) are given by (2.3); specifically, we assume a proportional hazards model for the hazard function of the censoring variable:

\[
h_γ (g_i | x_i) = h_0(g_i)e^{γ_0 + γ_1x_i},
\]

where \( h_0(·) \) is the baseline censoring hazard. Taking \( H_0(·) \) to be the cumulative baseline censoring hazard, the full likelihood is

\[
L_C(θ, γ_0, γ_1 ; y, h) = \prod_{i=1}^{n} \{ p_{θ}^{X_i} (x_i) \exp[-H_0(x_i)e^{γ_0 + γ_1x_i}] \}^{d_i} \cdot \int_{g_i}^{∞} p_{θ}^{X_i} (u) \exp[-H_0(g_i)e^{γ_0 + γ_1u}]h_0(g_i)e^{γ_0 + γ_1u} \ du \right)^{(1-d_i)},
\]

where \( d_i = 1 \) if \( x_i \leq g_i \) and 0 otherwise.

Under this model, a negative (positive) value of \( γ_1 \) implies that the posterior mean of survival under the ignorable model will be too large (small). In general, ISNI for the posterior mean of \( θ \) is

\[
\text{ISNI}(\tilde{θ}) = \text{COV}_I \left( θ, \sum_{i=1}^{n} (1 - d_i)E_θ[X_i | X_i > g_i][1 - e^{γ_0}H_0(g_i)] \right).
\]

Again, note that if no subjects are censored (all \( d_i = 1 \)), then ISNI is 0, and every inference about \( θ \) is completely insensitive to nonignorability.

6.1 Application to exponential survival data

Assume that the marginal distribution of \( X_i \) is exponential with mean \( e^{-θ} \) and that the conditional distribution of \( G_i \) given \( X_i = x_i \) is exponential with mean \( e^{-θ-γ_1x_i} \). (Formulas for censored data in other
Impact of nonignorable coarsening on Bayesian inference

733

models appear in Appendix C.) ISNI for the posterior mean survival \( e^{-\theta} \) in is general

\[
\text{ISNI}(e^{-\theta}) = \sum_{i=1}^{n} (1 - d_i) [1 - E(e^{\gamma_0}g_i)] \text{VAR}_I(e^{-\theta}).
\]

We consider 2 special cases:

**Case 1.** Under the Jeffreys prior for \( e^\theta \) and \( e^{\gamma_0} \), \( \pi(e^\theta, e^{\gamma_0}) = \frac{1}{e^{e^\theta}} \), we have that

\[
\text{ISNI}(e^{-\theta}) = \left[ \sum_{i=1}^{n} (1 - d_i) \left( 1 - \frac{\sum_i (1 - d_i)}{\sum_i t_i} g_i \right) \right] \cdot \frac{(\sum_i t_i)^2}{(\sum_i d_i - 1)^2(\sum_i d_i - 2)},
\]

where \( t_i = \min(x_i, g_i) \).

**Case 2.** Conjugate priors for \( e^\theta \) and \( e^{\gamma_0} \) take the prior for the censoring rate to be \( \Gamma(k, f) \) and the prior for the survival rate to be \( \Gamma(p, q) \). We can think of \( p \) and \( k \) as prior numbers of events and censorings, in prior lengths of follow-up \( q \) and \( f \), respectively, \( \pi(e^{\gamma_0}, e^\theta) \propto (e^{\gamma_0})^{k-1} \exp(-f e^{\gamma_0})(e^\theta)^{p-1} \exp(-q e^\theta) \). Then, ISNI for the posterior mean survival \( e^{-\theta} \) is

\[
\text{ISNI}(e^{-\theta}) = \left[ \sum_{i=1}^{n} (1 - d_i) \left( 1 - \frac{k + \sum_i (1 - d_i)}{k + \sum_i t_i} g_i \right) \right] \cdot \frac{(q + \sum_i t_i)^2}{(p + \sum_i d_i - 1)^2(p + \sum_i d_i - 2)},
\]

where \( t_i = \min(x_i, g_i) \).

**6.2 Sensitivity of mean time to relapse in a leukemia trial**

A bone marrow transplant (BMT) for leukemia is considered a failure if the patient either relapses or dies while remission. Oncologists postulated that the development of effective techniques for treating graft-versus-host disease and other complications of BMT would considerably reduce the incidence of death in remission. In this context, death in remission censors relapse, and because times to relapse and death in remission are possibly correlated, the censoring is potentially nonignorable and a sensitivity analysis is in order. We illustrate the procedure using data from a multicenter trial of BMT (Copeland and others, 1991). In up to 7 years of follow-up, 42 of 137 patients relapsed, 41 died in remission, and the rest survived without relapse. We assume that interest is in the mean time to relapse.

Let \( X_i \) be the relapse time, \( G_{1i} \) be the time to death in relapse, and \( G_{2i} \) be the maximum time of observation. We assume that \( X_i \) is exponential with mean \( e^{-\theta} \), \( G_{1i} \) given \( X_i = x_i \) is exponential with mean \( e^{-\gamma_0 - \gamma_1 x_i} \), and \( G_{2i} \) is uniform and independent of both \( X_i \) and \( G_{1i} \). The correct likelihood is

\[
L_C(\theta, \gamma_0, \gamma_1; y, h) = \prod_{i=1}^{n} \left[ e^{\theta} \exp(-e^\theta x_i) \exp(-e^{\gamma_0 + \gamma_1 x_i}) \right]^{d_{1i}(1-d_{2i})} \cdot \int_{g_{1i}}^{\infty} e^{\theta} \exp(-e^\theta u) e^{\gamma_0 + \gamma_1 u} \exp(-e^{\gamma_0 + \gamma_1 u} g_{1i}) du \right]^{(1-d_{1i})d_{2i}} \cdot \int_{g_{2i}}^{\infty} e^{\theta} \exp(-e^\theta u) \exp(-e^{\gamma_0 + \gamma_1 u} g_{2i}) du \right]^{(1-d_{1i})(1-d_{2i})},
\]
Table 3. Sensitivity analysis of posterior mean relapse time

<table>
<thead>
<tr>
<th>Model</th>
<th>$\tilde{\mu}$</th>
<th>SD($\mu$)</th>
<th>$\sim\psi$</th>
<th>SD($\gamma_1$)</th>
<th>ISNI($\tilde{\mu}$)</th>
<th>Bayes factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonignorable</td>
<td>1854.5</td>
<td>256.7</td>
<td>-0.0011</td>
<td>0.00026</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Ignorable</td>
<td>2550.9 ($\tilde{\mu}_0$)</td>
<td>398.4(SD($\mu_0$))</td>
<td>—</td>
<td>—</td>
<td>456114.4</td>
<td>3.69</td>
</tr>
</tbody>
</table>

Fig. 5. The dashed line is the hazard ratio of censoring (i.e. death in remission) for subjects with relapse time 0 compared to subjects with relapse time $x$, assuming a value $\gamma_1 = \gamma_1^*$ that can cause a change equal to one posterior SD in mean survival. The solid line is the hazard ratio under the ignorable model.

where

\[
d_{i1} = \begin{cases} 
1, & \text{if } X_i \leq G_{1i} \text{ and } X_i \leq G_{2i}, \\
0, & \text{otherwise},
\end{cases}
\]

\[
d_{2i} = \begin{cases} 
1, & \text{if } X_i > G_{1i} \text{ and } G_{2i} > G_{1i}, \\
0, & \text{otherwise}.
\end{cases}
\]

With this nonignorable model, it is possible (although not trivial) to calculate the posterior means of all the model parameters. Table 3 shows the posterior inferences from ignorable and nonignorable models. The Bayes factor is larger than 3, implying substantial evidence against $H_0: \gamma_1 = 0$. Because the posterior mean of $\tilde{\gamma}_1$ is negative, posterior mean relapse time from the ignorable model is evidently inflated. Moreover, the difference of posterior expectations from the ignorable and nonignorable models exceeds one SD, suggesting high sensitivity to nonignorability.

We calculated the ISNI for mean relapse time assuming a Jeffreys prior for $(e^\theta, e^{\gamma_0})$. To assess the sensitivity of posterior mean relapse time ($\tilde{\mu}$) to nonignorability without fitting the nonignorable model, we set $\gamma_1^* = \text{SD}(\mu_0)/\text{ISNI}(\tilde{\mu}) = 398.4/456114.4 = 0.00087$ and plotted ratios of the hazard of being censored by death in remission, $\exp(\gamma_1^* x^*)$ versus $x^*$, where $x^*$ ranges over the set of all plausible values of time to relapse (Figure 5). Because the ranges of plotted hazard ratios are plausible for the data, $\gamma_1^*$ is a plausible degree of nonignorability, and we conclude that posterior mean relapse time is sensitive to even this relatively modest amount of dependent censoring. In Table 4, we present the posterior mean of the censoring hazard ($\exp(\gamma_0 + \gamma_1^* x^*)$) and the posterior probability of the censoring hazard under the nonignorable model being less than the hazard under the ignorable model ($\exp(\gamma_0 + \gamma_1^* x^*) < \exp(\gamma_0)$), given $\gamma_1^*$, using the posterior distribution of $\gamma_0$ from the ignorable model and setting $x$ to the extreme
observed values of the relapse time. Because the ranges of the posterior means and probabilities are not implausible for the application, we again conclude sensitivity.

7. Discussion

When data are possibly nonignorably coarsened, a formal assessment of sensitivity can be an important first step in their analysis. We have presented here a simple method for conducting such analyses for moment-based Bayesian inferences under the coarse data model. The method has the advantages of accommodating any type of parametric model, avoiding the estimation of a nonignorable model, and allowing a separate sensitivity analysis for each model parameter or function of model parameters. We have presented formulas for ISNI under simple parametric models for common coarsening scenarios. Our exposition is intended to be illustrative rather than comprehensive, in that we conjecture that it is straightforward to derive analogous formulas for more complicated failure and coarsening models.

When an analytical posterior is not available, a straightforward numerical approach would calculate ISNI from repeated fits of the model at a range of values of the nonignorability parameter $\gamma_1$. Alternatively, if using a Markov chain Monte Carlo (MCMC) estimation algorithm, one could fit the full nonignorable model and take the necessary derivative numerically by evaluating $E(\eta|\gamma_1, y, h)$ at a range of convenient $\gamma_1$ values in the sample path. We do not counsel the latter strategy, however, as the posterior from the full nonignorable model may behave badly and be difficult to explore. Moreover, there is no guarantee that MCMC will sojourn long enough in the vicinity of $\gamma_1 = 0$ to impart sufficient accuracy to the computation of ISNI.

A potential criticism of the method is its dependence on a parametric model for the coarsening mechanism. In all cases that we are aware of, the nonignorable model is at best poorly identified from the data because we never know the precise values of the coarsened observations that would allow us to robustly estimate such a model. Fortunately, results obtained so far in the missing data context suggest that the sensitivity analysis itself is reasonably robust to assumptions about the missing data selection model (Xie, 2003). Thus, there is an empirical support for the conjecture that the choice of coarsening model has minimal influence on evaluations of sensitivity, at least locally. The validity of assumptions regarding the complete data model is equally a concern, but again our experience suggests that the degree of sensitivity to nonignorability is robust to the complete data model—which, if generally true, would simplify matters considerably (Xie, 2003; Zhang and Heitjan, 2006).

The sensitivity measures derived here are similar to those previously derived for MLEs in Troxel and others (2004) and Zhang (2004)—unsurprisingly, as it is well known that Bayesian and frequentist data summaries differ only slightly in large samples (e.g. Kass and others, 1989). If the objective is to execute Bayesian inference, the Bayesian ISNI approach has the advantage that one can readily apply it to other posterior moments such as the probability that a parameter $\theta$ lies below a constant $c$, that $\theta_1 < \theta_2$, and so on (see Appendix A). The MLE version of ISNI may also be inaccurate as a substitute for Bayesian ISNI in small-sample situations where the posterior mean and MLE have not converged.
One can generalize the method to cover more complicated coarsening models such as those involving multiple nonignorability parameters—for example, a separate nonignorability parameter for each treatment group in a randomized clinical trial. Xie (2003) and Xie and Heitjan (2004) have described a method that identifies the direction of maximal local sensitivity to nonignorability. Approaches that exploit prior information on the nonignorability parameters are also possible. We note moreover that although the analysis presented here relies on parametric models, in principle the same reasoning is applicable to nonparametric or semiparametric models such as those underlying the Kaplan–Meier curve and Cox regression.

When an ISNI-based analysis suggests modest sensitivity, one can use the ignorable model with some assurance of robustness to minor departures from ignorability. When the analysis suggests more pronounced sensitivity, one has a range of options: First, if there is sufficient prior information on the coarsening mechanism, one can fit the nonignorable model. If not, then one can either conduct global sensitivity analysis to explore the dependence of inferences on $\gamma_1$ more generally or decline to conduct further inferential analyses on the grounds that model-based inferences are evidently sensitive to things unknown. In the survival example, we were able to estimate the nonignorable model without heroic efforts, but we are not suggesting that the results are therefore reliable. If one has recourse to nonignorable modeling, it seems prudent to consider sensitivity of the results to a range of potential nonignorable models, relying on the findings only to the extent that reasonable models agree.

**ACKNOWLEDGMENTS**

We acknowledge with thanks the contributions of the reviewers. The United States Public Health Service supported this research under grant HL68074 and cooperative agreements AI038855 and NS032228. Conflict of Interest: None declared.

**APPENDIX A**

A.1 General

Letting $\pi(\eta)$ be the prior of $\eta$, the posterior mean of $\eta$ for a fixed $\gamma_1$ is

$$\tilde{\eta}(\gamma_1) = E(\eta|\gamma_1, y, h) = \frac{\int \eta L_C(\eta; \gamma_1, y, h) \pi(\eta) d\eta}{\int L_C(\eta; \gamma_1, y, h) \pi(\eta) d\eta},$$

where $L_C(\eta; \gamma_1, y, h)$ is the likelihood accounting for the coarsening. Define

$$\text{ISNI}(\tilde{\eta}(\gamma_1)) = \begin{bmatrix} \text{ISNI}(\tilde{\theta}(\gamma_1)) \\ \text{ISNI}(\tilde{\gamma}_0(\gamma_1)) \end{bmatrix} = \begin{bmatrix} \frac{\partial E(\eta|\gamma_1, y, h)}{\partial \gamma_1} \\ \frac{\partial E(\eta|\gamma_1, y, h)}{\partial \gamma_1} \end{bmatrix} = \begin{bmatrix} \int \eta \frac{\partial L_C(\eta; \gamma_1, y, h)}{\partial \gamma_1} |_{\gamma_1=0} L_I(\eta; y, h) \pi(\eta) d\eta \\ \int L_I(\eta; y, h) \pi(\eta) d\eta \end{bmatrix} \begin{bmatrix} \int \eta L_I(\eta; y, h) \pi(\eta) d\eta \\ \int L_I(\eta; y, h) \pi(\eta) d\eta \end{bmatrix}^2$$

$$= E_I \left( \eta, \frac{\partial L_C(\eta; \gamma_1, y, h)}{\partial \gamma_1} \bigg|_{\gamma_1=0} \right) - E_I(\eta)E_I \left( \frac{\partial L_C(\eta; \gamma_1, y, h)}{\partial \gamma_1} \bigg|_{\gamma_1=0} \right)$$

$$= \begin{bmatrix} \text{COV}_I \left( \theta, \frac{\partial L_C(\theta, \gamma_1, y, h)}{\partial \gamma_1} \bigg|_{\gamma_1=0} \right) \\ \text{COV}_I \left( \gamma_0, \frac{\partial L_C(\theta, \gamma_0, \gamma_1, y, h)}{\partial \gamma_1} \bigg|_{\gamma_1=0} \right) \end{bmatrix}.$$
where $\ell_C(\eta; \gamma_1, y, h)$ is the log of $L_C(\eta; \gamma_1, y, h)$, $L_I(\eta; y, h)$ is the likelihood when $\gamma_1 = 0$, and $E_I(\cdot)$ and $\text{COV}_I(\cdot)$ are the posterior expectation and covariance under the ignorable model.

For a function $Q(\cdot)$ of $\theta$, ISNI of $\widehat{Q}(\theta) = E(\theta | y)$ is therefore

$$\text{ISNI}(\widehat{Q}(\theta)) = \text{COV}_I \left( Q(\theta), \frac{\partial \ell_C(\theta; y, h)}{\partial \gamma_1} \bigg|_{\gamma_1 = 0} \right)$$

$$\approx \left( \frac{\partial Q(\theta)}{\partial \theta} \right) \bigg|_{\theta = \hat{\theta}(0)}' \cdot \text{VAR}_I(\theta) \left( \frac{\partial^2 \ell_C(\theta; y, h)}{\partial \gamma_1 \partial \theta} \bigg|_{\gamma_1 = 0, \hat{\theta}(0), \gamma_0(0)} \right).$$

To calculate the sensitivity of the probability that $\eta \in c$, where $c$ is a subset of sample space of $\eta$, we have

$$\text{ISNI}[\Pr[\eta \in c]] = \left. \frac{\partial \int c L_C(\eta; y, h) \pi(\eta) d\eta}{\partial \gamma_1} \right|_{\gamma_1 = 0}$$

$$= \int c \left. \frac{\partial L_C(\eta; y, h)}{\partial \gamma_1} \right|_{\gamma_1 = 0} L_I(\eta; y, h) \pi(\eta) d\eta$$

$$- \int c L_I(\eta; y, h) \pi(\eta) d\eta \int \left. \frac{\partial L_C(\eta; y, h)}{\partial \gamma_1} \right|_{\gamma_1 = 0} L_I(\eta; y, h) \pi(\eta) d\eta$$

$$= \left[ \int L_I(\eta; y, h) \pi(\eta) d\eta \right]^2$$

$$= E_I \left( \frac{\partial \ell_C(\eta; \gamma_1, y, h)}{\partial \gamma_1} \bigg|_{\gamma_1 = 0} \right) - \Pr_I[\eta \in c] E_I \left( \frac{\partial \ell_C(\eta; \gamma_1, y, h)}{\partial \gamma_1} \bigg|_{\gamma_1 = 0} \right),$$

where $\Pr_I[\cdot]$ is the posterior probability under the ignorable model.

**APPENDIX B**

**B.1 ISNI for missing data**

Note that with a $\beta(v, w)$ distribution for the prior of $h(\gamma_0)$, we can think of $v$ and $w$ as prior numbers of observed and missing outcomes. Then, we have

$$\text{ISNI}(\hat{\theta}) = - \left( \sum_{i=1}^n g_i \right) + \frac{v}{n + v + w} \left[ \sum_{i=1}^n (1 - g_i) \text{COV}_I \left( \theta, \int u p_0^{X_i}(u) du \right) \right].$$

If $\theta = (\theta_1, \theta_2)'$ and $p_0^{X_i}(x_i)$ is in the exponential family (McCullagh and Nelder, 1989)

$$p_0^{X_i}(x_i) = \exp\{[x_i \theta_1 - b(\theta_1)]a(\theta_2) + c(x_i, \theta_2)\}$$

for some $a(\cdot), b(\cdot), \text{and } c(\cdot)$, then we have

$$E_\theta(X_i) = b'(\theta_1)$$
\[
\text{ISNI}(\tilde{\theta}) = -\left(\frac{\sum_{i=1}^{n} g_i}{n + v + w}\right) + v \left[\sum_{i=1}^{n} (1 - g_i) \text{COV}_i(\theta, b'(\theta_1))\right].
\]

**Normal linear model.** Assume that we have predictors \(z_i = (z_{1i}, z_{2i}, \ldots, z_{pi})\) and that the \(X_i\) are independent and normally distributed with mean \(z_i \beta\) and variance \(\sigma^2\), which gives in the generalized linear model formulation
\[
b'(\theta_1) = z_i \beta.
\]

The ISNI for \(\tilde{\beta} = \text{E}[\beta|\gamma_1, y, g]\) is
\[
\text{ISNI}(\tilde{\beta}) = -\left(\frac{\sum_{i=1}^{n} g_i}{n + v + w}\right) + v \left[\sum_{i=1}^{n} (1 - g_i) \text{COV}_i(\beta, b'(\theta_1))\right]
\]
\[
= -\frac{(n_o + v)}{n + v + w} \text{VAR}_i(\beta) \left[\sum_{i=1}^{n} (1 - g_i) z_i'\right].
\]

We further assume that the prior distributions of \(\mu\) and \(\log \sigma\) are both uniform and that \(\mu\) and \(\log \sigma\) are a priori independent, yielding
\[
\pi(\beta, \sigma^2|Z) \propto \frac{1}{\sigma^2}.
\]

Then, we have
\[
\beta|y \sim t_{n_o-p}(\tilde{\beta}, s^2(\text{Z}_o'\text{Z}_o)^{-1}),
\]
where \(\text{Z}_o\) is the subset of \(Z\) consisting of all rows for which the corresponding outcome is observed and \(n_o\) is the number of observed subjects, and
\[
\text{ISNI}(\tilde{\beta}) = -\frac{(n_o + v)(n_o - p)}{(n + v + w)(n_o - p - 2)} s^2(\text{Z}_o'\text{Z}_o)^{-1} \sum_{i=1}^{n} (1 - g_i) z_i'.
\]

**Binomial.** Suppose the \(X_i\) are \(n\) independent Bernoulli random variables with \(\text{Pr}[X_i = 1] = \theta_i\). Then, \(\theta_1 = \ln\left(\frac{\theta}{1-\theta}\right), b(\theta_1) = \ln(1 + e^{\theta_1}), a(\theta_2) = 1,\) and \(b'(\theta_1) = \theta\). Furthermore, we assume that \(\theta\) has a conjugate beta prior with parameters \(e\) and \(f\). Therefore, the posterior distribution of \(\theta\) is also beta with parameters \(p\) and \(q\), where \(p = \sum_{i=1}^{n} g_i x_i + e\) and \(q = \sum_{i=1}^{n} g_i (1 - x_i) + f\). ISNI for \(\tilde{\theta} = \text{E}[\theta|\gamma_1, y, g]\) is then
\[
\text{ISNI}(\tilde{\theta}) = -\left(\frac{\sum_{i=1}^{n} g_i}{n + v + w}\right) + v \left[\sum_{i=1}^{n} (1 - g_i) \text{COV}_i(\theta, b'(\theta_1))\right]
\]
\[
= -\frac{(n_o + v)}{n + v + w} \left[\sum_{i=1}^{n} (1 - g_i)\right] \text{VAR}_i(\theta)
\]
\[
= -\frac{(n_o + v)(n - n_o)pq}{(n + v + w)(p + q)^2(p + q + 1)}.
\]
Logistic regression. Assume that we have predictors $z_i = (z_{i1}, z_{i2}, \ldots, z_{ip_i})$ and that the $X_i$ are $n$ independent Bernoulli variables with $\Pr[X_i = 1] = \expit(z_i\beta)$. Then, $b'(\theta_1) = \expit(z_i\beta)$. ISNI for $\tilde{\beta} = \mathbb{E}[\beta|\gamma_1, y, g]$ is

$$
\text{ISNI}(\tilde{\beta}) = -\frac{1}{n + v + w} \left[ \sum_{i=1}^{n} (1 - g_i)\text{COV}_I(\beta, b'(\theta_1)) \right]
$$

$$
= -\frac{n_0 + v}{n + v + w} \left[ \sum_{i=1}^{n} (1 - g_i)\text{COV}_I(\beta, \expit(z_i\beta)) \right]
$$

$$
\approx -\frac{n_0 + v}{n + v + w} \left[ \sum_{i=1}^{n} (1 - g_i) \frac{\exp(z_i\beta_0)}{[1 + \exp(z_i\beta_0)]^2} z_i \right] \text{VAR}_I(\beta),
$$

where $\beta_0$ is the posterior mean of $\beta$ from the ignorable model.

Poisson. Suppose the $X_i$ are $n$ independent Poisson variables with mean $\theta$. Then, $\theta_1 = \ln(\theta), b'(\theta_1) = e^{\theta_1}, a(\theta_2) = 1$, and $\mathbb{E}_0(X_i) = b'(\theta_1) = \theta$. Taking $\theta$ to have a conjugate gamma prior with parameters $e$ and $f$, ISNI for $\tilde{\beta} = \mathbb{E}[^{\theta_1}\gamma_1, y, g]$ is

$$
\text{ISNI}(\tilde{\theta}) = -\frac{1}{n + v + w} \left[ \sum_{i=1}^{n} (1 - g_i)\text{COV}_I(\theta, b'(\theta_1)) \right]
$$

$$
= -\frac{1}{n + v + w} \left[ \sum_{i=1}^{n} (1 - g_i) \text{VAR}_I(\theta) \right]
$$

$$
= -\frac{(n_0 + v)(n - n_0)(e + \sum_{i=1}^{n} g_i x_i)}{(n + v + w)(f + n_0)^2}.
$$

Log-linear regression. Assume that we have predictors $z_i = (z_{i1}, z_{i2}, \ldots, z_{pi})$ and that $X_i$ are $n$ independent Poisson variates with mean $\theta = \exp(z_i\beta)$. Then, $\mathbb{E}_0(X_i) = b'(\theta_1) = \exp(z_i\beta)$. ISNI for $\tilde{\beta} = \mathbb{E}[\beta|\gamma_1, y, g]$ is

$$
\text{ISNI}(\tilde{\beta}) = -\frac{1}{n + v + w} \left[ \sum_{i=1}^{n} (1 - g_i)\text{COV}_I(\beta, b'(\theta_1)) \right]
$$

$$
= -\frac{n_0 + v}{n + v + w} \left[ \sum_{i=1}^{n} (1 - g_i)\text{COV}_I(\beta, \exp(z_i\beta)) \right]
$$

$$
\approx -\frac{n_0 + v}{n + v + w} \left[ \sum_{i=1}^{n} (1 - g_i)z_i \exp(z_i\beta_0) \right] \text{VAR}_I(\beta),
$$

where $\beta_0$ is the posterior mean of $\beta$ from the ignorable model.
**Gamma.** Let $X_i$ be $n$ independent gamma variates with parameters $\theta_i = (\mu_i, \nu)$, where $\nu$ is a constant coefficient of variation and the mean is $\frac{\nu}{\mu_i}$. Then, $\theta_{1i} = -\frac{\mu_i}{\nu}, b(\theta_{1i}) = -\log(-\theta_{1i})$, and $a(\theta_2) = \theta_2 = \nu$. Furthermore, we assume that we have predictors $z_i = (z_{1i}, z_{2i}, \ldots, z_{pi})$ with $\nu = \frac{1}{\beta_i}$. ISNI for $\tilde{\beta} = E[\beta|\gamma_1, y, g]$ is then

$$\text{ISNI}(\tilde{\beta}) = -\left(\sum_{i=1}^{n} g_i + \nu \left[ \sum_{i=1}^{n} (1 - g_i) \text{COV}_I(\beta, b'(\theta_{1i})) \right] \right)$$

$$= -\frac{n_0 + \nu}{n + \nu + w} \sum_{i=1}^{n} (1 - g_i) \text{COV}_I \left( \beta, \frac{1}{\beta_i} \right)$$

$$\approx -\frac{n_0 + \nu}{n + \nu + w} \left[ \sum_{i=1}^{n} (1 - g_i) \frac{1}{\beta_i^3} \right] \text{VAR}_I(\beta),$$

where $\tilde{\beta}_0$ is the posterior mean of $\beta$ from the ignorable model.

**Inverse Gaussian.** Assume the $X_i$ are $n$ independent inverse Gaussian variables with parameters $\theta_i = (\mu_i, \sigma^2)$, where $\sigma^2$ is the variance. Then, $\theta_{1i} = -\frac{1}{2\mu_i}, b(\theta_{1i}) = (-2\theta_{1i})^{-1/2}$, and $a(\theta_2) = 1/\theta_2 = 1/\sigma^2$. Furthermore, we assume that we have predictors $z_i = (z_{1i}, z_{2i}, \ldots, z_{pi})$ with $\mu_i = \sqrt{\frac{1}{\beta_i}}$. ISNI for $\tilde{\beta} = E[\beta|\gamma_1, y, g]$ is then

$$\text{ISNI}(\tilde{\beta}) = -\left(\sum_{i=1}^{n} g_i + \nu \left[ \sum_{i=1}^{n} (1 - g_i) \text{COV}_I(\beta, b'(\theta_{1i})) \right] \right)$$

$$= -\frac{n_0 + \nu}{n + \nu + w} \sum_{i=1}^{n} (1 - g_i) \text{COV}_I \left( \beta, \sqrt{\frac{1}{\beta_i}} \right)$$

$$\approx -\frac{n_0 + \nu}{n + \nu + w} \left[ \sum_{i=1}^{n} (1 - g_i) \sqrt{\frac{1}{\beta_i^3}} \right] \text{VAR}_I(\beta),$$

where $\tilde{\beta}_0$ is the posterior mean of $\beta$ from the ignorable model.

**APPENDIX C**

C.1 **ISNI for censored data**

Using the notation of the body of the paper, ISNI for the parameter of a general censored data–selection model is

$$\text{ISNI}(\tilde{\theta}) = \text{COV}_I \left( \theta, \sum_{i=1}^{n} (1 - d_i) \left[ \int_{g_i}^{\infty} u p_{\theta}^{X}(u)du \right] \frac{1 - e^{\gamma_0 H_0(g_i)}}{\int_{g_i}^{\infty} p_{\theta}^{X}(u)du} \right)$$

$$= \text{COV}_I \left( \theta, \sum_{i=1}^{n} (1 - d_i) E_{\theta}[X_i|X_i > g_i] \frac{1 - e^{\gamma_0 H_0(g_i)}}{\int_{g_i}^{\infty} p_{\theta}^{X}(u)du} \right)$$
Impact of nonignorable coarsening on Bayesian inference

\[
\text{COV}_I \left( \theta, \sum_{i=1}^{n} (1 - d_i) \left( g_i + \frac{\int_{g_i}^{\infty} S_{\hat{\theta}}^X(u)du}{S_{\hat{\theta}}^X(g_i)} \right) [1 - e^{\gamma_0 H_0(g_i)}] \right)
\]

\[
= \text{COV}_I \left( \theta, \sum_{i=1}^{n} (1 - d_i) \frac{\int_{g_i}^{\infty} S_{\hat{\theta}}^X(u)du}{S_{\hat{\theta}}^X(g_i)} [1 - e^{\gamma_0 H_0(g_i)}] \right),
\]

where \( S_{\theta}^X(u) = \int_{u}^{\infty} p_{\theta}^X(l)dl \).

**Weibull survival with conditional exponential censoring.** We assume that the marginal distribution of \( X_i \) is Weibull with parameter \( \theta = (\sigma, \lambda_i) \) and mean \( \Gamma(1 + \frac{1}{\sigma})\lambda_i \), where \( \lambda_i = e^{z_i \beta} \):

\[
p_{\sigma, \lambda}^X(x_i) = \frac{\sigma}{\lambda_i} \left( \frac{x_i}{\lambda_i} \right)^{\sigma - 1} \exp \left( - \left[ \frac{x_i}{\lambda_i} \right]^{\sigma} \right).
\]

ISNI for \( \tilde{\beta} \) is then

\[
\text{ISNI} (\tilde{\beta}) = \text{COV}_I \left( \beta, \sum_{i=1}^{n} (1 - d_i) \frac{\int_{g_i}^{\infty} S_{\hat{\theta}}^X(u)du}{S_{\hat{\theta}}^X(g_i)} [1 - e^{\gamma_0 H_0(g_i)}] \right)
\]

\[
= \text{COV}_I \left( \beta, \sum_{i=1}^{n} (1 - d_i) \frac{\int_{g_i}^{\infty} \exp \left( \left[ \frac{u}{e^{\gamma_0 g_i}} \right]^{\sigma} \right)du}{\exp \left( \left[ \frac{g_i}{e^{\gamma_0 g_i}} \right]^{\sigma} \right)} [1 - e^{\gamma_0 g_i}] \right)
\]

\[
= \text{COV}_I \left( \beta, \sum_{i=1}^{n} (1 - d_i) \left[ 1 - k + \frac{\sum_{i}(1 - d_i)}{f + \sum_{i} t_i} \right] \right)
\]

\[
\approx \text{COV}_I \left( \beta, \sum_{i=1}^{n} (1 - d_i) \left[ 1 - k + \frac{\sum_{i}(1 - d_i)}{f + \sum_{i} t_i} \right] \frac{\int_{g_i}^{\infty} \exp \left( \left[ \frac{u}{e^{\gamma_0 g_i}} \right]^{\sigma} \right)du}{\exp \left( \left[ \frac{g_i}{e^{\gamma_0 g_i}} \right]^{\sigma} \right)} \right)_{\beta = \bar{\beta}_0},
\]

where \( \bar{\beta}_0 \) is the posterior mean of \( \beta \) from the ignorable model.

**Exponential survival with conditional Weibull censoring.** Assume that the marginal distribution of \( X_i \) is exponential with mean \( e^{-\theta} \), which has a gamma prior with parameters \( p \) and \( q \), and \( h_0(g_i) = \gamma_2 g_i^{\gamma_2 - 1} \).

ISNI for the posterior mean survival \( e^{-\theta} \) is

\[
\text{ISNI} (e^{-\theta}) = \text{COV}_I \left( e^{-\theta}, \sum_{i=1}^{n} (1 - d_i) \frac{\int_{g_i}^{\infty} S_{\hat{\theta}}^X(u)du}{S_{\hat{\theta}}^X(g_i)} [1 - e^{\gamma_0 H_0(g_i)}] \right)
\]

\[
= \text{COV}_I \left( e^{-\theta}, \sum_{i=1}^{n} (1 - d_i) e^{-\theta} (1 - e^{\gamma_0 g_i^{\gamma_2}}) \right)
\]
\[ \sum_{i=1}^{n} (1 - d_i) (1 - E_i [e^{\gamma_0 g_i^2}]) \text{VAR}_i (e^{-\theta}) \]
\[ = \sum_{i=1}^{n} (1 - d_i) (1 - E_i [e^{\gamma_0 g_i^2}]) \frac{(q + \sum_i t_i)^2}{(p + \sum_i d_i - 1)^2 (p + \sum_i d_i - 2)} , \]

where \( t_i = \min(x_i, g_i) \).

REFERENCES


Impact of nonignorable coarsening on Bayesian inference


[Received April 7, 2006; revised December 11, 2006; accepted for publication December 20, 2006]