*University of Michigan School of Public Health*

The University of Michigan Department of Biostatistics Working Paper Series

*Year* 2013 *Paper* 98

# MISSING AT RANDOM AND IGNORABILITY FOR INFERENCES ABOUT SUBSETS OF PARAMETERS WITH MISSING DATA

Roderick J. Little* Sahar Zanganeh†

*University of Michigan, rlittle@umich.edu
†University of Michigan - Ann Arbor



http://biostats.bepress.com/umichbiostat/paper98

# MISSING AT RANDOM AND IGNORABILITY FOR INFERENCES ABOUT SUBSETS OF PARAMETERS WITH MISSING DATA

Roderick J. Little and Sahar Zanganeh

**Abstract**

For likelihood-based inferences from data with missing values, Rubin (1976) showed that the missing data mechanism can be ignored when (a) the missing data are missing at random (MAR), in the sense that missingness does not depend on the missing values after conditioning on the observed data, and (b) the parameters of the data model and the missing-data mechanism are distinct; that is, there are no a priori ties, via parameter space restrictions or prior distributions, between the parameters of the data model and the parameters of the model for the mechanism. Rubin described (a) and (b) as the "weakest simple and general conditions under which it is always appropriate to ignore the process that causes missing data". However, these conditions are not always necessary. Also, they relate to the complete set of parameters in the model, but we argue that it would be useful to have definitions of MAR and ignorability for a subset of parameters of substantive interest. We propose such definitions, and apply them to a variety of examples where the missing data mechanism is missing not at random, but MAR or ignorable for the parameter subset.

Draft February 10, 2013

# MISSING AT RANDOM AND IGNORABILITY FOR INFERENCES ABOUT SUBSETS OF PARAMETERS WITH MISSING DATA

Roderick J.A. Little and Sahar Z Zangeneh

Departments of Biostatistics and Statistics, University of Michigan, Ann Arbor, MI 48105

## ABSTRACT

For likelihood-based inferences from data with missing values, Rubin (1976) showed that the missing data mechanism can be ignored when (a) the missing data are missing at random (MAR), in the sense that missingness does not depend on the missing values after conditioning on the observed data, and (b) the parameters of the data model and the missing-data mechanism are distinct; that is, there are no a priori ties, via parameter space restrictions or prior distributions, between the parameters of the data model and the parameters of the model for the mechanism. Rubin described (a) and (b) as the "weakest simple and general conditions under which it is always appropriate to ignore the process that causes missing data". However, these conditions are not always necessary. Also, they relate to the complete set of parameters in the model, but we argue that it would be useful to have definitions of MAR and ignorability for a subset of parameters of substantive interest. We propose such definitions, and apply them to a variety of examples where the missing data mechanism is missing not at random, but MAR or ignorable for the parameter subset.

**Key words:** Incomplete data, likelihood theory, missing data mechanism, partial likelihood, Bayes inference.

## 1. Introduction

We consider likelihood-based inference for parameters from data with missing values. Let *D* denote the set of complete data if there were no missing values, and *R* a set of binary variables indicating whether individual components of *D* are observed (1) or missing (0). We initially model the density of the joint distribution of *D* and *R* using the "selection model" factorization (Little and Rubin, 2002):

$$f_{D,R}(D,R \mid \theta,\phi) = f_D(D \mid \theta) f_{R|D}(R \mid D,\phi), \tag{1}$$

where $\theta$ is the parameter of the data model, and $\phi$ is the parameter of the model for the missing data mechanism. Let $D = (D_{\text{obs}}, D_{\text{mis}})$, where $D_{\text{obs}}$ is the observed part of *D* and $D_{\text{mis}}$ is the missing part of *D*. Then the full likelihood based on the observed data and the assumed model is

$$L(\theta,\phi \mid D_{\text{obs}}, R) = \text{const.} \times \int f_D(D \mid \theta) f_{R|D}(R \mid D,\phi) dD_{\text{mis}}, \tag{2}$$

treated as a function of the parameters $(\theta,\phi)$. The likelihood of $\theta$ *ignoring the missing-data mechanism* is

$$L(\theta \mid D_{\text{obs}}) = \text{const.} \times \int f_D(D \mid \theta) dD_{\text{mis}}, \tag{3}$$

which does not involve the model for *R*. In a landmark paper, Rubin (1976) noted that when the missing data are missing at random (MAR), defined as

$$f_{R|D}(R \mid D_{\text{obs}}, D_{\text{mis}}, \phi) = f_{R|D}(R \mid D_{\text{obs}}, \phi) \text{ for all } D_{\text{mis}}, \phi, \tag{4}$$

the full likelihood Eq. (2) factorizes as

$$L(\theta,\phi \mid D_{\text{obs}}, M) = \text{const.} \times f_D(D_{\text{obs}} \mid \theta) \times f_{R|D}(R \mid D, \phi)\,. \tag{5}$$

Hence, likelihood inference based on (3) is valid if the data are MAR, and fully efficient if $\theta$ and $\phi$ are distinct, that is, their joint parameter space is the product of the parameter space of $\theta$ and $\phi$. For Bayesian inference, distinctness involves the additional assumption that $\theta$ and $\phi$ are a priori independent (Little and Rubin, 2002).

These definitions are illustrated in the following simple example.

**Example 1. Monotone Bivariate Data**

Let $D = \{(y_{1i}, y_{2i}), i = 1,...,n\}$ denote an independent sample from two variables $Y_1, Y_2$ with probability density $f(y_{1i}, y_{2i} \mid \theta)$ indexed by unknown parameters $\theta$. Suppose $D_{\text{obs}} = \{(y_{1i}, y_{2i}), i = 1,...,m\}$ and $\{y_{1i}, i = m+1,...,n\}$, so that $Y_1$ is fully observed and $Y_2$ has missing values (Figure 1A). Let $R = \{r_i\}, i = 1,...,n$ where $r_i = 1$ if $y_{2i}$ is observed and $r_i = 0$ if $y_{2i}$ is missing. Missingness of $Y_2$ is assumed to depend only on $Y_1$, that is:

$$\Pr(r_i = 1 \mid y_{1i}, y_{2i}, \phi) = g(y_{1i}, \phi)\,, \tag{6}$$

where $g$ is a known function with support between 0 and 1. This mechanism meets definition (4) of MAR; likelihood inferences for $\theta$ are then ignorable if the parameters $\theta$ and $\phi$ are distinct, and Bayesian inferences are ignorable if in addition $\theta$ and $\phi$ are a priori independent.

Since Rubin (1976), MAR has come to be defined by Eq. (4), and ignorability of the missing-data mechanism by Eq. (4) together with the distinctness condition (see example Little and Rubin, 2002). However, Rubin described them as the "weakest simple and general conditions under which it is always appropriate to ignore the process that causes missing data", and it is important to note that these conditions are not necessary for ignoring the mechanism in all situations. Also, MAR and ignorability are defined in terms of the complete set of parameters $\theta$ in the model for *D*, but it would be useful to have a definition of MAR that applies to subsets of parameters of substantive interest. Here are some motivating examples.

**Example 2. Ignorability for parameters of distributions of fully-observed variables.** In Example 1, suppose we assume that missingness of $Y_2$ depends on both $Y_1$ and $Y_2$:

$$\Pr(r_i = 1 \mid y_{1i}, y_{2i}, \phi) = g(y_{1i}, y_{2i}, \phi), \tag{7}$$

where *g* is a known function with support between 0 and 1. This mechanism is missing not at random (MNAR), but is plausibly MAR for inference about the parameters of the marginal distribution of $Y_1$, since $Y_1$ is fully observed and missingness of $Y_2$ seems irrelevant for this inference. More specifically, suppose we write $\theta = (\theta_1, \theta_2)$ and factorize the joint distribution of $Y_1$ and $Y_2$ as

$$f(y_{1i}, y_{2i} \mid \theta) = f_1(y_{1i} \mid \theta_1) \times f_2(y_{2i} \mid y_{1i}, \theta_2), \tag{8}$$

where $\theta_1$ and $\theta_2$ denote the parameters for the marginal distribution of $Y_1$ and conditional distribution of $Y_2$ given $Y_1$ respectively. Then it seems that the mechanism should be MAR for $\theta_1$, and ignorable for $\theta_1$ if $\theta_1$ and $(\theta_2, \phi)$ are distinct sets of parameters. This is the case for the formal definitions we propose below. This example is somewhat trivial, but we extend it below to a more complex situation where $Y_1$ and $Y_2$ are blocks of (possibly incomplete) variables.

**Example 3. Outcome-dependent dropout in clinical trials, where valid treatment effects are estimated from respondents.** In randomized clinical trials where the outcomes are missing for participants who drop out, a similar response rate in the treatment arms is commonly thought to increase the possibility that biases from nonrandom nonresponse will cancel out. The resulting estimate of the parameter that compares the two treatments would therefore be valid, even though the estimate of the mean for each treatment group is biased. This motivates the question: is it possible to define mechanisms where the outcome is MNAR overall, but MAR for the parameters of interest, measuring treatment effects? The answer is yes, and we show below how our expanded definition of MAR and ignorability can be applied in this setting. We also show that in a discrete data setting, a more general condition in the response mechanism than "equal response rates" is sufficient for treatment effects to be estimable from the respondent sample.

**Example 4. Regression with the missing-data mechanism tailored to predictors.** In regression with missing predictors, data are MNAR when missingness depends on the underlying missing values of a predictor. For example, Little and Zhang (2011) analyze data from the 2003-2004 National Health and Nutrition Examination Survey (CDC 2004) to study the effect of socioeconomic status on blood pressure. Regressions of two outcome measures, systolic blood pressure (SBP) and diastolic blood pressure (DBP) are estimated on two socioeconomic status measures, household income (HHINC) and years of education (EDUC, in years) and three other covariates: AGE (in years), GENDER, and body mass index (BMI, kg/m$^2$). The covariates, AGE and GENDER are fully observed, but the predictor variables HHINC, EDUC and BMI have missing values. The MAR assumption (4) is considered implausible, since the probability of responding to HHINC is thought likely to depend on the underlying (sometimes missing) value of HHINC – individuals with high or low values of income are often considered less likely to respond to income than others. On the other hand, it may be reasonable to assume that education and BMI are “MAR” -- but the current definition does not allow some the mechanism to be MAR for coefficients of some variables and MNAR for others. Hence it would be useful to have distinct definitions of MAR and ignorability tailored to missingness of *W* = HHINC and *X* = (EDUC, BMI). Our proposed definitions accomplish this.

**Example 5. A sample with auxiliary data where the mechanism is MNAR but ignorable.** In Example 1, suppose $Y_1$ as well as $Y_2$ is missing when $r_i = 0$, so the respondent data consist only of the complete cases, $D_{\text{resp}} = \{(y_{1i}, y_{2i}), i = 1,...,m\}$. The joint distribution of $Y_1$ and $Y_2$ is factored as in Eq. (8), and missingness depends on $Y_1$ but not $Y_2$, as given by Eq. (6). Then the mechanism is MNAR, since missingness depends on values of $Y_1$ which are missing for the incomplete cases. Suppose, however, we also have auxiliary information on the marginal distribution of $Y_1$ from an external source, with observations not linked to the respondent sample (Figure 1B). The observed data are then $D_{\text{obs}} = (D_{\text{resp}}, D_{\text{aux}})$, where $D_{\text{aux}} = \{y^*_{1j}, j = 1,...,n\}$. The latter set includes the respondent values of $Y_1$, but we do not know which they are. Data of this form arise in sample surveys, where the external data are available for the whole sample or for the entire population from a census. The mechanism is technically MNAR, since the mechanism depends on $Y_1$, but we do not know the values of $Y_1$ for individual nonrespondents. However, intuitively the marginal distribution of $Y_1$ can be estimated from $D_{\text{aux}}$, and the conditional distribution of $Y_2$ given $Y_1$ can be estimated from $D_{\text{resp}}$, without modeling the missing-data mechanism. In our definitions below this mechanism is MAR for $\theta$, and ignorable for $\theta$ if the parameters $\theta$ and $\phi$ are distinct. This is an example where Rubin's conditions are not necessary.

In Section 2, we propose definitions of MAR and ignorability for likelihood inferences about subsets of model parameters, and relate them to Rubin's (1976) definitions for all the parameters. In Section 3 we then show how our definitions address the issues in our motivating examples. Conclusions are summarized in Section 4.

## 2. Definitions of MAR and Ignorability for Parameter Subsets

We propose a definition of missing at random for likelihood inferences for a subset $\theta_1$ of the parameters $\theta$ in a model.

**Definition 1**: Write $\theta = (\theta_1, \theta_2)$, where $\theta_1$ and $\theta_2$ are subsets of parameters, and let $\phi$ denote the parameters for a model for the missing-data mechanism *R*. The missing data mechanism is MAR for inference about $\theta_1$, denoted $\text{MAR}(\theta_1)$, if the likelihood (1) can be factorized as

$$L(\theta_1, \theta_2, \phi \mid D_{\text{obs}}, R) = \text{const.} \times L_1(\theta_1 \mid D_{\text{obs}}, R) \times L_{\text{rest}}(\theta_2, \phi \mid D_{\text{obs}}, R) \text{ for all } \theta_1, \theta_2, \phi. \tag{9}$$

**Definition 2.** The missing data mechanism is ignorable for likelihood inference about $\theta_1$, denoted $\text{LIGN}(\theta_1)$, if (a) the missing data mechanism is $\text{MAR}(\theta_1)$, and (b) $\theta_1$ and $(\theta_2, \phi)$ are distinct sets of parameters, in the sense defined by Rubin (1976).

Under $\text{MAR}(\theta_1)$, likelihood inference about $\theta_1$, or functions of $\theta_1$, can be based on $L_1(\theta_1 \mid D_{\text{obs}}, R)$, which does not involve a model for the mechanism *R*. Under $\text{LIGN}(\theta_1)$, inference based on $L_1(\theta_1 \mid D_{\text{obs}}, R)$ is fully efficient. If the mechanism is $\text{MAR}(\theta_1)$ but $\theta_1$ and $(\theta_2, \phi)$ are not distinct sets of parameters, likelihood inference based on $L_1(\theta_1 \mid D_{\text{obs}}, R)$ is valid but not fully efficient, and might still be entertained to avoid the additional assumptions involved in modeling the mechanism. For Bayesian inference, the mechanism can be ignored if, in addition, $\theta_1$ and $(\theta_2, \phi)$ are a-priori independent. The posterior distribution of $\theta_1$ is then

$$p(\theta_1 \mid D, R) = \text{const} \times \pi_1(\theta_1) \times L(\theta_1 \mid D_{\text{obs}}, R), \tag{10}$$

where $\pi_1(\theta_1)$ is the prior distribution of $\theta_1$. Note that (10) does not involve the parameters of the model for the mechanism.

When $\theta_1 = \theta$, these definitions deviate slightly from Rubin's (1976) original sufficient conditions. The distinctness condition reduces to distinctness between $\theta$ and $\phi$, as defined by Rubin. The $\text{MAR}(\theta_1)$ condition, Eq. (9), with $\theta_1 = \theta$ is less restrictive than

Rubin's MAR definition, Eq. (4), but it does imply Eq. (5), which is the key condition for validity of inferences about $\theta$ based on the ignorable likelihood Eq. (3).

## 3. The Proposed Definitions Applied to the Examples

We now apply these definitions to our motivating examples.

**Example 2 (ctd). Ignorability for parameters of distributions of fully-observed variables.** For the data pattern in Figure 1A, the factorization of the joint distribution of $Y_1$ and $Y_2$ in Eq. (8), and the mechanism defined by Eq. (7), the likelihood factors as

$L(\theta_1,\theta_2,\phi \mid D_{\text{obs}},R) = \text{const.} \times L_1(\theta_1 \mid D_{\text{obs}},R) \times L_{\text{rest}}(\theta_2,\phi \mid D_{\text{obs}},R)$ for all $\theta_1,\theta_2,\phi,$ where

$$L_1(\theta_1 \mid D_{\text{obs}},R) = \prod_{i=1}^{n} f_1(y_{1i},\theta_1)$$

$$L_{\text{rest}}(\theta_2,\phi \mid D_{\text{obs}},R) = \prod_{i=1}^{m} f_2(y_{2i} \mid y_{1i},\theta_2) g(y_{1i},y_{2i},\phi) \times \prod_{i=m+1}^{n} \int f_2(y_{2i} \mid y_{1i},\theta_2)\big((1-g(y_{1i},y_{2i},\phi)\big) dy_{2i}.$$

Hence the mechanism is MNAR, but it is $\text{MAR}(\theta_1)$, and $\text{LIGN}(\theta_1)$ if $\theta_1$ and $(\theta_2,\phi)$ are distinct sets of parameters. Without distinctness, there is potential information about $\theta_1$ in $L_{\text{rest}}(\theta_2,\phi \mid D_{\text{obs}},R)$, but recovering the information requires correctly specifying a model for the mechanism, with parameters that can be identified from the data. The simplicity of inference based on $L_1(\theta_1 \mid D_{\text{obs}},R)$ often outweighs the potential loss of information.

More generally, suppose $Y_1$ and $Y_2$ are sets of (possibly incomplete) variables, and let $r_i^{(1)}, r_i^{(2)}$ denote response indicators for $Y_1$ and $Y_2$ in observation *i*. We adopt a block-conditional factorization (Zhou, Kalbfleisch and Little, 2010) of the joint density:

$$f(y_{1i},y_{2i},r_i^{(1)},r_i^{(2)} \mid \theta_1,\theta_2,\phi_1,\phi_2) = \\ f_1(y_{1i} \mid \theta_1)\Pr(r_i^{(1)} \mid y_{1i},\phi_1) f_2(y_{2i} \mid y_{1i},r_i^{(1)},\theta_2)\Pr(r_i^{(2)} \mid r_i^{(1)},y_{1i},y_{2i},\phi_2),$$

and assume the missing-data mechanism:

$$\Pr(r_i^{(1)} \mid y_{1i};\phi_1) = g_1(y_{1,\text{obs},i},\phi_1) \text{ for all } y_{1,\text{mis},i}, \tag{11}$$

$$\Pr(r_i^{(2)} \mid r_i^{(1)}, y_{1i}, y_{2i};\phi_2) = g_2(r_i^{(1)}, y_{1i}, y_{2i},\phi_2), \tag{12}$$

where $y_{j.\text{obs},i}, y_{j.\text{mis},i}$ denote the observed and missing components of $Y_j$ for unit *i*. This mechanism is MNAR because of Eq. (12), but the resulting likelihood is

$L(\theta_1,\theta_2,\phi \mid D_{\text{obs}},R) = \text{const.}\times L_1(\theta_1 \mid D_{\text{obs}},R)\times L_{\text{rest}}(\theta_2,\phi_1,\phi_2 \mid D_{\text{obs}},R)$ for all $\theta_1,\theta_2,\phi_1,\phi_2$, where

$$L_1(\theta_1 \mid D_{\text{obs}},R) = \prod_{i=1}^{n} f_1(y_{1,\text{obs},i},\theta_1),$$

$$L_{\text{rest}}(\theta_2,\phi_1,\phi_2 \mid D_{\text{obs}},R) = \prod_{i=1}^{n} g_1(r_i^{(1)} \mid y_{1,\text{obs},i},\phi_1)\int f_2(y_{2i} \mid y_{1i}, r_i^{(1)},\theta_2) g_2(r_i^{(1)}, y_{1i}, y_{2i},\phi_2) dy_{1,\text{mis},i} dy_{2,\text{mis},i}$$

Hence the mechanism is MNAR, but it is $\text{MAR}(\theta_1)$, and $\text{LIGN}(\theta_1)$ if $\theta_1$ and $(\theta_2,\phi_1,\phi_2)$ are distinct sets of parameters.

**Example 3 (ctd). Outcome-dependent dropout in clinical trials, where valid treatment effects are estimated from respondents.** In a randomized clinical trial, let $X$ be a variable indicating $T + 1$ treatment groups ($X = 0,1,\ldots,T$), and $Y$ a categorical outcome variable with $K + 1$ distinct values $y = 0,1,\ldots K$. We assume missing data are confined to $Y$, and let $R = 1$ if $Y$ is observed and $R = 0$ if $Y$ is missing. We model the joint distribution of $(Y,R \mid X)$ using the pattern-mixture factorization (Little, 2003), with a logistic model for outcomes in the respondent and nonrespondent strata:

$$\Pr(Y = y, R = j \mid X = x,\theta,\phi) = \Pr(R = j \mid X = x,\phi)\Pr(Y = y \mid R = r, X = x,\theta)$$

$$\phi = (\phi_0, \phi_1, ..., \phi_T), \Pr(R = 1 \mid X = x, \phi) = 1 - \Pr(R = 0 \mid X = x, \phi) = \phi_x$$
$$\theta = (\theta^{(0)}, \theta^{(1)}), \theta^{(r)} = \left((\theta_{0y}^{(r)}, \theta_{1y}^{(r)}), y = 1, ..., K\right), r = 0,1$$
$$\log \frac{\Pr(Y = y \mid R = r, X = x, \theta)}{\Pr(Y = 0 \mid R = r, X = x, \theta)} = \theta_{0y}^{(r)} + \theta_{1y}^{(r)} x$$

The likelihood for the observed data is then $L(\theta,\phi) = L_1(\theta^{(1)} \mid Y_{\text{obs}}, R) \times L_2(\phi \mid R),$ where

$$L_1(\theta^{(1)} \mid Y_{\text{obs}}, R) = \prod_{i=1}^{r} \Pr(Y = y_i \mid r_i = 1, x_i, \theta),$$
$$L_2(\phi \mid R) = \prod_{t=0}^{T} \phi_t^{r_t} (1 - \phi_t)^{(n_t - r_t)},$$

and $n_t$ and $r_t$ are respectively the sample size and number of respondents in treatment group $X = t$. There is no information in the data for $\theta^{(0)}$. By the above definitions, the mechanism is MAR($\theta^{(1)}$) and LIGN($\theta^{(1)}$) if $\theta^{(1)}$ and $\phi$ are distinct. However, since $\theta^{(1)}$ concerns the distribution of $Y$ for respondents, it is generally not a valid measure of the quantities of interest, namely the effects of treatments in the whole sample. Suppose, however, we assume that the log odds of response is an additive function of treatment group and outcome, that is

$$\operatorname{logit} \Pr(R = 1 \mid X = x, Y = y) = \beta_0 + \beta_{1x} + \beta_{2y}, \tag{13}$$

where $\{\beta_0, \beta_{1x}, \beta_{2y}\}$ are functions of $\theta, \phi$. Equivalently, we assume the loglinear model [*XY*, *RX*, *RY*] for the contingency table defined by *R*, *X*, and *Y*, with the three-way associations set to zero. This assumption implies that

$$\log \frac{\Pr(Y = y \mid R = r, X = x, \theta)}{\Pr(Y = 0 \mid R = r, X = x, \theta)} = \theta_{0y}^{(r)} + \theta_{1y} x, \quad y = 1, ..., K,$$

That is, $\theta_{1y}^{(1)} = \theta_{1y}$, so the mechanism is MAR for $\theta_{1y}$, which measures the effects of treatments on the log odds ratio for $Y = y$ relative to $Y = 0$ in the whole sample. Thus, inferences based on the respondents are valid for these parameters. A special case of Eq. (13) is the assumption that missingness depends on *Y* but not *X* ($\beta_{1x} = 0$), which might be reasonable in a study where participants are blinded to treatment, and drop out is related to the value of the outcome but not the treatment received.

**Example 4 (ctd). Regression with missing-data mechanisms tailored to predictors.** Following Little and Zhang (2011), let (*Z*, *W*, *X* and *Y*) be (possibly vector-valued) variables, where interest concerns the regression of *Y* on predictors *Z*, *W* and *X*. The data are displayed in Figure 2; variables *Z* are fully observed, *W* and *X* have missing values, and *Y* may or may not have missing values. Let $R_{w_i}, R_{(x_i, y_i)}$ respectively denote the missing data pattern for $w_i$ and $(x_i, y_i)$ for observation *i*. The observations are grouped into two patterns: Pattern 1 (P1) consists of cases where *W* is fully observed ($R_{w_i} = u_w$), where $u_w$ denotes a vector of ones with the same dimension as *W*. Pattern 2 (P2) consists of cases with *W* missing or incomplete ($R_{w_i} = \overline{u}_w$). In both P1 and P2, the pattern of missing data for *X* and *Y* is arbitrary. Interest concerns the parameters $\theta_{y \cdot zwx}$ of the distribution of *Y* given (*Z*, *W*, *X*), say $f(y_i \mid z_i, w_i, x_i, \theta_{y.zwx})$.

The division of covariates into *W* and *X* is determined by the following assumptions about the missing data mechanism:

(a) <u>Covariate missingness of *W*</u>: the probability that *W* is fully observed depends only on the covariates and not *Y*, that is:

$$\Pr\left(R_{w_i} = u_w \mid z_i, w_i, x_i, y_i, \phi_w)\right) = \Pr\left(R_{w_i} = u_w \mid z_i, w_i, x_i, \phi_w)\right) \text{ for all } y_i \tag{14}$$

(b) <u>Subsample MAR of $X$, $Y$</u>: Missingness of $X$ and $Y$ is MAR within the subsample (P1) of cases for which $W$ is fully observed, that is:

$$\begin{aligned}&\Pr(R_{(x_i,y_i)} \mid z_i, w_i, x_i, y_i, R_{w_i} = u_w; \phi_{xy\cdot w}) = \\ &\qquad \Pr(R_{(x_i,y_i)} \mid z_i, w_i, x_{\text{obs},i}, y_{\text{obs},i}, R_{w_i} = u_w; \phi_{xy\cdot w}) \quad \text{for all } x_{\text{mis},i}, y_{\text{mis},i}\end{aligned} \tag{15}$$

The mechanism defined by Eqs. (14) and (15) is missing not at random, but we show that valid inferences for $\theta_{y\cdot zwx}$ can be based on likelihood for the data in P1, discarding the data in P2, without modeling the missing data mechanism. Little and Zhang (2011) model the distribution, conditional on the fully observed covariates $Z$, as

$$[W,X,Y,R_w,R_x,R_y \mid Z] = [R_w \mid Z,\phi_w][W,X,Y \mid R_w,Z,\theta][R_x,R_y \mid W,X,Y,R_w,Z,\phi_{xy}]$$

Here the joint distribution of $W$, $X$ and $Y$ given $Z$ is modeled separately in each pattern defined by $R_w$. Let $\theta$ denote the collective set of parameters of these distributions, and write $\theta = (\theta_1, \theta_2)$, where $\theta_1$ are the parameters of the distribution of $W$, $X$ and $Y$ given $Z$ in P1 $(R_{w_i} = u_w)$ and $\theta_2$ are the parameters of the distributions of $W$, $X$ and $Y$ given $Z$ in P2 $(R_{w_i} \neq u_w)$. The likelihood of the observed data factors as follows:

$$L(\theta,\phi_w,\phi_{xy} \mid \text{data}) = L_1(\theta_1) \times L_2(\phi_w,\phi_{xy}) \times L_{\text{rest}}(\theta_2,\phi_w,\phi_{xy}), \tag{16}$$

$$\begin{aligned}
L_1(\theta_1) &= \prod_{i \in P_1} f\left(w_i, x_{\text{obs},i}, y_{\text{obs},i} \mid z_i, R_{w_i} = u_w, \theta_1\right) \\
L_2(\phi_w,\phi_{xy}) &= \prod_{i \in P_1} \Pr(R_{w_i} = u_w \mid z_i, \phi_w) \Pr(R_{x_i}, R_{y_i} \mid R_{w_i} = u_w, z_i, w_i, x_{\text{obs},i}, y_{\text{obs},i}, \phi_{xy}) \\
L_{\text{rest}}(\theta_2,\phi_w,\phi_{xy}) &= \prod_{i \in P_2} \Pr(R_{w_i} \mid z_i, \phi_w) f(R_{x_i}, R_{y_i}, w_i, x_{\text{obs},i}, y_{\text{obs},i}, \mid R_{w_i}, z_i, \theta_2, \phi_{xy})
\end{aligned}$$

Here the likelihood from the data in P1 factors into $L_1(\theta_1)$ and $L_2(\phi_w,\phi_{xy})$ as a result of the "subsample MAR" condition Eq. (15). Hence, the missing data mechanism is

$\text{MAR}(\theta_1)$ and $\text{LIGN}(\theta_1)$ if $\theta_1$ and $(\theta_2, \psi_w, \psi_{xy})$ are distinct. By Eq. (14), the joint distribution of *W*, *X* and *Y* given *Z* in P1 $(R_{w_i} = u_w)$ factors as

$$[W, X, Y \mid R_w = u_w, Z, \theta_1] = [Y \mid X, W, Z, \theta_{y.zwx}(\theta_1)][W, X \mid R_w = u_w, Z, \theta_1],$$

Where $\theta_{y \cdot zwx} = \theta_{y \cdot zwx}(\theta_1)$ are the parameters of the regression of interest, namely the regression of *Y* on *X*, *W* and *Z* for the whole sample. Hence the mechanism is $\text{MAR}(\theta_{y \cdot zwx})$ and $\text{LIGN}(\theta_{y \cdot zwx})$ if $\theta_1$ and $(\theta_2, \phi_w, \phi_{xy})$ are distinct. That is, we have established that a likelihood-based analysis based on the data in P1 is valid for $\theta_{y \cdot zwx}$ without specifying the missing data mechanism. Little and Zhang (2011) call this approach subsample ignorable likelihood (SSIL) analysis. The omitted factor in the likelihood $L_{\text{rest}}(\theta_2, \phi_w, \phi_{xy})$ from P2 potentially has information about $\theta_{y \cdot zwx}$, but extracting it requires a model for the missing-data mechanism.

In the specific example cited above with outcome measures systolic blood pressure (SBP) and diastolic blood pressure (DBP), predictors with missing values household income (HHINC), years of education (EDUC, in years) and body mass index (BMI), and fully observed covariates age and gender, subsample MAR was considered plausible for EDUC and BMI, and covariate missingness was considered plausible for HHINC. Thus the above theory was applied with *Y* = (SBP, DBP), *W* = HHINC, *X* = (EDUC, BMI) and *Z* = (AGE, GENDER). The resulting SSIL method consists of applying a ignorable likelihood method to the subsample of cases with HHINC observed.

**Example 5 (ctd). A sample with auxiliary data where the mechanism is MNAR but ignorable.** Here the observed data are shown in Figure 1B, with $D_{\text{obs}} = (D_{\text{resp}}, D_{\text{aux}})$, where $D_{\text{resp}} = \{(y_{1i}, y_{2i}), i = 1,...,r\}$ and $D_{\text{aux}} = \{y_{1j}^*, j = 1,...,n\}$. The probability that $(y_{1i}, y_{2i})$ is observed in the sample is given by Eq. (6). The data is missing not at random according to Rubin's definition since missingness depends on $y_{i1}$, which is missing for the incomplete cases. The joint distribution of $(y_{1i}, y_{2i})$ is factored as in Eq. (8). Let $\mathbb{S}$ denote the set of permutations of the external data $\pi(1,...,n) = (\pi(1),...,\pi(n))$ that map $D_{\text{resp}}$ into the set of respondent values of $Y_1$, in the sense that $y_{1,\pi(i)}^* = y_{1i}, i = 1,...,r$. Let $\|\mathbb{S}\|$ be the size of this set. The observed likelihood is then

$$
\begin{aligned}
L(\theta_1, \theta_2, \phi \mid D_{\text{obs}}, M) &= \text{const.} \times \prod_{i=1}^{r} f_1(y_{1i} \mid \theta_1) f_2(y_{2i} \mid y_{1i}, \theta_2) \left( (1 - g(y_{1i}, \phi)) \right) \\
&\quad \times \sum_{\pi \in \mathbb{S}} \prod_{i=r+1}^{n} f_1(y_{1,\pi(i)}^*) g(y_{1,\pi(i)}, \phi) / \|\mathbb{S}\| \\
&= \text{const.} \prod_{j=1}^{n} f_1(y_{1j}^* \mid \theta_1) \times \prod_{i=1}^{r} f_2(y_{2i} \mid y_{1i}, \theta_2) \times \prod_{i=1}^{r} \left( (1 - g(y_{1i}, \phi)) \right) \times \prod_{j=r+1}^{n} g(y_{1j}, \phi),
\end{aligned}
$$

since each of the $\|\mathbb{S}\|$ permutations has the same probability, and the aggregate of the product from $r+1$ to $n$ is the same for each permutation. Hence the mechanism is $\text{MAR}(\theta)$ and $\text{LIGN}(\theta)$ if $\theta$ and $\phi$ are distinct.

## 4. CONCLUSION

We have proposed definitions of MAR and ignorability for likelihood inference about subsets of model parameters. This is useful since in many problems the primary focus is on a particular parameter or subset of parameters, and weaker conditions suffice for a subset. Our definitions differ slightly from Rubin (1976) when applied to all the model parameters, in that cases like Example 5 can be formulated where the mechanism is MAR and ignorable for all the parameters, but the mechanism is not MAR according to Rubin's definition. This example of auxiliary information is important in survey settings, where auxiliary data is available from external data sources; in the future we plan to extend this example to situations with item nonresponse, and more extensive auxiliary information.

## REFERENCES

Little, R.J.A. (1993). Pattern-Mixture Models for Multivariate Incomplete Data. *Journal of the American Statistical Association*, 88, 125-134. DOI: 10.2307/2290705.

Little, R. J. A., and Rubin, D. B. (2002). *Statistical Analysis with Missing Data (2nd ed.)* New Jersey: John Wiley.

Little, R. J. and Zhang, N. (2011). Subsample ignorable likelihood for regression analysis with missing data. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, 60, 4, 591–605. DOI: 10.1111/j.1467-9876.2011.00763.x

Rubin, D. B. (1976). Inference and Missing Data. *Biometrika* 63, 581-592.

Zhou, Y., Kalbfleisch, J.D. & Little, R.J. (2010). Block-Conditional MAR Models for Missing Data. *Statistical Science*, 25, 4, 517-532. DOI: 10.1214/10-STS344.

Figure 1. Patterns of Missing Data in the Examples

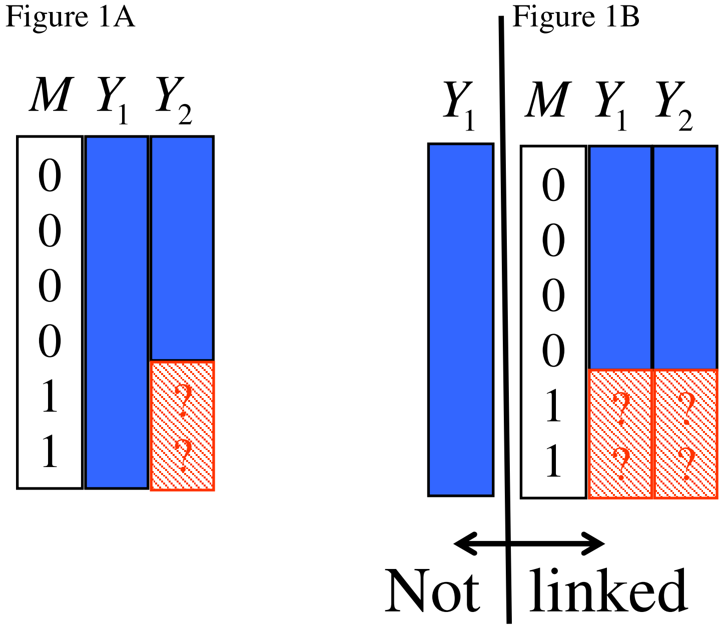

**Figure 2. General Missing Data Structure for Example 4**

| Pattern | Observation, $i$ | $z_i$ | $w_i$ | $x_i$ | $y_i$ | $R_{w_i}$ | $R_{(x_i, y_i)}$ |
|---|---|---|---|---|---|---|---|
| 1 | $i = m+1,\ldots,m+r$ | √ | √ | ? | ? | $u_w$ | $u_{(x,y)}$ or $\bar{u}_{(x,y)}$ |
| 2 | $i = m+r+1,\ldots,n$ | √ | ? | ? | ? | $\bar{u}_w$ | $u_{(x,y)}$ or $\bar{u}_{(x,y)}$ |

Key: √ denotes observed, ? denotes observed or missing