sciendo

Journal of Official Statistics, Vol. 36, No. 3, 2020, pp. 703–728, http://dx.doi.org/10.2478/JOS-2020-0035

Proxy Pattern-Mixture Analysis for a Binary Variable Subject to Nonresponse

Rebecca R. Andridge[1] *and Roderick J.A. Little*[2]

Given increasing survey nonresponse, good measures of the potential impact of nonresponse on survey estimates are particularly important. Existing measures, such as the R-indicator, make the strong assumption that missingness is missing at random, meaning that it depends only on variables that are observed for respondents and nonrespondents. We consider assessment of the impact of nonresponse for a binary survey variable Y subject to nonresponse when missingness may be not at random, meaning that missingness may depend on Y itself. Our work is motivated by missing categorical income data in the 2015 Ohio Medicaid Assessment Survey (OMAS), where whether or not income is missing may be related to the income value itself, with low-income earners more reluctant to respond. We assume there is a set of covariates observed for nonrespondents and respondents, which for the item nonresponse (as in OMAS) is often a rich set of variables, but which may be potentially limited in cases of unit nonresponse. To reduce dimensionality and for simplicity we reduce these available covariates to a continuous proxy variable X, available for both respondents and nonrespondents, that has the highest correlation with Y, estimated from a probit regression analysis of respondent data. We extend the previously proposed proxy-pattern mixture (PPM) analysis for continuous outcomes to the binary outcome using a latent variable approach for modeling the joint distribution of Y and X. Our method does not assume data are missing at random but includes it as a special case, thus creating a convenient framework for sensitivity analyses. Maximum likelihood, Bayesian, and multiple imputation versions of PPM analysis are described, and robustness of these methods to model assumptions is discussed. Properties are demonstrated through simulation and with the 2015 OMAS data.

Key words: Missing data; nonignorable nonresponse; nonresponse bias; survey data; bayesian methods.

1. Introduction

Response rates for large-scale surveys have been steadily declining in recent years (Curtain et al. 2005; Brick and Williams 2013), increasing the need for methods to analyze the impact of nonresponse on survey estimates. Andridge and Little (2011) argue that the assessment of impact depends primarily on three features of the available data: the nonresponse rate, differences between respondents and nonrespondents on characteristics that are observed for the entire sample, and the relationship between these fully observed covariates and the survey outcome of interest. Current methods for handling nonresponse in surveys have tended to focus on a subset of these features, but all three are important.

[1] The Ohio State University College of Public Health Division of Biostatistics, 242 Cunz Hall, 1841 Neil Ave., Columbus, OH 43210, U.S.A. Email: andridge.1@osu.edu
[2] University of Michigan Department of Biostatistics, M4071 SPH II, 1415 Washington Heights, Ann Arbor, MI 48109, U.S.A. Email: rlittle@umich.edu

© Statistics Sweden

In addition, missing-data methods have usually assumed that the missing data are missing at random (MAR, see Rubin 1976), which means that missingness does not depend on missing values in the data, after conditioning on the variables observed for respondents and nonrespondents. This assumption is often questionable, particularly in the case of unit nonresponse, where the set of variables observed for nonrespondents and respondents is limited. Data that are not MAR are called missing not at random (MNAR) or nonignorable. We propose measures to assess the impact of nonresponse for binary survey variables, when data are potentially MNAR.

A limited set of methods have been developed for MNAR nonresponse for categorical outcomes in survey data. Stasny (1991) used a MNAR hierarchical Bayes selection model to study victimization in the National Crime Survey. This work was extended by Nandram and Choi (2002a, 2002b). Similar methods are developed for multinomial outcomes in Nandram et al. (2002) and Nandram et al. (2005). Previous methods for categorical nonresponse have tended to require that auxiliary data are also categorical, but our methods allow auxiliary variables to be continuous, that is, do not require that continuous variables be categorized before inclusion in the model.

Outside of survey applications there have been more recent developments in MNAR nonresponse for binary outcomes. Magder (2003) formulated departures from MAR (towards MNAR) in terms of a response probability ratio, defined as the ratio of the probability of the outcome being observed ("response probability") comparing subjects with and without the outcome. In a similar vein, Hedeker et al. (2007) and Jackson et al. (2014) formulate departures from MAR in terms of odds ratios comparing the probability of success for subjects with and without missing outcomes, and Higgins et al. (2008) used this approach (and referred to it as "informative missingness odds ratios") for meta-analyses. More recently, Liublinska and Rubin (2014) used tipping point displays to perform sensitivity analyses for binary outcomes in two-arm clinical trials.

Little and Rubin (2019) describes five different strategies for handling MNAR problems:

1. Follow up a sample of nonrespondents and incorporate this information into the main analysis.
2. Adopt a Bayesian approach, assigning the parameters prior distributions. Bayesian inference does not generally require that the data provide information for all the parameters, although inferences tend to be sensitive to the choice of prior distribution.
3. Impose additional restrictions on model parameters, such as on the regression coefficients in regressions for missingness or the survey variables.
4. Conduct analysis to assess sensitivity of inferences for quantities of interest to different choices of the values of parameters poorly estimated from the data.
5. Selectively discard data to avoid modeling the missingness mechanism.

Our methods apply approaches (2) and (4) to handle MNAR missingness in binary survey outcomes. The work is motivated by missing income data in the 2015 OMAS (RTI International 2015), though it has broad applicability to missing binary outcomes in survey data. For example, a recent analysis used survey data from the 2016 Behavioral Risk Factor Surveillance System (BRFSS) to estimate the proportion of adults who served as caregivers in the United States, but approximately 10% of survey respondents did not answer the

question about caregiving (Barnhart et al. 2019). The resulting analysis assumed missingness was completely at random; our methods would allow investigation into how estimates would change under varying missingness assumptions, including if caregivers were less likely to respond to the question (MNAR missingness). Other possible applications include estimation of the prevalence of socially undesirable or risky behaviors, as these often are assumed to be MNAR (i.e., people who engage in the behavior are more likely to have missing data on these behaviors). For example, estimating the prevalence of intimate partner violence among pregnant women using survey data from the Pregnancy Risk Assessment Monitoring System (PRAMS), or estimating opioid drug use using survey data from the National Survey on Drug Use and Health (NSDUH).

The 2015 OMAS has been conducted seven times since 1998, and is one of the largest state-sponsored health surveys in the United States. The sampling design was a stratified (by county/sub-county), dual-frame (cell phone and landline) random digit dialing sample of Ohio's non-institutionalized population, with oversampling of certain minority populations. Clusters were defined as a household/ family, and within each cluster one adult was randomly selected to participate. Details on the design and implementation of the 2015 OMAS are available elsewhere (RTI International 2015).

A total of $n = 42{,}876$ adults provided responses to some or all of the survey, but 22.2% of subjects ($n = 9{,}511$) were missing the categorical income variable. Income is particularly important for OMAS, as the primary focus of the survey is to assess the health and health care utilization of Ohio's Medicaid, Medicaid-eligible, and non-Medicaid populations, and Medicaid eligibility is partially determined by income. Having a measure of income was even more important for the 2015 OMAS, since this was the first administration of OMAS after Ohio expanded Medicaid coverage in January 2014. Post-expansion, Medicaid was available for all adults aged 19 through 64, regardless of parental status, with family incomes at or below 138% of the Federal Poverty Level (FPL). Thus, for the 22.2% of survey participants with missing income data, eligibility for Medicaid based on income under the new criteria could not be established.

We extend to categorical outcomes our previously-described proxy pattern-mixture (PPM) analysis for a continuous outcome (Andridge and Little 2011), in order to estimate the potential impact of MNAR nonresponse on the proportion of people at various income levels estimated by 2015 OMAS data. In Section 2 we describe the binary PPM model and in Section 3 we discuss three estimation approaches: maximum likelihood (ML), Bayes, and multiple imputation (MI). In Section 4 we describe the sensitivity of each method to model misspecification and propose modifications to produce more robust estimates. The method is illustrated through simulation in Section 5 and then applied to the 2015 OMAS data in Section 6. Section 7 presents some concluding remarks.

2. Proxy Pattern-Mixture Model for a Binary Outcome

Two main approaches for formulating MNAR models can be distinguished. For unit i in the survey, let Y_i be a survey variable subject to missing data, M_i a missing data indicator with value 1 if Y_i is missing and 0 if Y_i is observed, and Z_i known variables. Assuming independent units, selection models factor the joint distribution of M_i and Y_i as

$$f(M_i, Y_i|Z_i, \theta, \psi) = f(Y_i|Z_i, \theta)f(M_i, |Z_i, Y_i, \psi), \quad (1)$$

where densities are distinguished by their arguments. The first factor characterizes the distribution of Y_i in the population, the second factor models the missingness mechanism, and θ and ψ are parameters. Alternatively, pattern-mixture models factor the joint distribution as

$$f(M_i, Y_i|Z_i, \xi, \omega) = f(Y_i|Z_i, M_i, \xi)f(M_i, |Z_i, \omega), \tag{2}$$

where the first distribution characterizes the distribution of Y_i given Z_i in the strata defined by different patterns of missingness, M_i; the second distribution models the probabilities of the different patterns (Rubin 1977; Little 1993), and ξ and ω are parameters. The selection model formulation was used to characterize MAR in Rubin (1976), and early MNAR models were based on this factorization; see, for example Heckman (1976). However, the selection Equation (1) requires specifying the missingness model via the density $f(M_i|Z_i, Y_i, \psi)$. In contrast, pattern-mixture Equation (2), which we use as the basis of our proposed method, avoids the need for an explicit parametric model for the missingness mechanism. For more discussion comparing these modeling approaches, see, for example Little and Rubin (2019) or Muthen et al. (2011).

Andridge and Little (2011) first introduced the proxy pattern-mixture (PPM) model for assessing the potential for nonresponse bias for continuous outcomes. Their model was based on an assumption of multivariate normality, which is not suitable for categorical survey outcomes. We now extend the normal proxy pattern-mixture model to a binary outcome Y using a latent variable framework. Consider, initially, a simple random sample of size n from an infinite population. Let Y_i, denote the value of a binary survey outcome and $Z_i = (Z_{i1}, Z_{i2}, \ldots, Z_{ip})$ the values of p covariates for unit i in the sample. Only r of the n sampled units provide a response for Y, so observed data consist of (Y_i, Z_i) for $i = 1, \ldots, r$ and Z_i for $i = r + 1, \ldots, n$. This data pattern could be seen with item nonresponse, as in our 2015 OMAS application, where Y is the indicator for being below 138% FPL (missing for 22% of respondents) and Z is a rich set of variables collected in the remainder of the survey. This data structure could also occur with unit nonresponse, where Y is a single survey item and the covariates Z are design variables known for the entire sample, paradata (e.g., interviewer observations), or data available through linkage with administrative registers. Our main interest is assessing the impact of nonresponse on inference for the proportion of units in the population with $Y = 1$.

In order to use the framework of Andridge and Little (2011), we assume that Y is related to a normally distributed latent variable U through the rule that $Y = 1$ when the latent variable $U > 0$. Following the approach of previous PPM models (Andridge and Little 2011; Andridge and Thompson 2015), we reduce the covariates Z to a single proxy variable X that is a linear combination of the Z using a regression model for $Y|Z$. Letting M denote the missingness indicator (i.e., $M = 1$ when Y is missing), we take the regression of Y on Z for the respondents to be a probit regression model given by

$$\Pr(Y = 1|Z, M = 0) = \Phi(\alpha_0 + \alpha Z) \tag{3}$$

where $\Phi(\cdot)$ denotes the standard normal CDF. We then create the proxy X as the linear predictor from the probit regression, $X = \hat{a}_0 + \hat{a}Z$. In doing this, covariates that are not associated with Y (among respondents) will have $\hat{a} \approx 0$ and thus will not contribute to the proxy variable X. Though the $\hat{a}$ are estimated using respondent data only, the proxy X can be

created for nonrespondents as well, since Z is fully observed. The regression coefficients $\{\hat{a}_0, \hat{a}\}$ are subject to sampling error, so in practice X is estimated rather than known.

We now apply the normal PPM model of Andridge and Little (2011) to X, U and M. Specifically, we assume that the joint distribution of $[U, X, M]$ follows the bivariate pattern-mixture model discussed in Little (1994):

$$\begin{aligned} (U, X|M = m) &\sim N_2\left((\mu_u^{(m)}, \mu_x^{(m)}), \Sigma^{(m)}\right) \\ M &\sim Bernoulli(1 - \pi) \\ \Sigma^{(m)} &= \begin{bmatrix} \sigma_{uu}^{(m)} & \rho^{(m)}\sqrt{\sigma_{uu}^{(m)}\sigma_{xx}^{(m)}} \\ \rho^{(m)}\sqrt{\sigma_{uu}^{(m)}\sigma_{xx}^{(m)}} & \sigma_{uu}^{(m)} \end{bmatrix}, \end{aligned} \tag{4}$$

where N_2 denotes the bivariate normal distribution. In this pattern-mixture model, the parameters of the joint distribution of U and X are allowed to differ across patterns, as indicated by the superscript (m).

Of primary interest is the marginal mean of Y, which can be obtained from the pattern-mixturemodel by averaging over the two patterns ($m = 0, 1$). Specifically, we want to estimate

$$\begin{aligned} \mu_y = \Pr(Y = 1) &= \Pr(U > 0) \\ &= \Pr(U > 0|M = 0) \times \Pr(M = 0) + \Pr(U > 0|M = 1) \times \Pr(M = 1) \\ &= \pi\Phi\left(\mu_u^{(0)}/\sqrt{\sigma_{uu}^{(0)}}\right) + (1 - \pi)\Phi\left(\mu_u^{(1)}/\sqrt{\sigma_{uu}^{(1)}}\right). \end{aligned} \tag{5}$$

For parameters of the respondent distribution ($m = 0$), estimates are easily obtained, though since U is a latent variable we cannot estimate both its mean and variance. Without loss of generality we fix the variance at $\sigma_{uu}^{(0)} = 1$ as is conventional for latent variables and estimate the mean $\mu_u^{(0)}$ (for details on estimation methods, see Section 3). For nonrespondents, we can estimate the mean and variance of X, $\{\mu_x^{(1)}, \sigma_{xx}^{(1)}\}$, but there is no information in the data with which to estimate the parameters $\mu_u^{(1)}$, $\sigma_{uu}^{(1)}$, and $\rho^{(1)}$. Thus the model is underidentified.

Importantly, note that $\rho^{(m)}$ is the correlation between X and the latent variable U, not the correlation between X and Y. This correlation between latent U and X is referred to as the biserial correlation between binary Y and continuous X, and is always larger than the Pearson correlation between Y and X. This correlation is defined for both respondents ($\rho^{(0)}$) and nonrespondents ($\rho^{(1)}$), but it is only estimable (without further assumptions) for respondents. We refer to this correlation as the "strength of the proxy", as higher $\rho^{(0)}$ indicates a stronger model for Y, that is, that the covariates Z are more predictive of Y in the probit model for respondents.

In order to identify the model and obtain estimates for the parameters $\mu_u^{(1)}$, $\sigma_{uu}^{(1)}$, and $\rho^{(1)}$ we use parameter restrictions induced by assumptions on the missing data mechanism (Little 1994; Andridge and Little 2011). Specifically, we assume that the probability that $M = 1$ is an unspecified function f of a linear combination of X and U:

$$\Pr(M = 1|U, X) = f((1 - \phi)X^* + \phi U), \tag{6}$$

where $X^* = X/\sqrt{\sigma_{xx}^{(0)}}$ is the proxy rescaled to have unit variance.

Here ϕ is a sensitivity parameter that ranges from 0 to 1 and determines the missingness mechanism. A value of $\phi = 0$ corresponds to missingness being entirely dependent on the proxy X, which is fully observed, implying missingness is at random (MAR). At the other extreme, $\phi = 1$ corresponds to missingness being entirely dependent on U, the completely unobserved latent variable, and thus a missing not at random (MNAR) mechanism. Intermediate values of ϕ allow for "less extreme" MNAR mechanisms, for example, $\phi = 0.5$ which would equally weight the contributions of X and U. For MNAR mechanisms, missingness is a function of the latent variable U, thus allowing for a "smooth" missingness function. In other words, conditional on X the probability of missingness may lie on a continuum instead of only taking two values, as would be the case if missingness depended on Y itself. We note that in previous PPM models the sensitivity parameter was defined differently (as $\lambda \in [0, \infty]$ in Andridge and Little 2011; Andridge and Thompson 2015; Andridge et al. 2017); we elect to use the more recent reparameterization of Little et al. (2019). We note that the sensitivity analyses in the earlier papers using $\lambda = \{0, 1, \infty\}$ correspond to $\phi = \{0, 0.5, 1\}$.

With the missingness assumption in Equation (6), the unidentified parameters can be expressed as functions of the identified parameters and the sensitivity parameter ϕ. The mean and variance of U for nonrespondents (given $M = 1$) are given by

$$\begin{aligned} \mu_u^{(1)} &= \mu_u^{(0)} + \left(\frac{\phi + (1-\phi)\rho^{(0)}}{\phi\rho^{(0)} + (1-\phi)}\right)\frac{\mu_x^{(1)} - \mu_x^{(0)}}{\sqrt{\sigma_{xx}^{(0)}}} \\ \sigma_{uu}^{(1)} &= 1 + \left(\frac{\phi + (1-\phi)\rho^{(0)}}{\phi\rho^{(0)} + (1-\phi)}\right)^2\frac{\sigma_{xx}^{(1)} - \sigma_{xx}^{(0)}}{\sigma_{xx}^{(0)}}. \end{aligned} \tag{7}$$

These formulae follow from the corresponding formulae for the normal proxy pattern-mixture model, with U in place of Y; see Andridge and Little (2011) and the appendix of Sullivan and Andridge (2015) for details on their derivation.

Looking closely at these formulas, we see that the mean and variance for nonrespondents ($\mu_u^{(1)}$, $\sigma_{uu}^{(1)}$) are shifted from the mean and variance for respondents ($\mu_u^{(0)}$,$\sigma_{uu}^{(0)} = 1$), with the amount of shift based on the sensitivity parameter ϕ, the correlation between U and X for respondents ($\rho^{(0)}$), and the differences between the respondent and nonrespondent distributions of X. Larger shifts correspond to larger nonresponse bias for the respondent mean of Y, that is, the overall mean is further from the respondent mean. The size of the bias is also determined by the response rate π, since as shown in Equation (5) the overall mean is a weighted combination of the respondent and nonrespondent means.

A few observations can be made about the formulae in Equation (7). Regardless of ϕ or $\rho^{(0)}$ or π, if the mean and variance of X are the same for respondents and nonrespondents (i.e., $\mu_x^{(0)} = \mu_x^{(1)}$ and $\sigma_{xx}^{(0)} = \sigma_{xx}^{(1)}$), then the mean and variance of U are identical for respondents and nonrespondents. Thus the marginal mean of Y in Equation (5) equals the respondent mean of Y. This property seems reasonable, as if there is no evidence of nonresponse bias in X, which is a proxy for Y, then there is no evidence of nonresponse bias in Y. If there is a difference between respondents and nonrespondents in means or variances of X, then the difference in the respondent and nonrespondent means and

variances of U are functions of the differences for X, the strength of the proxy as measured by $\rho^{(0)}$, and the sensitivity parameter ϕ. The higher the correlation $\rho^{(0)}$, the closer the shift in means and variances for U is to the shift for X, regardless of ϕ. This makes sense, as in the extreme, if X is a perfect proxy for U, that is, $\rho^{(0)} = 1$, then the shift in the mean and variance of U should be exactly equal to the (standardized) shift for X (i.e., the nonresponse bias for U is equal to the nonresponse bias for X). To assess the effect of the sensitivity parameter ϕ, note that as ϕ goes from 0 to 1, we move from an MAR assumption ($\phi = 0$) towards an increasingly strong MNAR assumption. For a given difference in means and variances for X and correlation $\rho^{(0)}$, moving further away from MAR (increasing ϕ towards 1) leads to larger differences between respondent and nonrespondent means of U. This, in turn, leads to larger differences between the respondent mean and the overall mean, that is, a larger nonresponse bias for the overall mean. To help visualize these effects, Subsection 5.1 shows an illustration of how these parameters impact nonresponse bias under this model.

For a given partially missing survey outcome Y and covariates Z, all components of the PPM model will be identified under the missingness assumption in Equation (6) with a specified value of ϕ. Thus, to use our model to assess the sensitivity of the mean of Y to varying assumptions about the missingness mechanism, we specify values of the sensitivity parameter ϕ and estimate the marginal mean of Y for each ϕ value. In order to capture a range of nonignorability, we suggest using the values $\phi = \{0, 0.5, 1\}$, which correspond to MAR ($\phi = 0$), the most "extreme" MNAR ($\phi = 1$) where missingness depends entirely on unobserved U, and an intermediate value ($\phi = 0.5$) that lies between these two extremes. How much or little the estimates vary, along with the size of confidence intervals for the mean, can provide insight into the potential magnitude of nonresponse bias.

3. Estimation Methods

3.1. Maximum Likelihood

Maximum likelihood (ML) estimates for the pattern-mixture model are obtained by substituting estimates for the parameters into Equations (5) and (7). Estimators of $\mu_x^{(m)}$ and $\sigma_{xx}^{(m)}$ for $m = 0,\ 1$ are the usual sample means and covariance matrices for X for respondents and nonrespondents. The estimate for π is the response rate, that his, the proportion of cases with Y observed. ML estimates of the biserial correlation $\rho^{(0)}$ and $\mu_u^{(0)}$ (referred to as the "cutpoint") do not have closed form (Tate 1955a, 1995b) and require an iterative algorithm such as Newton-Raphson. Substituting these estimates into Equations (5) and (7) yields the ML estimate of the mean of Y. The large-sample variance estimate of μ_y is obtained through Taylor series expansion and inversion of the information matrix.

Though quite easy to compute, ML estimation for the PPM model has several drawbacks. First, the resulting ML estimate of $\mu_u^{(0)}$ is not the inverse probit of the respondent mean of Y, and thus the ML estimate of the respondent mean of Y is not $\overline{y}_R$. We propose an alternative approach that avoids this consequence in Subsection 4.1. Additionally, the proxy X is constructed using estimated coefficients from the probit regression of Y on Z, but the ML estimate treats these coefficients as known. A more principled approach is to incorporate this uncertainty with a Bayesian approach, described

below. Finally, incorporation of survey design features, such as weights, in the ML estimates is not straightforward, leading us to consider a multiple imputation approach that enables easily application of design-based inference.

3.2. Bayesian Inference

In order to obtain Bayesian estimates for the parameters of the PPM model, we place non-informative priors on the probit regression parameters α (which includes the intercept) and use a Gibbs sampler to draw the latent U for respondents (Albert and Chib 1993). At the jth iteration of the Gibbs sampler, U follows a truncated normal distribution conditional on Y and $\alpha_{(j)}$ (and the proxy $X_{(j)} = \alpha_{(j)}Z$),

$$(U_{(j)}|Y, \alpha_{(j)}, M = 0) = (U_{(j)}|Y, X_{(j)}, M = 0) \sim N(X_{(j)}, 1) = N(\alpha_{(j)}Z, 1)$$

$$\text{truncated at the left by 0 if } Y = 1 \quad (8)$$

$$\text{truncated at the right by 0 if } Y = 0,$$

where the subscript (j) denotes the jth draws of the parameters. Note that the truncation arises because $Y = 1$ if $U > 0$ and $Y = 0$ otherwise. Then, given the augmented continuous $U_{(j)}$ we draw $\alpha_{(j+1)}$ from its posterior distribution, which also follows a normal distribution,

$$(\alpha_{(j+1)}|Y, U_{(j)}, M = 0) \sim N((Z^TZ)^{-1}Z^TU_{(j)}, (Z^TZ)^{-1}), \quad (9)$$

and recreate the proxy $X_{(j+1)} = \alpha_{(j+1)}\,Z$.

The data augmentation of Y to U allows us to exploit the straightforward Bayesian estimation methods for the normal PPM model. For a chosen value of ϕ, we apply the PPM model algorithm as described in Andridge and Little (2011) to the pair (X, U) to obtain draws of the parameters of the joint distribution of X and U, $\{\mu_x^{(0)}, \sigma_{xx}^{(0)}, \mu_x^{(1)}, \sigma_{xx}^{(1)}, \mu_u^{(0)}, \rho^{(0)}\}$. Since U is unobserved even for the respondents, after each draw of the parameters from the PPM model, X is recreated for the entire sample and U is redrawn for the respondents given the current set of parameter values as described in the data augmentation approach above. Note that this does not require a draw of the latent data for nonrespondents. Draws from the posterior distribution of μ_y are obtained by transforming the draws from the Gibbs sampler as in Equation (7) to obtain $\{\mu_u^{(1)}, \sigma_{uu}^{(1)}\}$, and then substituting these draws into Equation (5).

3.3. Multiple Imputation

An alternative method of inference is multiple imputation (MI) (Rubin 1978). For a selected ϕ we create K complete data sets by filling in missing (binary) Y values with draws from the posterior distribution based on the pattern-mixture model. At the jth draw of the model parameters from their posterior distribution as described in Subsection 3.2 we draw the latent U for nonrespondents based on the conditional distribution,

$$\left[u_{i(j)}|x_{i(j)}, m_i = 1, \phi_{(j)}\right] \sim N\left(\mu_{u(j)}^{(1)} + \frac{\sigma_{ux(j)}^{(1)}}{\sigma_{xx(j)}^{(1)}}\left(x_{i(j)} - \mu_{x(j)}^{(1)}\right), \sigma_{uu(j)}^{(1)} - \frac{\sigma_{ux(j)}^{(1)\,2}}{\sigma_{xx(j)}^{(1)}}\right). \quad (10)$$

The missing y_i are then imputed as $y_{i(j)} = I(u_{i(j)} > 0)$, where $I()$ is an indicator function taking the value 1 if the expression is true. In order to reduce auto-correlation between the imputations due to the Gibbs sampler, we thin the chain for the purpose of creating the imputations. For the kth completed data set the estimate of μ_y is the imputed sample mean $\bar{Y}_k$ with estimated variance W_k. A consistent estimate of μ_y is then given by $\hat{\mu}_y = \frac{1}{K}\sum_{k=1}^{K} \bar{Y}_k$ with $\text{Var}(\hat{\mu}_y) = \bar{W}_K + \frac{K+1}{K} B_K$, where $\bar{W}_K = \frac{1}{K}\sum_{k=1}^{K} W_k$ is the within-imputation variance and $B_K = \frac{1}{K-1}\sum_{k=1}^{K}\left(\bar{Y}_k - \hat{\mu}_y\right)^2$ is the between-imputation variance.

As with the normal PPM analysis, an advantage of the MI approach is that complex design features like clustering, stratification and unequal sampling probabilities can be readily incorporated in the complete-data component of the MI combining rules. Once the imputation process has created complete data sets, design-based methods can be used to estimate μ_y and its variance; for example the design-weighted Horvitz-Thompson estimator can be used to calculate $\bar{Y}_k$. It has been shown that the multiple imputation variance estimator can be biased when applied to data from a complex sample survey unless the sample weights are included as a predictor in the imputation (Kim et al. 2006), and thus we include the sample weights as part of Z when performing imputation with the binary PPM model. For stratified designs, strata can be included as a set of indicator variables in Z.

4. Reducing Sensitivity to Violations of Normality

The normal PPM model is relatively robust to departures from normality, since it relies on first and second moments in estimating the mean of Y. However, for binary outcomes the normality of X plays a more crucial role, as can be seen in the following example. Under the latent variable framework, the conditional distribution of U given X for respondents is normal by definition. Thus, if X has a normal distribution, the marginal distribution of U is also normal, and $\mu_y^{(0)} = \Pr(U > 0|M = 0) = \Phi(\mu_u^{(0)})$. Substituting the ML estimate for $\mu_u^{(0)}$ into the expression for $\mu_y^{(0)}$yields an unbiased estimate of the mean of Y. Suppose, however, that X does not have a normal distribution among the respondents. The conditional distribution $[U\,|X, M = 0]$ remains normal (by definition), but the marginal distribution $[U\,|M = 0]$ is no longer normal. As a consequence, the ML estimate of $\mu_u^{(0)}$ is unbiased, but the ML estimate of $\mu_y^{(0)}$ obtained as $\hat{\mu}_y^{(0)} = \Phi(\hat{\mu}_u^{(0)})$ is biased. The transformation from U to Y gives $\Pr(Y = 1|M = 0) = \Pr(U > 0|M = 0) = \int_0^\infty f_U(u)du$ where $f_U^{(0)}(u)$ is the convolution of the normal distribution of $[U\,|X, M = 0]$ and the non-normal distribution of $[X|M = 0]$. Thus, the ML estimate of the respondent mean of Y is biased, despite the fact that Y is fully observed for the respondents. The Bayesian approach produces biased estimates of the respondent mean for the same reason; the transformation of draws of $\mu_u^{(0)}$ to draws of $\mu_y^{(0)}$ produces biased results.

Both the ML and Bayesian estimation approaches will produce increasingly biased estimates of $\mu_y^{(0)}$ the further X deviates from normality. MI is less sensitive to departures from normality since imputations are based on the conditional distribution $[U|X, M]$ which is normal by definition of the latent variable and is not affected by non-normal X. In the next section, we describe modifications to the ML and Bayesian methods that are robust against deviations from normality of the proxy X.

4.1. Robust Estimation Methods

To reduce sensitivity to non-normal X we use the *two-step method* proposed by Olsson et al. (1982) to estimate the biserial correlation coefficient, $\rho^{(0)}$. In the first step, the cutpoint$\mu_u^{(0)}$ is estimated by $\hat{\mu}_u^{(0)} = \Phi^{-1}(\bar{y}_R)$, so that the estimate of the respondent mean of Y is $\bar{y}_R$. Then a conditional ML estimate of $\rho^{(0)}$ is computed, given $\hat{\mu}_u^{(0)}$. ML estimates of the other parameters of the PPM model are computed as before. This method is computationally simpler than ML, has the attractive property of returning the natural estimate $\hat{\mu}_y^{(0)} = \bar{y}_R$, and is less sensitive to non-normality of the proxy. We estimate the variance using the bootstrap. We refer to this method as Modified ML (MML).

We modify the Bayesian method to reduce sensitivity to non-normal X by applying the Gibbs sampler to draw the latent U for nonrespondents conditional on X and the current parameter values at each iteration, which is not required for the standard Bayesian approach (but is part of the MI approach, see Equation (10)). Draws of the nonrespondent mean of Y, $\mu_y^{(1)}$, are then taken to be $\mu_y^{(1)} = \frac{1}{n-r}\sum_{i=r+1}^{n} I(U_i > 0)$, instead of transformations of the parameters $\{\mu_x^{(0)}, \sigma_{xx}^{(0)}, \mu_x^{(1)}, \sigma_{xx}^{(1)}, \mu_u^{(0)}, \sigma_{uu}^{(0)}, \rho^{(0)}\}$ as given in Equations (7) and (5). A similar modification for the respondent mean is not appropriate, as draws of U for the respondents in the Gibbs sampler are conditional on the observed Y (as in Equation (8)) and thus the resulting draw of $\mu_u^{(0)}$ will always be $\bar{y}_R$. To avoid this, we propose two approaches. An obvious extension is to draw U *without* conditioning on Y, and condition only on the current draws of the proxy X and the parameters. With the subsequent draw of $\mu_y^{(0)}$ taken to be $\mu_y^{(0)} = \frac{1}{n-r}\sum_{i=r+1}^{n} I(U_i > 0)$, the draws of $\mu_y^{(0)}$ will not always be $\bar{y}_R$. To distinguish this method from the unmodified posterior distribution intervals (PD), we call this method Modification 1: PD-redraw. The drawback of this method is that variances may actually be *overestimated* since we are essentially imputing the observed binary outcome Y for the respondents. An alternative approach is to not use the latent U, but instead directly use the draws of the proxy X to calculate predicted probabilities from the probit model. The average of these predicted probabilities for the respondents can then be taken as a draw of the respondent mean, that is, $\mu_y^{(0)} = \frac{1}{r}\sum_{i=1}^{r} \Phi^{-1}(X_i)$. This is actually a draw of the conditional mean of Y (conditional on X) and so its posterior distribution will underestimate the variance of $\mu_y^{(0)}$. To combat this, we take a simple random sample with replacement of the X_i before calculating the mean of the predicted probabilities (Modification 2: PD-predprob).

5. Simulation Studies

We now describe a set of simulation studies designed to (1) illustrate the effects of proxy strength $\rho^{(0)}$, differences in the mean of X between respondents and nonrespondents, and sample size on PPM model estimates of the mean of a binary outcome Y, (2) assess confidence coverage of ML, Bayes and MI inferences when model assumptions are met, and (3) assess confidence coverage of the robust estimation methods when the normality assumption is incorrect. All simulations and data analysis were performed using the software package R (R Core Team 2017). Code for implementing the PPM model and example analyses are available at http://github.com/randridge/PPMA.

5.1. Numerical Illustration of Binary PPM Analysis

We first illustrate the taxonomy of evidence concerning bias based on the strength of the proxy ($\rho^{(0)}$) and the standardized difference between the mean of the proxy for the entire sample and the mean for respondents, $\delta^* = \left(\mu_x - \mu_x^{(0)}\right)/\sqrt{\sigma_{xx}^{(0)}}$. We created a total of eighteen artificial data sets in a 3x3x2 factorial design with a fixed nonresponse rate of 50%. A single data set was generated for each combination of $\rho^{(0)} = \{0.8, 0.5, 0.2\}$, $\delta^* = \{0.1, 0.3, 0.5\}$, and $n = \{100, 400\}$ (with corresponding $r = \{50, 200\}$). A single covariate Z was generated for both respondents and non-respondents, with $z_i \sim N(0, 1)$, $i = 1, \ldots, r$ for respondents and $z_i \sim N\ (\delta^*/(1-r/n),\ 1),\ i = r+1, \ldots, n$ for nonrespondents. For respondents only, a latent variable u_i was generated as $[u_i|z_i] \sim N(a_0 + a_1 z_i,\ 1)$, with an observed binary Y then created as $y_i = 1$ if $u_i > 0$. We set $a_1 = \rho^{(0)}/\sqrt{1-\rho^{(0)^2}}$ so that $\text{Corr}(U,X|M=0) = \rho^{(0)}$ where the proxy $X = a_0 + a_1 Z$. We chose $a_0 = \Phi^{-1}(0.3)\sqrt{1+a_1^2}$ so that the expected value of Y for respondents was 0.3. In this and all subsequent simulations, the latent variable U was used for data generation and then discarded; only Y and Z were used for the proxy pattern-mixture analysis.

For each of the eighteen data sets, estimates of the mean of Y and its variance were obtained using the PPM model for $\phi = (0, 0.5, 1)$. For each value of ϕ, three 95% intervals were calculated:

1. ML: the ML estimate ± 2 standard errors (large-sample approximation),
2. PD: the posterior median and 2.5th to 97.5th posterior interval based on 2000 cycles of the Gibbs sampler as outlined in Subsection 3.2, with a burn-in of 20 iterations,
3. MI: mean ± 2 standard errors from 50 multiply imputed data sets, with a burn-in of 20 iterations and imputing on every hundredth iteration of the Gibbs sampler.

The robust estimation methods described in Subsection 4.1 designed to handle non-normal proxies were also calculated. Since the simulated covariate data were normally distributed, the modified estimators yield similar results and are not shown. The complete case estimate, that is, the respondent mean, was also computed (± 2 standard errors) for each data set.

5.1.1. Results

Figure 1 shows the resulting 95% intervals using each of the three estimation methods for the nine data sets with n=400, plotted alongside the respondent mean (complete case). The relative performances of each method for the data sets with $n = 100$ are similar to the results with $n = 400$ (with larger interval lengths); results are not shown. We note that in this simulation the true mean of Y is not known (indeed, we did not simulate Y for nonrespondents); we simply illustrate the effect of various values of proxy strength ($\rho^{(0)}$) and proxy mean deviation (δ^*) on the sensitivity analysis and compare the different estimation methods.

For populations with strong proxies ($\rho^{(0)} = 0.8$), ML, PD, and MI give nearly identical results. For these populations, there is not a noticeable increase in the length of the intervals as we move from $\phi = 0$ to $\phi = 1$, suggesting that even in the case of a large deviation ($\delta^* = 0.5$) there is good information to correct the potential bias.

For weaker proxies, we see differences among the three methods. When $\phi = 0$ (MAR) the three methods yield similar inference, but for MNAR mechanisms the intervals for PD

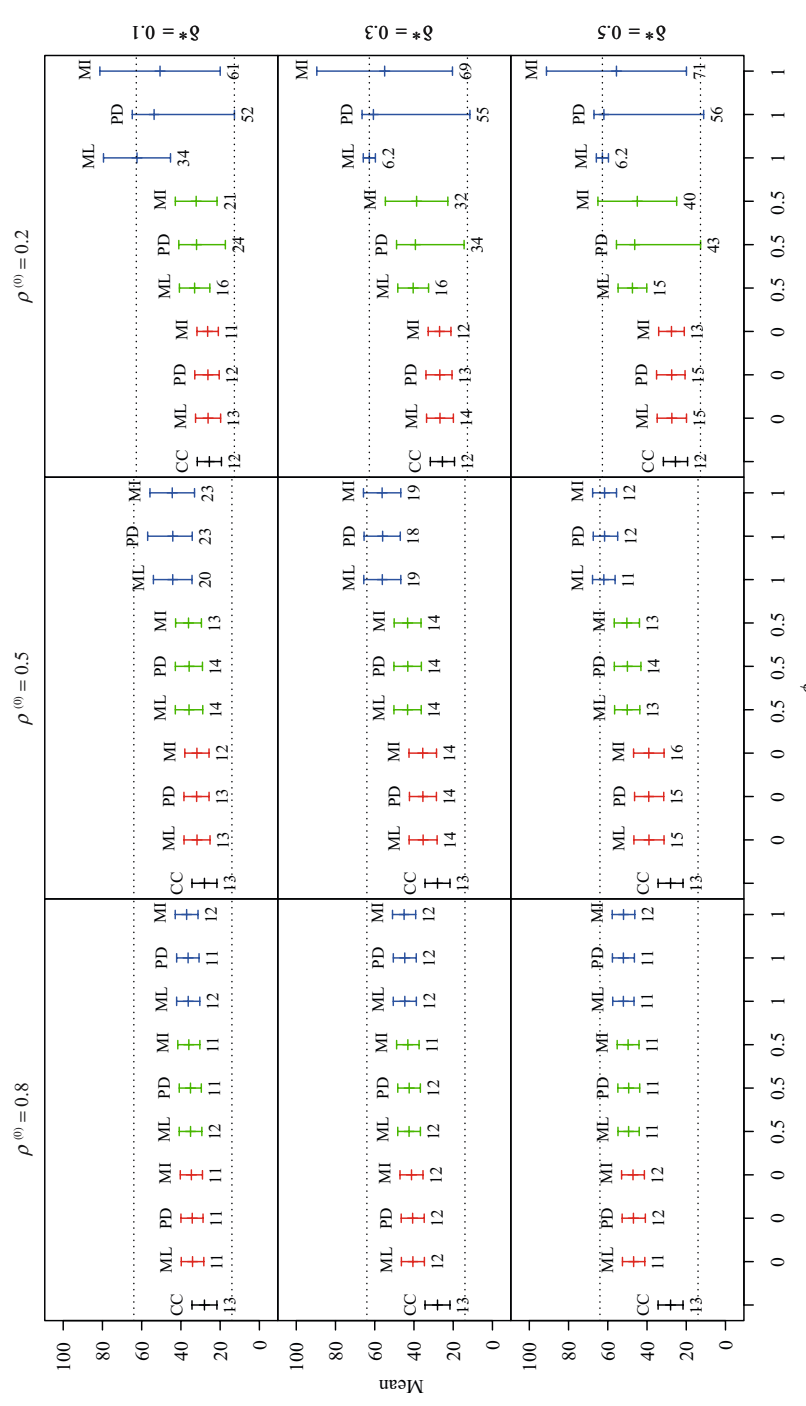


Fig. 1. 95% intervals for nine generated data sets ($n = 400$) for $\phi = (0, 0.5, 1)$. Numbers below intervals are the interval length. Dotted lines are the point estimates obtained by filling in all ones or all zeros for missing values. CC: Complete cases; ML: Maximum likelihood; PD: Posterior distribution; MI: 50 multiply imputed data sets.

and MI tend to be wider than those for ML. For both Bayesian methods (PD, MI) the interval width increases as we move from $\phi = 0$ to $\phi = 1$, with a marked increase in length when $\rho^{(0)} = 0.2$. The ML inference displays different behavior; its intervals actually get very small for the weak proxies and large d^*. This is caused by the unstable behavior of the ML estimate near the boundary of the parameter space. For weak proxies (small $\rho^{(0)}$), the ML estimate of $\sigma^{(1)}_{uu}$ can be zero if the nonrespondent proxy variance is smaller than the respondent variance (see Equation (7)). Since the ML estimate of the mean of Y is given by $\hat{\mu}^{(1)}_y = \Phi\left(\hat{\mu}^{(1)}_u / \sqrt{\hat{\sigma}^{(1)}_{uu}}\right)$, a zero value for $\hat{\sigma}^{(1)}_{uu}$ causes $\hat{\mu}^{(1)}_y$ to be exactly 0 or 1 depending on the sign of $\hat{\mu}^{(1)}_u$. The large sample variance will then be small since the estimate of $\sigma^{(1)}_{uu}$ is zero, and interval widths will be small relative to the PD or MI intervals.

Since the outcome is binary, we obtain a natural upper and lower bound for the mean of Y by filling in all missing values with zeros or all with ones. These bounds are sometimes referred to as Manski bounds (Manski 2016) and are shown with dotted lines in Figure 1. For strong proxies, even with a large deviation this upper bound is not reached, suggesting that the overall mean would not be this extreme even in the worst-case MNAR scenario, where missingness depends entirely on Y through U. However, for the weakest proxy ($\rho^{(0)} = 0.2$) we see that even for the smallest deviation ($\delta^* = 0.1$) the intervals for PD and MI cover or nearly cover these bounds. This is due to the weak information about Y contained in the proxy. The PD intervals are highly skewed and the MI intervals are exaggerated in length. The posterior distribution of μ_y is bimodal, with modes at each of the two bounds obtained when missing values are all zeros or all ones. Thus the posterior interval essentially covers the entire range of possible values of μ_y. Similarly, for MI the imputed data sets have imputed values that are either all zeros or all ones. This causes very large variance and thus large intervals, and since by construction the intervals are symmetric for MI, they are even larger than the posterior intervals from PD. As previously discussed, the ML method gives extremely small intervals for the weak proxies, with the point estimate at the upper bound.

5.2. *Confidence Coverage, Normally Distributed Proxy*

We now assess coverage properties for each of the three estimation methods when the PPM model is correct. We generated data using the same set-up as Subsection 5.1. We fixed $\delta^* = 0.3$ and varied $\rho^{(0)} = \{0.8, 0.5, 0.2\}$ and $n = \{100, 400\}$ for a total of six scenarios. We generated 1,000 replicate data sets for each scenario and applied the binary PPM model using each of $\phi = \{0, 0.5, 1\}$ to each one, assuming the value of ϕ is the true value. This led to a total of eighteen hypothetical populations, and for each we computed the actual coverage of a nominal 95% interval and median interval length. We also calculated the relative empirical bias for each estimator. Treating the assumed value of ϕ as correct is unrealistic, but coverages are clearly not valid when the value of ϕ is misspecified, and uncertainty in choice of ϕ is captured by the sensitivity analysis.

A total of six estimators for the mean of the binary outcome Y and its variance were obtained for each replicate data set. These included the unmodified ML, unmodified Bayesian (PD), and MI estimators, as well as the three modified estimators described in Subsection 4.1: the modified ML estimator (MML) and two modifications to the

Bayesian estimator (PD-redraw, PD-predprob). Confidence intervals for the modified ML estimator were based on 500 bootstrap samples. Posterior intervals for all three PD methods were based on 1,000 draws from the Gibbs sampler as the chains were quick to converge.

5.2.1. Results

Table 1 displays the average empirical relative bias, nominal coverage, and median CI width for the eighteen populations. For the smaller sample size ($n = 100$; Table 1a), all methods suffer from slight undercoverage, even when the proxy is strong. Undercoverage is worst in the populations with the weakest proxy ($\rho^{(0)} = 0.2$) and with $\phi = 1$, where all the methods are negatively biased. With 50% nonresponse, these small samples have only 50 observed data points, and estimation of the distribution of the latent variable is challenging. With this small sample size, confidence intervals are slightly wider for the modified ML than for unmodified ML, and thus coverage is better for the modified ML procedure. No method displays consistently better performance for the small sample size, though the larger interval lengths of PD-redraw and MI yield slightly closer to nominal coverage.

More differences between the methods emerge with the larger sample size ($n = 400$; Table 1b). All methods perform well when the proxy is strong ($\rho^{(0)} = 0.8$), though the second modification to the Bayesian method that resamples the predicted probabilities (PD-predprob) shows a small amount of undercoverage. As expected, the interval widths for the first modification to the Bayesian method (PD-redraw) are wider than the unmodified Bayesian method (PD), with PD-redraw actually overcovering for several populations, most notably when $\phi = 0.5$ or 1 and with weaker proxies. As was evident in the previous simulation, when $\rho^{(0)} = 0.2$ and $\phi = 1$ the confidence interval length for the standard ML procedure is much smaller than any of the other methods, and this leads to undercoverage. In this worst-case scenario, the modified ML, PD-redraw, and MI procedures have nominal coverage despite having some bias and have the largest median confidence interval widths.

5.3. Confidence Coverage, Non-Normally Distributed Proxy

The final objective of the simulation study was to assess the performance of the methods when the normality assumption of the proxy was violated. Complete data were generated and missing values created using a selection model with a variety of missing data mechanisms. The sample size was fixed at $n = 400$, and three different distributions for a single covariate Z were selected: (a) Normal(0, 1), (b) Gamma(4, 0.5), (c) Exponential(1). These distributions were chosen to evaluate the effect of both moderate skew (Gamma) and severe skew (Exponential). We note that the selection model implies marginal normality, whereas the PPM model assumes conditional normality, so even with a normally distributed covariate the distributional assumptions of the PPM model are violated.

Data were generated as follows. For each of the three Z distributions the covariate z_i, $i = 1,..., n$ was generated. Then, for each of $\rho = \{0.8, 0.5, 0.2\}$ the latent u_i was generated from $[u_i|z_i] \sim N(a_0 + a_1 z_i, 1)$, with $a_1 = \rho/\sqrt{1-\rho^2}$ so that $\text{Corr}(U, X) = \rho$, where the

Table 1. Average relative empirical bias, 95% interval coverage and median interval length for eighteen artificial populations with $\delta* = 0.3$ and $\rho^{(0)} = \{0.8, 0.5, 0.2\}$ for (a) $n = 100$ and (b) $n = 400$. ML: Maximum likelihood; PD: Posterior distribution (Bayesian method); PD-redraw: Modification 1 to PD; PD-predprob: Modification 2 to PD; MI: 20 multiply imputed data sets. Results over 1,000 replicates.

(a) $n = 100$

		$\rho^{(0)} = 0.8$			$\rho^{(0)} = 0.5$			$\rho^{(0)} = 0.2$		
ϕ	Method	Relative bias (%)	Coverage (%)	CI width	Relative bias (%)	Coverage (%)	CI width	Relative bias (%)	Coverage (%)	CI width
0	ML	−0.4	**92.9**	0.24	0.4	**92.7**	0.27	0.5	**93.4**	0.27
	Modified ML	−0.1	94.5	0.26	0.5	**93.2**	0.28	0.5	93.8	0.28
	PD	0.3	**93.3**	0.23	1.2	93.7	0.26	1.7	94.2	0.26
	PD-redraw	0.1	**93.4**	0.24	1.0	96.3	0.30	1.5	*97.2*	0.31
	PD-predprob	−0.2	**90.9**	0.22	0.7	**91.9**	0.25	1.2	**93.3**	0.25
	MI	−0.3	**92.2**	0.25	0.6	**93.4**	0.27	1.3	93.8	0.27
0.5	ML	−0.5	**93.6**	0.24	0.0	**92.2**	0.28	−6.7	**80.4**	0.29
	Modified ML	−0.1	94.5	0.25	0.1	**92.6**	0.28	−6.7	**81.5**	0.29
	PD	−0.3	93.7	0.24	−0.5	94.8	0.29	−9.6	95.4	0.37
	PD-redraw	−0.5	**93.0**	0.25	−0.8	*96.5*	0.33	−9.9	*97.3*	0.41
	PD-predprob	−0.6	**92.1**	0.23	−0.8	**93.1**	0.28	−9.7	95.0	0.36
	MI	−0.6	93.8	0.25	−1.2	**93.1**	0.30	−10	**93.2**	0.41
1	ML	−0.3	**93.3**	0.26	0.8	**76.2**	0.37	−18	**75.4**	0.30
	Modified ML	0.1	94.7	0.29	0.8	**93.4**	0.43	−18	**82.2**	0.65
	PD	−0.5	93.8	0.26	−4.3	95.3	0.36	−25	**81.8**	0.49
	PD-redraw	−0.7	93.9	0.27	−4.1	*96.6*	0.38	−25	**86.2**	0.52
	PD-predprob	−0.7	**92.0**	0.26	−4.1	94.7	0.36	−25	**82.4**	0.50
	MI	−0.4	**93.6**	0.28	−5.1	93.9	0.38	−27	**86.6**	0.60

Bolded values are below 1.96 simulation standard errors.
Italicized values are above 1.96 simulation standard errors.

Table 1. Continued.

(b) $n = 400$

		$\rho^{(0)} = 0.8$			$\rho^{(0)} = 0.5$			$\rho^{(0)} = 0.2$		
ϕ	Method	Relative bias (%)	Coverage (%)	CI width	Relative bias (%)	Coverage (%)	CI width	Relative bias (%)	Coverage (%)	CI width
0	ML	0.0	**93.3**	0.12	−0.1	95.4	0.14	0.0	95.7	0.14
	Modified ML	0.1	96.2	0.13	−0.1	95.3	0.14	0.0	95.7	0.14
	PD	0.1	95.3	0.12	0.1	94.6	0.13	0.3	95.1	0.13
	PD-redraw	0.0	95.4	0.13	0.0	*97.7*	0.15	0.3	*98.6*	0.16
	PD-predprob	0.0	**93.4**	0.11	−0.1	93.9	0.13	0.2	94.0	0.13
	MI	−0.1	95.1	0.13	0.0	95.3	0.14	0.2	95.4	0.14
0.5	ML	0.0	94.8	0.12	−0.1	95.1	0.14	−0.6	94.0	0.15
	Modified ML	0.1	95.4	0.13	−0.1	95.6	0.14	−0.6	94.5	0.15
	PD	0.0	95.1	0.12	−0.1	95.7	0.14	−1.3	*97.1*	0.21
	PD-redraw	−0.1	95.5	0.13	−0.2	*97.8*	0.16	−1.4	*98.5*	0.23
	PD-predprob	−0.1	94.1	0.11	−0.2	94.9	0.14	−1.3	*96.8*	0.21
	MI	−0.1	95.1	0.13	−0.3	95.8	0.15	−3.0	*97.2*	0.23
1	ML	0.0	94.4	0.13	1.2	**92.1**	0.22	−5.0	**90.7**	0.16
	Modified ML	0.1	95.3	0.13	1.3	**92.9**	0.21	−5.0	94.7	0.38
	PD	0.0	94.7	0.13	0.1	95.8	0.20	−8.3	**91.0**	0.32
	PD-redraw	−0.1	94.8	0.13	0.3	*96.7*	0.21	−8.2	93.9	0.33
	PD-predprob	−0.1	**93.1**	0.12	0.2	95.6	0.19	−8.2	**90.8**	0.32
	MI	0.0	94.9	0.13	0.4	94.6	0.21	−11	95.3	0.37

Bolded values are below 1.96 simulation standard errors.
Italicized values are above 1.96 simulation standard errors.

Table 2. Average relative empirical bias, 95% interval coverage and median interval length for eighteen artificial populations with n = 400 with covariate distributions (a) Normal, (b) Gamma, and (c) Exponential. ML: Maximum likelihood; PD: Posterior distribution (Bayesian method); PD-redraw: Modification 1 to PD; PD-predprob: Modification 2 to PD; MI: 20 multiply imputed data sets. Results over 1000 replicates.

(a) $Z \sim$ Normal(0,1)

		MAR $\Pr(M = 1\|Z, U) = f(Z)$			MNAR $\Pr(M = 1\|Z, U) = f(U)$		
ρ	Method	Relative bias (%)	Coverage (%)	CI width	Relative bias (%)	Coverage (%)	CI width
0.8	ML	−0.1	*96.4*	0.12	0.3	95.8	0.13
	Modified ML	0.0	96.2	0.12	0.4	96.1	0.14
	PD	0.3	95.9	0.12	0.3	95.6	0.13
	PD-redraw	0.1	96.3	0.12	−0.1	95.6	0.14
	PD-predprob	0.0	94.9	0.11	0.0	94.6	0.13
	MI	0.0	96.2	0.12	0.3	95.7	0.14
0.5	ML	0.0	95.0	0.13	0.5	**91.7**	0.28
	Modified ML	0.0	95.0	0.13	0.6	*96.6*	0.31
	PD	0.4	94.5	0.13	1.6	94.8	0.24
	PD-redraw	0.2	*97.6*	0.15	1.5	*96.5*	0.25
	PD-predprob	0.1	**93.1**	0.12	1.5	93.8	0.24
	MI	0.1	95.2	0.13	1.5	94.9	0.25
0.2	ML	−0.1	94.9	0.13	4.3	**64.7**	0.49
	Modified ML	−0.1	95.1	0.14	4.4	96.2	0.54
	PD	0.2	94.2	0.13	6.5	*98.7*	0.34
	PD-redraw	0.1	*98.6*	0.16	6.4	*99.1*	0.35
	PD-predprob	0.1	**93.2**	0.12	6.4	*98.6*	0.34
	MI	0.1	94.7	0.14	5.8	*97.7*	0.36

Bolded values are below 1.96 simulation standard errors.
Italicized values are above 1.96 simulation standard errors.

proxy $X = a_0 + a_1 Z$. Note that in this simulation ρ is the correlation in the entire sample, not just among respondents. The binary outcome Y was then created as $y_i = 1$ if $u_i > 0$, with values of a_0 chosen so that $E[Y] = 0.3$. The missing data indicator m_i was generated according to a logistic model,

$$\text{logit}(\Pr(m_i = 1|u_i, z_i)) = \gamma_0 + \gamma_Z z_i + \gamma_U u_i, \tag{11}$$

and values of y_i were deleted when $m_i = 1$. The two different missingness mechanisms selected were MAR, with $\gamma_Z = 0.5$, $\gamma_U = 0$, and extreme MNAR, with $\gamma_Z = 0$, $\gamma_U = 0.5$. Aside from the discrepancy of marginal versus conditional normality, these two mechanisms correspond to ϕ values of 0 and 1, respectively. For both scenarios, values of γ_0 were selected to induce approximately 50% missingness.

The process of generating $\{z_i, u_i, y_i, m_i\}$, and inducing missingness was repeated 1,000 times for each of the eighteen populations (3 Z distributions × 3 ρ values × 2 missingness mechanisms). The same six estimators for the mean of the binary outcome Y and its variance were obtained for each of the eighteen data sets as in the previous simulation. For the MAR mechanism, ϕ was taken to be zero, and for MNAR $\phi = 1$.

Table 2. Continued.

(b) $Z \sim$ Gamma(4,0.5)

		MAR $\Pr(M = 1 \mid Z, U) = f(Z)$			MNAR $\Pr(M = 1 \mid Z, U) = f(U)$		
ρ	Method	Relative bias (%)	Coverage (%)	CI width	Relative bias (%)	Coverage (%)	CI width
0.8	ML	7.4	**87.7**	0.13	11	**79.6**	0.12
	Modified ML	2.3	94.2	0.12	9.3	**84.3**	0.12
	PD	7.8	**86.1**	0.12	11	**78.9**	0.12
	PD-redraw	0.1	95.5	0.12	4.0	**93.5**	0.12
	PD-predprob	0.0	**92.9**	0.11	4.0	**91.2**	0.11
	MI	−0.1	94.7	0.12	4.2	95.3	0.13
0.5	ML	2.1	94.2	0.13	7.7	**86.7**	0.18
	Modified ML	0.8	95.2	0.13	6.8	**90.1**	0.21
	PD	2.5	**92.8**	0.13	8.1	**86.8**	0.17
	PD-redraw	0.4	*97.7*	0.15	4.9	**92.2**	0.18
	PD-predprob	0.3	**92.5**	0.12	4.9	**89.3**	0.17
	MI	0.2	94.5	0.13	5.2	**92.6**	0.19
0.2	ML	0.3	95.5	0.14	−2.2	**64.8**	0.33
	Modified ML	0.1	95.4	0.14	−2.3	94.2	0.46
	PD	0.6	94.8	0.13	4.8	*97.5*	0.30
	PD-redraw	0.3	*98.7*	0.16	3.8	*98.4*	0.31
	PD-predprob	0.2	93.7	0.12	3.9	*97.6*	0.30
	MI	0.4	94.8	0.14	4.0	96.2	0.32

Bolded values are below 1.96 simulation standard errors.
Italicized values are above 1.96 simulation standard errors.

5.3.1. Results

When Z is normally distributed, results are similar to the previous simulation, as seen in Table 2a. All methods have negligible bias across all scenarios, except when $\rho = 0.2$ under MNAR. For this population, there is a small bias but all methods except unmodified ML still achieve nominal coverage, and in fact many show higher than nominal coverage. The consistently best performing methods are Modified ML, PD-redraw, and MI, which achieve close to nominal coverage in all scenarios. PD-predprob shows undercoverage when the proxy is strong under MAR. As was previously seen, the unmodified ML procedure has intervals that are too short under MNAR with a weak proxy ($\rho = 0.2$), and thus exhibits very poor coverage. Modified ML fixes this problem, since the bootstrap is used for variance estimation instead of the large-sample approximation, though the intervals are more than 50% longer than with other methods.

Table 2b shows results for the slightly skewed proxy, when Z has a Gamma distribution. In general, all estimation methods perform better with weaker proxies when the covariate is skewed. The methods that rely the most on the underlying normality assumption of the PPM analysis, unmodified ML and PD, show bias for the stronger proxies under both missingness mechanisms and hence tend to undercover. When missingness is at random, as

Table 2. Continued.

(c) $Z \sim$ Exponential(1)

		MAR $\Pr(M = 1 \mid Z, U) = f(Z)$			MNAR $\Pr(M = 1 \mid Z, U) = f(U)$		
ρ	Method	Relative bias (%)	Coverage (%)	CI width	Relative bias (%)	Coverage (%)	CI width
0.8	ML	15	**71.7**	0.13	20	**42.0**	0.11
	Modified ML	4.5	**91.8**	0.12	16	**60.8**	0.11
	PD	15	**69.6**	0.13	20	**42.3**	0.11
	PD-redraw	−0.6	94.4	0.12	4.6	**91.5**	0.11
	PD-predprob	−0.6	**90.9**	0.10	4.7	**88.1**	0.10
	MI	−0.7	94.2	0.12	4.8	**93.5**	0.12
0.5	ML	3.1	93.9	0.14	13	**71.5**	0.14
	Modified ML	0.9	94.7	0.13	11	**83.0**	0.19
	PD	3.6	**93.2**	0.13	13	**71.5**	0.14
	PD-redraw	−0.8	*96.7*	0.14	5.4	**93.2**	0.15
	PD-predprob	−0.9	**92.7**	0.12	5.4	**89.2**	0.14
	MI	−0.9	94.0	0.13	5.6	**93.3**	0.16
0.2	ML	−0.6	95.0	0.14	0.2	**66.0**	0.24
	Modified ML	−0.8	94.9	0.14	0.1	94.5	0.41
	PD	−0.1	95.1	0.13	5.7	95.0	0.26
	PD-redraw	−0.9	*98.1*	0.15	3.2	*97.2*	0.27
	PD-predprob	−0.9	**93.6**	0.12	3.3	96.3	0.26
	MI	−0.9	94.5	0.13	3.4	96.0	0.28

Bolded values are below 1.96 simulation standard errors.
Italicized values are above 1.96 simulation standard errors.

before the best performers are Modified ML, PD-redraw, and MI, with some overcoverage by PD-redraw. The other PD modification (PD-predprob) slightly undercovers for stronger proxies. The more difficult populations are under MNAR. For both $\rho = 0.8$ and $\rho = 0.5$ all methods exhibit some bias, with the unmodified ML and PD methods showing the most bias. Subsequently, almost all methods fail to achieve nominal coverage. The exception is MI, which is at nominal coverage for all but one scenario. Under MNAR, for the weakest proxy ($\rho = 0.2$) the unmodified ML again shows undercoverage, while Modified ML corrects this problem. However, it does so with very large confidence intervals relative to the PD and MI methods, which reach nominal coverage.

Results for Z having an Exponential distribution are displayed in Table 2c. The results are similar to the Gamma case, with larger biases and lower coverage rates across all populations. For both mechanisms, coverage actually increases for all methods as ρ decreases (the exception being ML under MNAR with $\rho = 0.2$). Under MNAR, it is difficult for any estimation method to reach nominal coverage except when the proxy is weak. In that scenario, the modified ML, both modified PD, and MI methods reach nominal coverage, while the unmodified ML and PD methods remain biased and have poor coverage. Of course, confidence intervals are very wide, especially Modified ML, which has nearly 50% longer intervals.

Overall, the best performing method is MI, which achieves nominal or just under nominal coverage for all three distributions of Z, including the severely skewed Exponential, and under both missingness mechanisms with all strengths of proxies. This result is not unexpected. Even though MI uses the fully parametric PPM model to generate posterior draws of the parameters, these draws are subsequently used to impute the missing Y values via the conditional distribution of $[U|X, M = 1]$. Even if the proxy is not normally distributed, the conditional distribution of the latent variable given the proxy is normal by definition, and so MI is the least sensitive to departures from normality in the proxy.

The one other method that does reasonably well in most scenarios is the first modification to the Bayesian draws, PD-redraw. As with MI, this method conditions on the proxy and draws the latent U and thus outperforms the unmodified Bayesian method that relies entirely on the joint normality of U and the proxy X. PD-redraw achieves at or near nominal coverage for strong proxies across all levels of skewness, but exhibits overcoverage for weaker proxies. This is to be expected, since in this modification the latent U for respondents are redrawn unconditional on the observed Y, which is effectively imputing the observed Y, and certainly has the potential to add unnecessary variability, as was noted in Subsection 4.1.

6. Application to the 2015 Ohio Medicaid Assessment Survey

We now return to the missing income data for the 2015 OMAS (RTI International 2015) and apply the PPM model to assess the potential nonresponse bias in the estimated proportion of people at various income levels. Overall, 22.2% of subjects ($n = 9,511$) were missing the categorical income variable, and there were low levels of missingness for other variables important for the construction of design weights (e.g., age, gender, race). In order to focus on the potential bias in the income variable, we used the publicly available data that already had these other variables singly imputed (using a weighted sequential hot deck procedure).

We created two dichotomized income variables to analyze separately with the PPM model. Low income was defined as 138% or lower of the Federal Poverty Level (FPL), and high income was defined as 300% or more of the FPL. These two binary variables were analyzed separately using the PPM model; see Section 7 for a discussion of the extension of the model to ordinal outcomes. Covariates that were fully observed (or completed by single-imputation) and used in the analysis were stratum, region, household composition (number of adults, number of children), respondent age, respondent gender, respondent race, education level, insurance type, an indicator for fair/poor health status, and the sample weight. Probit regression was used to estimate the proxy, with the final models chosen based on the cases with income observed, using backwards selection starting from a model that contained all second-order interactions. Data and code for implementing the PPM model using the 2015 OMAS data is available at http://github.com/randridge/PPMA.

The ML estimates of the strength of the proxies were $\hat{\rho}^{(0)} = 0.71$ for low income, and $\hat{\rho}^{(0)} = 0.71$ for high income, reflecting a moderate to strong amount of information in the auxiliary data for predicting income. The standardized difference between the mean of the proxy X for the entire sample compared to that of the respondents, $d^* = \left(\bar{x} - \bar{x}^{(0)}\right)/\sqrt{S_{xx}^{(0)}}$, was $d^* = 0.066$ for low income and $d^* = -0.064$ for high income. Estimates of the proportion of low and high income and 95% intervals for each of $\phi = (0, 0.5, 1)$ were

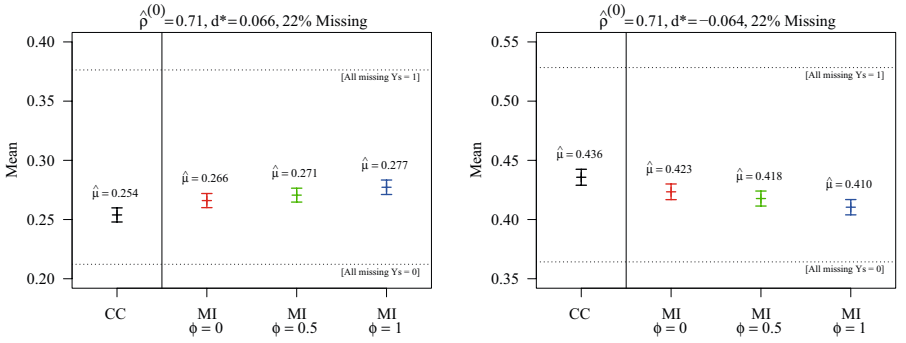


Fig. 2. Estimates of the proportion (a) low income and (b) high income based on 2015 Ohio Medicaid Assessment Survey data. CC: Complete cases; MI: 20 multiply imputed data sets using the proxy pattern mixture model with $\phi = 0$ (MAR), $\phi = 0.5$ (intermediate MNAR), and $\phi = 1$ (extreme MNAR)

obtained using the MI estimation procedure with $K = 20$ data sets. The burn-in period was 20 draws due to quick convergence and imputation occurred on every hundredth iteration. Since OMAS has a complex survey design, we used design-based estimators of the proportion using the survey weights via the "survey" routines in R, which estimate variances using Taylor series linearization (Lumley 2004).

Estimated proportions and 95% confidence intervals are displayed in Figure 2, with Manski bounds (obtained by filling in all zeros or all ones) denoted by dotted lines. For low income, the complete case (respondent) estimate underestimates the percent of subjects relative to the MAR and MNAR estimates, with the amount of underestimation increasing as ϕ increases. For high income, using the complete cases yields an estimate of the percent of high income earners that is too large, and as we move from MAR to MNAR the estimated proportion decreases. Since the proxies are relatively strong ($\hat{\rho}^{(0)}$), and the bias in the proxy (as measured by d^*) is relatively small, the confidence interval widths do not get drastically larger for $\phi = 1$.

While the potential impact of MNAR nonresponse on the estimated proportions do not seem very large, they are potentially large when put into context. Recall that 2015 was the first year OMAS was conducted after the Medicaid expansion in Ohio. Estimating the proportion of adults whose incomes are below 138% FPL is necessary to quantify the number of adults who are newly eligible for Medicaid. The difference between 26.6% low income and 27.7% low income could have large policy implications.

7. Discussion

In this article, we have extended the previously developed normal PPM analysis to handle binary data, which are ubiquitous in sample surveys. As with a continuous outcome, the new method integrates the three key components that contribute to nonresponse bias: the amount of missingness, differences between respondents and nonrespondents on characteristics that are observed for the entire sample, and the relationship between these fully observed covariates and the survey outcome of interest. The analysis includes, but does not make the assumption, that missingness is MAR, allowing the user to investigate a

range of MNAR mechanisms and the resulting potential for nonresponse bias. For the binary case, it is common to investigate what the estimates would be if all nonresponding units were zeros (or ones), and in fact the binary PPM analysis produces these two extremes (Manski bounds) when the proxy is weak, that is, when good predictors of Y among respondents do not exist. With even just moderately strong covariate information, as in the OMAS application, our new PPM method produces intervals that are considerably shorter than the Manski bounds, thus reducing the size of the potential nonresponse bias.

In the binary PPM model, ϕ is a sensitivity parameter and there is no information in the data with which to estimate it. Thus, we have proposed a sensitivity analysis using $\phi = \{0, 0.5, 1\}$ to produce separate estimates; comparing these estimates paints a picture of the potential nonresponse bias across a range of missingness mechanisms. An alternative approach would be to put a beta prior on f, as was done for the normal PPM model by Little et al. (2019). While this would produce a "single answer", the result is liable to be sensitive to how this prior is specified. If, for example, a noninformative Uniform(0,1) prior is used, the resulting posterior distribution interval for the mean of Y will effectively span the range from the estimate at $\phi = 0$ to the estimate at $\phi = 1$. This approach would produce a single interval and may be preferred by some practitioners.

There are some drawbacks to the binary PPM model, relative to the normal PPM on which it is based. The ease of implementation of the PPM model is lost in the binary case; closed-form ML estimates are no longer available and Bayesian and MI methods require iteration using Gibbs sampling. However, the ML solutions are good starting points for the Gibbs sampler and only very short burn-in periods were required in the examples we considered. An additional level of complexity in the binary case is the effect of skewed proxies. The normal PPM model estimates are relatively robust to departures from bivariate normality in the proxy and outcome, but the binary model relies more heavily on normality – even slight deviations away from normality in the latent variable lead to biased results. To address this weakness, we introduced modified estimators that perform better when the normality assumption is violated, while maintaining good performance when the normality assumption holds.

We have described three different estimation methods for the binary PPM model: ML, fully Bayesian (PD), and MI. In our investigations MI is consistently the best of these approaches. Unlike ML and Bayesian methods, it does not require a modification to handle skewed proxies, and complex design features like design weights are readily incorporated in the complete-data component of the MI combining rules. However, the ML and Bayesian methods do have one potential advantage: they can be applied in scenarios where microdata are available for respondents only (not available for the nonrespondents). If one has only the sample mean and covariance matrix of Z (the auxiliary variables) for nonrespondents, the unmodified ML and Bayesian methods could be used. In fact, if one only has the summary mean and covariance matrix of Z for the population, along with microdata for the respondents only, as in poststratification, the nonrespondent mean and covariance could be "backed out" of the population values, thus allowing ML and Bayesian estimation of the binary PPM model.

The binary PPM model can be extended to ordinal outcomes. Suppose instead of a binary outcome, we observe a partially missing ordinal outcome Y, where Y_i takes one of J ordered values, $1, \ldots, J$. As with the binary case, we assume there is an underlying latent

continuous variable U, related to the observed Y through the rule that $Y = j$ if $\gamma_{j-1} < U < \gamma_j$ for $j = 1, \ldots, J$, with $\gamma_0 = -\infty$ and $\gamma_J = \infty$. This latent structure allows an extension of probit regression to ordinal outcomes (e.g., Agresti 2002, chap. 7), such that $\Pr(Y \leq j|Z, M = 0) = \Pr(U \leq \gamma_j) = \Phi(\gamma_j + \alpha Z)$. For the PPM we take the proxy $X = \hat{\alpha} Z$, and apply the proxy pattern-mixture Model (4) to the joint distribution of the proxy X and latent U. Resulting ML estimates of the parameters $\mu_u^{(1)}$ and $\sigma_{uu}^{(1)}$ have the same form as in the binary case. The ML estimates of the parameters of the distribution of X for respondents and nonrespondents are the usual estimators. This leaves $\mu_u^{(0)}, \sigma_{uu}^{(0)}$, $\rho^{(0)}$ and $\gamma = \{\gamma_j\}$ to be estimated. Without loss of generality we take $\mu_u^{(0)} = 0$ and $\sigma_{uu}^{(0)} = 1$ and obtain ML estimates for the correlation $\rho^{(0)}$ and cutpoints γ. This reduces to the problem of estimating the polyserial correlation between the ordinal Y and continuous X, first considered by Cox (1974). As with the binary case, there is no closed-form solution and an iterative solution is required. The Bayesian and MI estimation methods follow as direct extensions of the binary case.

In future work we plan to extend PPM analysis to domain estimation, an important issue in practice. When the domain indicator is fully observed (for example, region in the OMAS data), application of the PPM model is straightforward; the domain indicator can be included in the model that creates the proxy, or the entire PPM method can be applied separately within each domain. If using multiple imputation, the product of the sampling weights and the domain indicator should be included as covariates in constructing the proxy (i.e., as part of Z) so MI variance estimates are unbiased (Kim et al. 2006). The more complex case is when the domain indicator and outcome are jointly missing, for example, income and current Medicaid status in the OMAS data. The methods of Little and Wang (1996), who extend the bivariate pattern-mixture model to the multivariate case when there are two patterns of missingness, may be useful in this scenario.

8. References

Agresti, A. 2002. *Categorical Data Analysis*. New York: Wiley.

Albert, J.H. and S. Chib. 1993. "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association* 88: 669–679. DOI: https://doi.org/10.1080/01621459.1993.10476321.

Andridge, R., A.M. Noone, and N. Howlader. 2017. "Imputing estrogen receptor (ER) status in a populationbased cancer registry: a sensitivity analysis." *Statistics in Medicine* 36: 1014–1028. DOI: https://doi.org/10.1002/sim.7193.

Andridge, R.R. and R.J.A. Little. 2011. "Proxy Pattern-Mixture Analysis for Survey Nonresponse." *Journal of Official Statistics* 27: 153–180. DOI: https://doi.org/10.1214/15-AOAS878SUPP.

Andridge, R.R. and K.J. Thompson. 2015. "Assessing nonresponse bias in a business survey: proxy pattern-mixture analysis for skewed data." *The Annals of Applied Statistics* 9: 2237–2265. DOI: https://doi.org/10.1214/15-AOAS878.

Barnhart, W.R., D. Ellsworth, A.C. Robinson, J. Myers, R.R. Andridge, and S.M. Havercamp. 2019. "Caregiving in the shadows: National analysis of health outcomes and intensity and duration of care among those who care for people with mental illness

and for people with developmental disabilities." *Disability and Health Journal* 3: 100837. DOI: https://doi.org/10.1016/j.dhjo.2019.100837.

Brick, J.M. and D. Williams. 2013. "Explaining Rising Nonresponse Rates in Cross-Sectional Surveys." *The ANNALS of the American Academy of Political and Social Science* 645: 36–59. DOI: https://doi.org/10.1177/0002716212456834.

Cox, N.R. 1974. "Estimation of the Correlation Between a Continuous and a Discrete Variable." *Biometrics* 30: 171–178. DOI: https://doi.org/10.2307/2529626.

Curtain, R., S. Presser, and E. Singer. 2005. "Changes in Telephone Survey Nonresponse over the Past Quarter Century." *Public Opinion Quarterly* 69: 87–98. DOI: https://doi.org/10.1093/poq/nfi002.

Heckman, J.J. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models." *The Annals of Economic and Social Measurement* 5: 475–492.

Hedeker, D., R.J. Mermelstein, and H. Demirtas. 2007. "Analysis of binary outcomes with missing data: missing = smoking, last observation carried forward, and a little multiple imputation." *Addiction* 102: 1564–1573. DOI: https://doi.org/10.1111/j.1360-0443.2007.01946.x.

Higgins, J.P.T., I.R. White, and A.M. Wood. 2008. "Imputation methods for missing outcome data in meta-analysis of clinical trials." *Clinical Trials* 5: 225–239. DOI: https://doi.org/10.1177/1740774508091600.

Jackson, D., I.R. White, D. Mason, and S. Sutton. 2014. "A general method for handling missing binary outcome data in randomized controlled trials." *Addiction* 109: 1286–1993. DOI: https://doi.org/10.1111/add.12721.

Kim, J.K., J.M. Brick, W.A. Fuller, and G. Kalton. 2006. "On the Bias of the Multiple-Imputation Variance Estimator in Survey Sampling." *Journal of the Royal Statistical Society B* 68: 509–521. DOI: https://doi.org/10.1111/j.1467-9868.2006.00546.x.

Little, R.J.A. 1993. "Pattern-Mixture Models for Multivariate Incomplete Data." *Journal of the American Statistical Association* 88: 125–134. DOI: https://doi.org/10.2307/2533148.

Little, R.J.A. 1994. "A Class of Pattern-Mixture Models for Normal Incomplete Data." *Biometrika* 81: 471–483. DOI: https://doi.org/10.1093/biomet/81.3.471.

Little, R.J.A. and D.B. Rubin. 2019. "Statistical Analysis with Missing Data." 3rd edition. Wiley: New York.

Little, R.J.A. and Y. Wang. 1996. "Pattern-Mixture Models for Multivariate Incomplete Data with Covariates." *Biometrics* 52: 98–111. DOI: https://doi.org/10.1080/01621459.1993.10594302.

Little, R.J.A., B.T. West, P.S. Boonstra, and J. Hu. 2019. "Measures of the Degree of Departure from Ignorable Sample Selection." *Journal of Survey Statistics and Methodology*. DOI: https://doi.org/10.1093/jssam/smz023.

Liublinska, V. and D.B. Rubin. 2014. "Sensitivity analysis for a partially missing binary outcome in two-arm randomized clinical trial." *Statistics in Medicine* 33: 4170–4185. DOI: https://doi.org/10.1002/sim.6197.

Lumley, T. 2004. "Analysis of Complex Survey Samples." *Journal of Statistical Software* 9: 1–19.

Magder, L.S. 2003. "Simple approaches to assess the possible impact of missing outcome information on estimates of risk ratios, odds ratios, and risk differences." *Controlled Clinical Trials* 24: 411–421. DOI: https://doi.org/10.1016/s0197-2456(03)00021-7.

Manski, C.F. 2016. "Credible Interval Estimates for Official Statsitics with Survey Nonresponse." *Journal of Econometrics* 191: 293–301. DOI: https://doi.org/10.1016/j.jeconom.2015.12.002.

Muthen, B., T. Asparouhov, A.M. Hunter, and A.F. Leuchter. 2011. "Growth Modeling With Nonignorable Dropout: Alternative Analyses of the STAR∗D Antidepressant Trial." *Psychological Methods* 16: 17–33. DOI: https://doi.org/10.1037/a0022634.

Nandram, B. and J.W. Choi. 2002a. "A Bayesian Analysis of a Proportion Under Non-Ignorable Non-Response." *Statistics in Medicine* 21: 1189–1212. DOI: https://doi.org/10.1002/sim.1100.

Nandram, B. and J.W. Choi. 2002b. "Hierarchical Bayesian Nonresponse Models for Binary Data from Small Areas with Uncertainty about Ignorability." *Journal of the American Statistical Association* 97: 381–388. DOI: https://doi.org/10.1198/016214502760046934.

Nandram, B., G. Han, and J.W. Choi. 2002. "A Hierarchical Bayesian Nonignorable Nonresponse Model for Multinomial Data from Small Areas." *Survey Methodology* 28: 145–156.

Nandram, B., N. Liu, J.W. Choi, and L. Cox. 2005. "Bayesian Non-response Models for Categorical Data from Small Areas: An Application to BMD and Age." *Statistics in Medicine* 24: 1047–1074. DOI: https://doi.org/10.1002/sim.1985.

Olsson, U., F. Drasgow, and N.J. Dorans. 1982. "The Polyserial Correlation Coefficient." *Psychometrika* 47: 337–347. DOI: https://doi.org/10.1007/BF02294164.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org (accessed June 2020).

RTI International. 2015. *2015 Ohio Medicaid Assessment Survey Methodology Report*. Technical report, RTI International. Available at: http://grc.osu.edu/sites/default/files/inline-files/12015OMASMethReptFinal121115psg.pdf (accessed June 2020).

Rubin, D.B. 1976. "Inference and Missing Data" (with Discussion). *Biometrika* 63: 581–592. DOI: https://doi.org/10.1093/biomet/63.3.581.

Rubin, D.B. 1977. "Formalizing Subjective Notions About the Effect of Nonrespondents in Sample Surveys." *Journal of the American Statistical Association* 72: 538–542. DOI: https://doi.org/10.1080/01621459.1991.10475033.

Rubin, D.B. 1978. "Multiple Imputation in Sample Surveys." A Phenomenological Bayesian Approach to Nonresponse. In Proceedings of the Survey Research Methods Section, American Statistical Association (San Diego, CA): 20–34. DOI: https://doi.org/10.1002/9780470316696.

Stasny, E.A. 1991. "Hierarchical Models for the Probabilities of a Survey Classification and Nonresponse: An Example from the National Crime Survey." *Journal of the American Statistical Association* 86: 296–303. DOI: https://doi.org/10.1080/01621459.1991.10475033.

Sullivan, D. and R. Andridge. 2015. "A hot deck imputation procedure for multiply imputing nonignorable missing data: The proxy pattern-mixture hot deck."

Computational Statistics and Data Analysis 82: 173–185. DOI: https://doi.org/10.1016/j.csda.2014.09.008.

Tate, R.F. 1955a. "Applications of Correlation Models for Biserial Data." *Journal of the American Statistical Association* 50: 1078–1095. DOI: https://doi.org/10.1080/01621459.1955.10501293.

Tate, R.F. 1955b. "The Theory of Correlation Between Two Continuous Variables When One is Dichotomized." *Biometrika* 42: 205–216. DOI: https://doi.org/10.2307/2333437.

Received August 2018
Revised May 2019
Accepted October 2019