Robust Methods for Estimating the Mean with Missing Data

By

Ye Yang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in the University of Michigan
2015

Doctoral Committee:

Professor Roderick J. Little, Chair
Professor Michael R. Elliott
Professor Trivellore E. Raghunathan
Professor Naisyin Wang

To my parents, and Shady

Acknowledgements

I would like to express my utmost appreciation and gratitude to my advisor, Rod Little. Thank you for your invaluable guidance, mentorship, and patience throughout the years. You have made me a better student, a better researcher, and helped me learn so much. It's been truly a pleasure and honor to work with you. Thank you, my committee members Mike Elliott, Trivellore Raghunathan, and Naisyin Wang, for your constructive input to this research.

To my parents, Ning Yang and Ying Chen, thank you for your encouragement and support that have helped make this dissertation possible. This is dedicated to you.

Finally, to all my friends in Ann Arbor, thank you for making the past six years enjoyable and memorable. I will miss all the nights we played basketball at NCRB together, the dinners, card games, barbeques, road trips, and most of all, your company. I will forever cherish these memories. Hope one day our lives cross paths again.

TABLE OF CONTENTS

LIST OF FIGURES

<u>Figure</u>

LIST OF TABLES

Table

CHAPTER I

Introduction

Missing data are a common problem in many empirical studies. In surveys, sampled units may be difficult to reach, or may refuse to respond to some or all of the survey questions, leading to unit or item nonresponse. In many cases we can obtain fully observed auxiliary variables, which may be used to predict the missing values. Useful auxiliary variables are predictive of the missing variables as well as the probability of observing these variables.

In this dissertation we consider the problem of estimating the mean of an outcome variable subject to nonresponse, under a setting in which we have one or more fully observed covariates. Some commonly used methods to address the issue of nonresponse include complete case analysis, which estimates the mean using only observed values of the outcome, and multiple imputation, which models the outcome parametrically on the observed covariates. In these methods, we make assumptions regarding the relationship between the outcome and the predictors as well as the missing data mechanism. Violation of these assumptions, i.e. model misspecification, will lead to biased estimates of the mean. In the following chapters, we attempt to address this issue by proposing methods for estimating the mean that are more robust to model misspecification.

In the second chapter, we consider data in which the outcome is missing at random (MAR), where missingness depends only on the observed covariates. We explore a variety of doubly robust estimators (DR), which specify both a model for the mean and a model for the propensity to respond. While DR estimators are consistent if either the mean or propensity model is correctly specified, it is not clear which will perform best under different settings. We attempt to answer this question through simulations under a variety of scenarios, and compare the performances of each DR estimator with respect to its root mean squared error (RMSE), confidence interval width (CIW), and coverage rate. Finally, we apply the methods to an asthma study conducted at the University of Michigan.

In many situations, the outcome may be missing not at random (MNAR), in which traditional MAR-based methods are biased. Chapter III proposes a modification of the pattern mixture model in Little (1994) for assessing nonresponse bias under MNAR. We assume a continuous outcome variable Y and a fully observed covariate X. The method adopts a Bayesian approach, and utilizes a robust spline model to estimate the mean of the outcome assuming that missingness depends on the value of $X + \lambda Y$ for some λ. Estimates under different values of λ are presented to assess for sensitivity and potential for bias from MNAR. We then extend this analysis to a set of covariates Z. For simplicity, we reduce the set of Z into a single proxy X that is the best predictor of Y, obtained by regressing Y on Z for the respondents, and apply the method to the proxy X

and Y. We explore the properties of the proposed method and the original pattern mixture model in simulations and data from the asthma study conducted at the University of Michigan.

In some cases we may be interested in estimating the mean of a binary variable. In the fourth chapter, we extend the analysis discussed in Chapter III to binary outcomes using a latent variable approach. Performances of the proposed extension are illustrated through simulations.

CHAPTER II

A Comparison of Doubly Robust Estimators of the Mean with Missing Data

2.1. Introduction

In this chapter we consider the situation where we have a continuous survey outcome *Y* with *r* observations ($\{y_i\}$, i = 1, …, *r*) and *n* - *r* nonresponses and a set of *p* covariates X_1, …, X_p that are fully observed ($\{x_{i1}$, …, $x_{ip}\}$, i = 1, …, *n*). Suppose *R* is an indicator variable that takes a value of 1 if *Y* is observed and 0 if *Y* is missing. We assume that *Y* is missing at random (MAR), so that the missingness of *Y* depends only on X_1, …, X_p. The goal is to estimate μ, the mean of *Y*.

A simple and common approach is to estimate μ using only the complete cases. The complete-case mean is inefficient if *X* is predictive of *Y*, because information from incomplete cases is lost, and biased if missingness of *Y* depends on the observed covariates *X*. An alternative to CC analysis is weighted complete case analysis (WCC), which estimates the mean by $\hat{\mu} = \sum_{i=1}^{r} w_i y_i / n$, where w_i is a weight calculated as the inverse of the estimated probability that *R* = 1 given a fully observed set of covariates *X*. WCC is consistent under MAR, but is less efficient than CC analysis if *X* is not associated with *Y*, particularly if *X* is highly associated with *R* (Little and Vartivarian, 2005).

Parametric imputation models for the distribution of Y given X can also be applied to impute or multiply impute the missing values of Y. While imputation can increase precision by exploiting information on the covariates, it is vulnerable to misspecification of the regression model, which may lead to bias.

Doubly-robust (DR) estimators have been developed to protect against the effects of model misspecification and improve the robustness of estimates. An estimator of μ is DR if it is consistent when either the regression model for the mean function or the propensity to respond (the "propensity model") is correctly specified. In this chapter we consider the following DR estimators of the mean:

1. Penalized spline of propensity prediction (PSPP), an approach that regresses Y on the estimated response propensity score flexibly via a penalized spline.
2. Calibration (CAL) methods, with estimates of the form

   $$\hat{\mu} = n^{-1}(\sum_{i=1}^{n} \hat{y}_i) + n^{-1}[\sum_{i=1}^{r} w_i(y_i - \hat{y}_i)],$$

   a function of the predicted mean of the respondents and nonrespondents and a weighted average of the residuals.
3. Modified calibration methods (MCAL), where the division of n in the weighted sum of residuals is replaced by $(\sum_{i=1}^{r} w_i)$.

In CAL and MCAL methods, $\hat{y}_i$ may be estimated using either ordinary (OLS) or weighted least squares (WLS). In addition, Cao, et al. (2009) proposed a DR calibration estimator that has the smallest asymptotic variance among all calibration methods if the propensity score is correctly specified.

While these DR estimators are asymptotically consistent and efficient when either the model for the propensity or the mean function is well specified, it is not clear how to choose between them in applied problems, and in particular, their properties for finite sample sizes are of interest. Zhang and Little (2011) compared in a simulation study performances of PSPP, CAL, WCC, and a linear in weight prediction method (LWP), where Y is regressed linearly on the weights, under various scenarios of correct and incorrectly specified mean and propensity models. Results showed that PSPP yielded better estimates of μ in terms of root mean square error and confidence interval coverage than both LWP and CAL, with all three DR methods having large gains over WCC when the estimated propensity is incorrect. Although CAL and LWP generally yielded similar estimates of μ, CAL had superior precision at small sample sizes. In many applications it is advantageous to substitute the sum of response weights, $\sum_{i=1}^{r} w_i$, for *n* as in MCAL, since $\sum_{i=1}^{r} w_i$ provides some protection against large weights caused by small propensities. Moreover, using WLS in the regression of y_i may offer improvements over OLS in CAL when the regression is not linear, where WLS helps to reduce bias in the mean. Furthermore, under a correct propensity model the alternative calibration method proposed in Cao, et al. (2009) promises superior asymptotic variance than both CAL and MCAL. Thus, it is of interest to further explore the properties of the various forms of calibration under different scenarios and how they compare with PSPP. In this chapter we expand the comparisons of CAL and PSPP in Zhang and Little (2011) to include other simulation scenarios, and the alternatives to CAL described above. Specifically, through simulations we will attempt to answer the following questions:

1. How do the bias and root mean squared error of the estimates of μ compare under different forms of model misspecification, as sample sizes vary from small to large? In particular:
   (a) How do the performances of DR estimates that multiply the weighted residuals by $(\sum_{i=1}^{n} w_i)^{-1}$ compare with analogous estimates that multiply the weighted estimates by n^{-1}?
   (b) How does robustness and efficiency of the calibration methods compare when the calibration means are predicted by OLS vs. WLS, for various choices of regression weights (as discussed below)?
   (c) How do these calibration methods compare with the PSPP, which uses predictions from a robust model rather than calibration to achieve robustness?
2. How wide and how close to the nominal coverage are the associated confidence intervals for the various methods? Additionally, what are the repeated sampling properties of the posterior distribution of μ based on a Bayesian implementation of PSPP, with reference prior distributions, and how do they compare with the PSPP method using a bootstrap estimate of the variance?

In section 2.2, we present the various alternative methods in more detail. In section 2.3.1-2.3.5, we describe simulation studies designed to compare the methods under correctly and incorrectly specified regressions for the mean and propensity to respond, and evaluate the results based on root mean square error (RMSE) of estimates and width and coverage of confidence intervals under repeated sampling. In section 2.4

we apply the methods to data from an asthma intervention study. Concluding remarks and discussion are provided in section 2.5.

2.2. Doubly Robust Estimators

All methods assume that Y is MAR, that is, Y and R are independent given X_1, …, X_p, and are based on two regressions: (a) a regression model for Y on X_1, …, X_p, estimated from the subsample of respondents, and (b) a regression for the propensity to respond, $Pr(R = 1|X_1, \ldots, X_p)$, estimated from a logistic regression of R on X_1, …, X_p using all the data. The DR property refers to consistent estimation of the mean of Y provided one of these two regressions is correctly specified.

2.2.1. Calibration prediction by OLS – dividing by n

Robins, et al. (1994) proposed a class of augmented inverse probability weighted estimators for the mean that calibrates predictions from a linear regression model with a weighted average of the residuals from observed outcomes. This method combines information from complete and incomplete cases by modeling both the outcome and propensity using a set of fully observed covariates X. The calibration estimator takes the form:

$$\hat{\mu} = n^{-1}(\textstyle\sum_{i=1}^{n} \hat{y}_i^{ols}) + n^{-1}[\textstyle\sum_{i=1}^{r} w_i(y_i - \hat{y}_i^{ols})] \quad (2.1)$$

where $\hat{y}_i^{ols} = E(y_i|X_1, \ldots, X_p)$ is the predicted mean from the linear regression of Y on X, fitted by OLS, and $w_i = 1/\widehat{Pr}(R_i = 1|X_1, \ldots, X_p)$ is the estimated inverse of the probability of response for the i^{th} subject. The estimator is DR as it yields consistent estimates if either the model for the prediction or propensity is correctly specified. Setting the predicted means to 0 results in the WCC estimate $\hat{\mu} = \sum_{i=1}^{r} w_i y_i / n$.

2.2.2. Calibration prediction by OLS – dividing by sum of weights

An alternative form of the calibration estimator is to replace n^{-1} in the second part of (2.1) by $(\sum_{i=1}^{r} w_i)^{-1}$, the inverse of the sum of the estimated respondent weights, yielding:

$$\hat{\mu} = n^{-1}(\sum_{i=1}^{n} \hat{y}_i^{ols}) + (\sum_{i=1}^{r} w_i)^{-1}[\sum_{i=1}^{r} w_i(y_i - \hat{y}_i^{ols})] \quad (2.2)$$

In most cases $\sum_{i=1}^{r} w_i$ will be approximately equal to *n*. However, (2.2) tends to reduce the effects of extreme weights caused by small propensity scores for some subjects, since in (2.2) these weights are propagated in the dominator of the second term.

2.2.3. Calibration prediction by WLS

A third variation in the calibration estimator is to predict the outcome using WLS:

$$\hat{\mu} = n^{-1}(\sum_{i=1}^{n} \hat{y}_i^{wls}) + (\sum_{i=1}^{r} w_i)^{-1}[\sum_{i=1}^{r} w_i(y_i - \hat{y}_i^{wls})] \quad (2.3)$$

where $\hat{y}_i^{wls}$ is the predicted value of *Y* for the i^{th} individual obtained by WLS with weights w_i. The property of WLS for a linear regression with an intercept implies $\sum_{i=1}^{r} w_i(y_i - \hat{y}_i^{wls}) = 0$, thus (2.3) reduces to $\hat{\mu} = n^{-1}(\sum_{i=1}^{n} \hat{y}_i^{wls})$, the mean of the weighted predictions for the entire sample. WLS regression helps to reduce bias in the mean when the regression is not linear, and hence may be more effective than calibrating OLS estimates by the average of the weighted residuals.

2.2.4. Calibration prediction by Cao, et al.

In Robins, et al. (1994) and Tsiatis and Davidian (2007), all consistent and asymptotically normal estimators of μ when the propensity score model is correct are asymptotically equivalent to the estimator

$$\hat{\mu} = n^{-1}\sum_{i=1}^{n}\left[\frac{R_i Y_i}{\pi(X_i,\hat{\gamma})} - \frac{R_i - \pi(X_i,\hat{\gamma})}{\pi(X_i,\hat{\gamma})} h(X_i)\right] \quad (2.4)$$

where $\pi(X_i, \hat{\gamma}) = \widehat{Pr}(R_i = 1|X)$, estimated by logistic regression via maximum likelihood. This is equal to the estimator in (2.1) when $h(X_i) = \hat{y}_i = m(X_i, \beta^{ols})$, where $m(X_i, \beta^{ols})$ is the mean regression model with $\hat{\beta}^{ols}$ estimated by ordinary least squares. This estimator has the smallest asymptotic variance among those in class (2.4) when the mean regression model is correct, but not when the mean regression model is misspecified, even if the propensity model is correct. Cao et al. (2009) propose an estimator of the form (2.4) with $h(X_i) = \hat{y}_i = m(X_i, \beta^{opt})$ that is DR and has the smallest asymptotic variance if either the propensity or mean regression model is correctly specified. To achieve this, $\hat{\beta}^{opt}$ for the outcome regression model is estimated by solving jointly for (β, c):

$$\sum_{i=1}^{n}\left[\frac{R_i}{\pi(X_i,\hat{\gamma})}\frac{1-\pi(X_i,\hat{\gamma})}{\pi(X_i,\hat{\gamma})}\begin{Bmatrix} m'(X_i,\beta) \\ \frac{\pi'(X_i,\hat{\gamma})}{1-\pi(X_i,\hat{\gamma})}\end{Bmatrix}\left\{Y_i - m(X_i,\beta) - c^T\frac{\pi'(X_i,\hat{\gamma})}{1-\pi(X_i,\hat{\gamma})}\right\}\right] = 0$$

where $m'(X_i, \beta) = d/d\beta\ [m(X_i, \beta)]$ and $\pi'(X_i, \hat{\gamma}) = d/d\gamma\ [\pi(X_i, \hat{\gamma})]$.

If the mean regression model is misspecified, this estimator will still have smaller asymptotic variance than any estimator of the form (2.4), as long as the propensity model is correct.

2.2.5. Penalized spline of propensity prediction(PSPP)

The PSPP method predicts the missing values of Y from the following mixed-effects model:

$$Y = s(P^*) + g(X_2, \ldots, X_p) + \varepsilon,\ \varepsilon \sim N(0, \sigma^2) \tag{2.5}$$

Where $P^* = logit[Pr(R = 1 | X_1, \ldots, X_p)]$, and $s(P^*)$ is a penalized spline of the form

$$s(P^*) = \beta_0 + \beta_1 P^* + \sum_{k=1}^{K} \gamma_k (P^* - \kappa_k)_+ \tag{2.6}$$

where $a_+ = a$ if $a > 0$ and $a_+ = 0$ otherwise, and $\kappa_1 < \ldots < \kappa_K$ are K equally spaced, fixed knots. γ_k are assumed normal with mean 0 and variance τ^2. One of $X_1, \ldots, X_p$, here X_1, is omitted in $g()$ to avoid collinearity. In practice P^* is unknown and estimated from the logistic regression of R on $X_1, \ldots, X_p$.

The PSPP model may be fitted as a linear mixed model treating the splines as random effects. In our simulation, we adopt a Bayesian version of the PSPP model, where we assign a uniform prior for β and inverse gamma priors with parameters (10^{-5}, 10^{-5}) and (10^{-5}, 10^{-5}) for σ^2 and τ^2, respectively. We choose small values for these prior parameters in order to result in non-informative but proper priors. Inferences are based on the posterior predictive distribution of the mean of Y, computed using the Gibbs sampler (see Appendix for details of the algorithm).

Since a property of propensity scores is that missingness of Y is independent of Y given P^*, to limit bias it is sufficient to model the relationship between Y and P^* correctly. As a result, predictions from the PSPP model have a DR property. That is, predictions of Y are consistent if either:

a. $E(Y | P^*, X_1, \ldots, X_p) = \beta_0 + \beta_1 P^* + g(X_2, \ldots, X_p)$, or

b. $E(Y|P^*, X_1, \ldots, X_p) = s(P^*)$ and P^* is correctly specified.

The spline allows flexible modeling of the mean as a function of the propensity score. The parametric component $g()$ is designed to increase precision, and the mean is consistently estimated even if the function $g()$ is not correctly specified, if the propensity model is correctly specified. The PSPP model may be extended to non-normal outcomes using generalized linear models and appropriate link functions.

2.3. Simulation studies

We study the performance of the estimators by comparing root mean square error (RMSE), and confidence interval width and coverage rate, under eight scenarios. We simulate 1000 data sets with sample sizes of 50, 100, 200, 400, and 800. The first four scenarios are adapted from Zhang and Little (2011), where either the regression of Y on X or the propensity model is misspecified. For each of these scenarios, we vary degrees of misspecification of both propensity and mean functions. The last four scenarios are taken from Kang and Schafer (2007), which were further studied by Cao et al. (2009). To address some limitations in the choices of misspecified models in Kang and Schafer (2007), we study four different combinations of correct and misspecified mean and propensity functions.

We compare the performance of the estimators by their RMSE relative to the RMSE of the (infeasible) before deletion (BD) analysis, which estimates the average of all the values of Y with none of the values deleted. We define relative RMSE of an estimator as

$$RRMSE_{est} = 100 \times \frac{RMSE_{est} - RMSE_{BD}}{RMSE_{BD}}$$

where RMSE is the square root of the average mean square error over the 1000 samples.

We estimate the variances of the marginal means of *Y* for the methods via the bootstrap. For each replicate simulation, we apply the methods to 200 bootstrap samples, and variance is estimated as

$$Var_{boot}(\hat{\mu}) = \frac{1}{199}\sum_{b=1}^{200}(\hat{\mu}^{(b)} - \bar{\mu}_{boot})^2$$

where $\hat{\mu}^{(b)}$ is the estimated marginal mean of *Y* for the b^{th} bootstrap sample and $\bar{\mu}_{\text{boot}}$ is the average of estimated marginal means of *Y* over all bootstrap samples.

For the Bayesian method of PSPP, we impute the missing values of *Y* by taking draws from the posterior predictive distribution of *Y* given *X*. This is implemented by drawing $Y^{(d)}|X, \beta^{(d)}, \gamma^{(d)} \sim N(s(P^*) + g(X_2, \ldots, X_p), \sigma^{2(d)})$, where superscript (*d*) denotes the conditional draw of the parameter in the d^{th} iteration of the Gibbs sampling algorithm. Applying the algorithm over a total of 10000 iterations and deleting the first 1000 for burn-in, we obtain *D*=9000 imputed data sets, and the variance of the marginal mean is estimated as

$$Var(\hat{\mu}_{PSPP}) = \frac{1}{D}\sum_{d=1}^{D} W_d + \frac{D+1}{D(D-1)}\sum_{d=1}^{D}(\hat{\mu}_d - \bar{\mu}_D)^2$$

where W_d is the marginal variance in the d^{th} imputed data set, $\hat{\mu}_d$ is the estimated marginal mean in the d^{th} imputed data set, and $\bar{\mu}_D = \frac{1}{D}\sum_{d=1}^{D}\hat{\mu}_d$.

We construct 95% confidence intervals (CI) for each of the 1000 samples and estimate the coverage rate as the proportion of the 1000 confidence intervals that cover the true value, where CI = $(\hat{\mu} - t_{n-1,0.975}\sqrt{Var(\hat{\mu})}, \hat{\mu} + t_{n-1,0.975}\sqrt{Var(\hat{\mu})})$, and $t_{n-1,0.975}$ is the

97.5^{th} percentile of the t distribution with n-1 degrees of freedom. Confidence interval widths, computed as CIW = $2*t_{n-1,0.975}\sqrt{Var(\hat{\mu})}$, are averaged over the 1000 samples. As with RMSE, we compare the interval width of the estimators relative to those of the BD analysis. The relative confidence interval width (RCIW) is defined as

$$RCIW_{est} = 100 \times \frac{CIW_{est} - CIW_{BD}}{CIW_{BD}}$$

2.3.1. Simulation 1: misspecified quadratic mean function and correct propensity model

In this scenario we generate missing values of Y under the following propensity model:

$$logit[Pr(R = 1 | X_1, X_2)] = \alpha_1 X_1 \qquad (2.7)$$

and the true mean structure:

$$Y | X_1 \sim N(1 + X_1 + \alpha_2 X_1^2, 1) \qquad (2.8)$$

where X_1 is a fully observed covariate with a standard normal distribution. For simulations 2.1-2.4, we vary the degree of dependence of R on X, setting α_1=0.1 for low dependence and α_1=0.5 for high dependence. In both cases the expected overall response probability is 0.5. We estimate the propensity (2.7) using a correctly specified logistic regression model, $logit[Pr(R = 1 | X_1)] = \hat{\alpha}_0 + \hat{\alpha}_1 X_1$. The mean function for Y in (2.8) in the calibration methods is misspecified as a linear rather than quadratic function of X_1. We set α_2=0.8 and α_2=4 in (2.8) to simulate respectively low and high degrees of misspecification of the mean function. Figure A2.1 of the Appendix displays relationship between Y and X_1 for respondents by levels of dependence of mean and propensity models, which shows clear misspecification of the mean model when X_1^2 is not included as a predictor, particularly at α_2=4. The marginal mean is then estimated by the following methods:

1. Calibration method (CAL) described in (2.1).
2. The modified version of calibration (MCAL) described in (2.2).
3. Calibration method (WCAL) where $\hat{y}_i$ is now the prediction for the i^{th} subject using a WLS regression of Y on X_1, with weights being the inverse of the estimated response probability.
4. The robust form of calibration (RCAL) proposed by Cao, et al. (2009).
5. The PSPP method from (2.5) and (2.6) with a null g function. We choose the number of fixed, equally spaced knots for the penalized spline to be equal to 5, 10, and 15 for sample sizes 50, 100, and 200 or more, respectively. We adopt both a Bayesian and maximum likelihood approach to this model. We note that this simulation set-up (unlike later ones) tends to favor PSPP over the other methods, since the spline on the propensity allows the true curvilinear relationship between Y and X_1 to be approximated.

Figure 2.1 displays the RRMSE for low (L) and high (H) degrees of dependence of the propensity model on X_1, and low and high degrees of misspecification of the mean function. Thus, LH represents low dependence in the propensity model and high misspecification in the mean model. For comparison purposes we include inferences under two regression models without propensity adjustments: the correctly specified model (CORR) where the quadratic term is included in the regression for Y, and the incorrectly specified regression model where the quadratic term is omitted (MISS). Due to similarity between CAL and MCAL, results for MCAL are not displayed in our figures.

In all scenarios and sample sizes, PSPP yields the smallest RRMSE compared to the other methods, with the exception of the correctly specified model CORR. There was very little difference in RRMSE between Bayes and maximum likelihood versions of PSPP (see Appendix). The advantages of PSPP are most apparent when the mean of Y is strongly associated with X_1^2 (LH and HH), as the splines help mimic the true quadratic relationship, as seen by RRMSEs close to those of the correctly specified regression model CORR. Among the calibration methods, RRMSEs for RCAL are consistently lower than those of CAL, MCAL, and WCAL, and approaches the RRMSE of PSPP and the correct model when the sample size and strength of association between R and X_1 is high. Fitting the regression model using weighted least squares in WCAL results in similar RRMSE as MISS in LL and LH (where WCAL lines overlap with those of MISS in Figure 2.1), as weighting has minimal effect in correcting bias due to a weak dependence of propensity on X_1. However, WCAL shows slight but consistent improvements in RRMSE over both CAL and MCAL, with higher gains when the propensity is strongly dependent on X_1. MCAL has minimal gains over CAL in RMSE, as the sum of respondent weights is approximately equal to the sample size in this scenario.

Relative width of 95% confidence intervals and coverage rates are shown in figures 2.2 and 2.3, respectively. Relative performances of the methods with respect to confidence interval widths are similar to RRMSE. PSPP has the smallest confidence interval in all scenarios, other than inferences under the correctly-specified model CORR.

RCAL has the smallest confidence interval width among all calibration methods except when sample size is 50 in LL and HL, which reflects its property of having the least asymptotic variance of its class when either the propensity or mean model is correctly specified. Both WCAL and MCAL perform similarly to MISS in LL and LH and show reductions in confidence interval widths over CAL, particularly at smaller sample sizes where response weights may be more variable. In such cases WLS estimation and summing the respondent weights may help to stabilize estimates of μ.

All methods have coverage rates close to the nominal 95% in LL and HL and in LH and HH when sample sizes are 400 or more. Under-coverage is most apparent in smaller sample sizes when the degree of misspecification of the mean model is high (LH and HH). Both PSPP and RCAL tend to have coverage rates closer to the nominal 95% than CAL, MCAL, and WCAL.

2.3.2. Simulation 2: misspecified mean function with interaction and correct propensity model

For this scenario, the missing values of *Y* are generated under the model:

$$logit[Pr(R = 1|X_1, X_2)] = 0.25X_1 - \alpha_1 X_2$$

where α_1 takes a value of 0.1 and 0.5 for respectively low and high degree of dependence of *R* on the covariates. As in scenario 1, we estimate the propensity by a correctly specified linear additive logistic regression model of *R* on X_1 and X_2. The regression of *Y* on the covariates is misspecified by including an interaction term in the true distribution of Y given X_1 and X_2 that is included in correct model (CORR) but excluded in the other fitted models:

$$Y|X_1, X_2 \sim N(1 + X_1 + X_2 + \alpha_2 X_1 X_2, 1)$$

The strength of dependency of Y on the interaction between X_1 and X_2, is varied from low (α_2 = 0.8) to high (α_2 = 4) levels (see Figure A2.2 of Appendix for plots of Y on X_1 and X_2 for respondents). For the calibration methods, we predict $\hat{y}$ using a linear regression of Y on X_1 and X_2. For PSPP we include only the linear term in X_2 in the $g()$ function, omitting X_1 to avoid collinearity. Unlike the first simulation, the regression of Y is not well approximated in the PSPP model since the interaction term is omitted.

Figures 2.4-2.6 displays the RRMSE, relative confidence interval width, and coverage rates. As in the previous simulation, results for correctly specified (CORR) and misspecified (MISS) regressions are included for comparison, and MCAL is omitted since its results are similar to those of CAL.

Figure 2.4 indicates that CORR has superior RRMSE to the other methods, so there is some penalty for misspecification regardless of method. In LL, where the propensity is weakly associated with X and the mean function is only slightly misspecified, all methods other than CORR achieved similar RRMSE except RCAL, which was inferior at smaller sample sizes. In other scenarios, RCAL has the lowest RRMSE at sample sizes of 200 to 800, while having the highest RRMSE when sample size is 50. PSPP yields similar or lower RRMSE than the CAL, MCAL and WCAL calibration methods, with larger gains in RRMSE in the HH situation. WCAL, while performing similarly to MISS in LL and LH as shown by their overlapping lines in Figure 2.4, consistently outperforms CAL and MCAL particularly in the HH scenario.

Figure 2.5 shows results for confidence interval widths. Aside from the (correctly specified) CORR, PSPP yielded the narrowest confidence intervals, except for sample size 800 in the HH scenario, where RCAL yielded narrower intervals. The asymptotic properties of RCAL are again illustrated in this simulation, where RCAL was the best of the calibration methods at large sample sizes but yielded much wider confidence intervals than the other methods at small samples, particularly sample size 50. MCAL shows reduction in interval widths over CAL when n = 50 in all four scenarios, but differences are minimal as sample size becomes large. As seen in RRMSE, WCAL results in large improvements in precision over CAL in HH. In LL, MISS yields lower confidence interval widths at small samples than the DR methods. This is perhaps due to the low dependence of Y on the interaction of X_1 and X_2, which MISS omits, resulting in only a slight departure from the correct model. In terms of coverages (Figure 2.6), RCAL tended to be conservative at small sample sizes. The other methods tended to have close to nominal or slightly anti-conservative coverage, with differences between PSPP, CAL, MCAL, and WCAL in coverage generally being minor. Improvements in coverage over the misspecified model (MISS) are evident at large sample sizes in the HH scenario, illustrating gains in the robust modeling methods.

2.3.3. Simulation 3: misspecified discontinuous mean function and correct propensity model

For this scenario, the missing values of Y are generated from the model $logit[Pr(R = 1|X_1,X_2)] = \alpha_1 X_1$ as in Simulation 1 and the distribution of Y given X_1 and X_2 is:

$$Y|X_1, X_2 \sim N(5 + X_1 + X_1^2, 1) \text{ if } X_1<0, \text{ and}$$

$$Y|X_1, X_2 \sim N(-5 + 2X_1, 1) \text{ if } X_1 \geq 0$$

when the degree of misspecification is low and:

$$Y|X_1, X_2 \sim N(5 + X_1 + X_1^2 + X_2 + 5X_1X_2, 1) \text{ if } X_1<0, \text{ and}$$

$$Y|X_1, X_2 \sim N(-5 + 2X_1 + X_2 + 5X_1X_2, 1) \text{ if } X_1 \geq 0$$

when the degree of misspecification is high (see Figure A2.3 of Appendix for plots of Y on X_1 and X_2 for respondents). Here, we introduce a discontinuity in the mean function of Y at $X_1 = 0$. We estimate the propensity by a correctly specified logistic regression and the marginal mean of Y by the same methods as in Simulation 2.

RRMSE, relative width of confidence interval, and coverage rates are shown in figures 2.7-2.9. In LL and HL cases, where Y is dependent only on X_1, the penalized spline resembles the true mean function and consequently PSPP yields lower RRMSE than all methods other than CORR. However, in LH and HH when Y depends on both X_2 and its interaction with X_1, the two DR conditions of PSPP no longer hold, so the spline fails to model the true mean function correctly leading to RRMSE similar or worse than RCAL, WCAL, and CAL. Similar to previous scenarios, when the propensity is highly correlated with X, RCAL yields large RRMSEs in small sample sizes and decreases as sample size becomes larger. Overall, there is no distinguishable difference in RRMSE between the calibration estimators except in HH, where CAL and MCAL (omitted from figures due to similarity with CAL) have slightly higher RRMSEs than WCAL and RCAL. Moreover, as seen in every scenario, there is little difference between MISS and WCAL estimates when the propensity is weakly correlated with *X*.

Due to overfitting a discontinuous mean function, the correct regression model yields extreme confidence interval widths at sample size of 50. Similar to previous scenarios, we notice a sharp decrease in confidence interval width of RCAL as sample size increases. PSPP yields smaller confidence intervals than the calibration methods in LL, LH, and HL. In HH, however, all methods except CAL and MCAL yield similar interval widths at sample sizes greater than 100, again demonstrating gains in using WLS when both mean and propensity are highly correlated with X. In all cases, the methods cover the true value at a rate close to the nominal 95%, though we notice slight under-coverage for PSPP in HH.

2.3.4. Simulation 4: correctly specified mean function and misspecified propensity model

Unlike the previous simulations, in which we estimate the propensity under the correct model but use a wrongly specified prediction model, we now examine the performance of the estimators when the model for propensity is incorrectly specified but the mean function is correct. Y and R are generated under the same models as in Simulation 2. We then estimate the marginal means by regressing R on X_1, Y on X_1, X_2, and X_1X_2, and applying the DR estimators. For PSPP, we include only X_2 and X_1X_2 in the $g()$ function, as $s(P^*)$ is linear on X_1.

Results of RRMSE, relative confidence interval width, and coverage rate are displayed in figures 2.10-2.12. Under all sample sizes and situations, all methods yield similar RRMSE close to those of the correct model, though RCAL tends to have slightly higher RRMSE when sample size is 50. This simulation illustrates that when the mean model is correctly specified, there are negligible differences among the calibration methods as $\hat{y}$ has no or negligible bias. Confidence interval widths are comparable for all methods at sample sizes greater than 50. As noted before, RCAL experiences a greater variation of estimates when the number of subjects is small, but precision increases drastically once sample size reaches 100. All methods yield coverage rates close to 95%.

2.3.5. Simulations 5-8: scenarios from Kang and Schafer (2007)

We adopted scenarios from Kang and Schafer (2007) where we have a set of standard normal covariates $Z_1, \ldots, Z_4$ and an outcome variable $Y|Z_1, \ldots, Z_4 \sim N(210 + 27.4Z_1 + 13.7Z_2 + 13.7Z_3 + 13.7Z_4, 1)$. The true propensity model is $logit[Pr(R = 1|\ Z_1, \ldots, Z_4)] = -Z_1 + 0.5Z_2 - 0.25Z_3 - 0.1Z_4$, and an additional set of covariates $X_1, \ldots, X_4$ are defined as $X_1 = exp(Z_1/2)$, $X_2 = Z_2/(1+exp(Z_1)) + 10$, $X_3 = (Z_1Z_3/25 + 0.6)^3$, and $X_4 = (Z_2 + Z_4 + 20)^2$. Correctly specified mean and propensity models are fitted using a linear and logistic regression on Z, respectively, while incorrectly specified models are regressed on X. The scenario is designed such that a misspecified model is still nearly correct. We will apply the methods (CAL, MCAL, WCAL, RCAL, PSPP) to each of the four combinations of correctly and incorrectly specified mean and propensity models.

In these scenarios, we indicate a correctly specified model by C and a misspecified model by M. Thus, CM indicates a correctly specified propensity model but an incorrect mean function. Figure 2.13 displays the RRMSE of methods under the four combinations of correctly and incorrectly specified propensity and mean models. In CC and MC where the mean function is correctly specified but the propensity model may or may not be, all methods yield similar RRMSE close to those of the correct model, resulting in overlapping lines in Figure 2.13, and perform significantly better than the misspecified model. In CM, RCAL yields the lowest RRMSE at sample size of 100 and converges to those of the correct model as sample size increases. PSPP also outperforms both WCAL and CAL regardless of sample size. When both the propensity and mean models are misspecified, the DR methods yield large RRMSEs that are worse than those of the misspecified model, although RCAL tends to perform better at larger sample sizes. In both CM and MM, WCAL shows significant gains in RRMSE over its OLS counterparts CAL and MCAL (see Appendix). MCAL also yields noticeable improvements in RRMSE over CAL at sample size of 50 in CM, but differences become small afterwards.

Results for relative width of confidence intervals are shown in Figure 2.14. Comparable to RRMSE, there is little difference in interval width between PSPP, RCAL, and WCAL when the mean model is correctly specified. When the mean mode is incorrect, however, both RCAL and PSPP tend to yield lower interval widths at large sample sizes, with WCAL significantly outperforming both CAL and MCAL. All methods cover the true parameter at approximately the nominal 95% rate when either the propensity or mean model is correctly specified (Figure 2.15). When both models are wrong, all methods exhibit bias that result in significant under-coverage, particularly at larger sample sizes.

2.4. Example: asthma intervention study

We apply the methods to an asthma study conducted by the University of Michigan Schools of Public Health and Medicine, funded by the National Heart, Lung, and Blood Institute. The data consists of asthmatic children from Detroit elementary and middle schools randomized to the intervention or control group. The study aims to evaluate the effectiveness of the intervention, an education program, in reducing asthma symptoms within one year. However, since it is a well-known fact that asthma symptoms in children naturally decline as age increases, the focus of our analysis is to estimate the one-year change in asthma symptoms in the control group. Two primary measures were collected at baseline and one-year follow-up: the average number of days per month the subject experiences severe asthma symptoms, and the average number of nights per month the subject is waken up from asthma symptoms. Our goal is to estimate the mean change in days and nights with symptoms per month from baseline to one-year follow-up in the control group.

Baseline control data were collected from 696 children ages 6 to 14, out of which 437 participated in the follow-up measurements. We assume the data are MAR. Only age at baseline was found to be significantly associated with response, as subjects who remained in the study were older than those who dropped out (9.9 vs. 9.4; $P<0.001$). Moreover, baseline age is significantly associated with the outcomes given the respective baseline measurements, which are negatively associated with the outcomes. We first estimate the propensity by a logistic regression model on baseline age. Next, we apply the calibration estimators to estimate the mean one-year change in days and nights of symptoms per month separately using age and the respective baseline measurement as predictors. For PSPP, we model mean change via a spline on the estimated propensity and the baseline measurement in the g function. We then compare the DR estimates with those obtained from CC analysis

Table 2.1 displays the results for each estimator. For complete-case analysis and calibration methods, we construct 95% confidence intervals based on standard errors estimated from 200 bootstrap samples. For PSPP, we obtain the 95% credibility interval from the posterior distribution of the mean. Based on CC analysis, subjects on average experienced a decrease in both days and nights of symptoms per month from baseline to year one, with a larger decrease in days per month (-0.87 days per month vs. -0.44 nights per month). Moreover, only decrease in days of symptoms per month is significantly different from 0, as 95% confidence intervals for nights per month cover 0 for all methods. In general the estimated decrease in symptoms is smaller among DR methods than CC analysis, which is expected as older children tend to experience a greater decline in asthma symptoms and, in our sample, are more likely to remain in the study. The DR methods yield similar estimates of change in days and nights per month and there are only minor differences among CAL, MCAL, and WCAL. RCAL yields estimates closer to those of CC analysis while PSPP tends to fall in between RCAL and CAL.

The similarity of the DR estimates can be explained by the fact that the key to effective propensity weighting, an element in all DR methods in this study, is modelling variables that are associated with both outcome and response. In our example, differences between respondents and non-respondents can be well-explained by the subjects' age as age is the only variable highly associated with both response and decrease in symptoms.

2.5. Discussion

DR estimators should yield consistent estimates of the mean as long as either the propensity or mean model is correct. In our simulations we compared five DR estimators for estimating the mean with missing data. Performances of these estimators are evaluated based on their root mean square errors, 95% confidence interval widths, and their associated rate of covering the true parameter. Overall, the DR methods tended to yield better inference than the incorrect model when either the propensity or mean models are correctly specified, as promised by the DR property. However, the DR methods were less successful for sample size $n = 50$, where the asymptotic DR property is less consequential. Also, if neither the propensity nor mean models are correct, the DR methods can yield estimates of the mean that are worse than those of an incorrect regression model.

When the mean function is correctly specified, we see little difference in prediction and precision between the DR methods. In other settings, PSPP and RCAL tended to outperform the other DR methods, both in terms of RMSE and confidence coverage. When only the propensity model is correct, PSPP consistently yields better RMSE and precision than CAL, MCAL, and WCAL and outperforms RCAL when sample size is small or when the mean function has a smooth relationship with the propensity, such as in simulation 1 and in LL and HL of simulation 3. On the other hand, RCAL showed some gains in RMSE over PSPP for larger sample sizes in simulations 2 and 3. PSPP tended to have narrower confidence intervals, with coverage that was slightly anti-conservative for small sample sizes; RCAL tended to have wider confidence intervals that were conservative in terms of confidence coverage. Among the calibration methods, RCAL yields lower RMSE and interval widths than CAL, MCAL, and WCAL, perhaps a reflection of its asymptotic property of having the least variance of its class. However, RCAL was noisier and tended to have wide confidence intervals for sample size n = 50. MCAL shows small but consistent gains in prediction and precision over CAL. This is especially true at smaller sample sizes, suggesting that dividing the weighted residuals by sum of the weights provides some protection against large weights caused by small propensities. The gains over CAL are even higher in WCAL when we estimate the regression coefficients by weighted least squares, suggesting that correcting bias in the regression coefficients via WLS is more effective than calibrating estimates by the weighted average of the residuals. However, when the correlation between propensity and X is low, WLS regression based on response weighting has little impact on bias of the regression coefficients, as witnessed by the similarity between MISS and WCAL in these situations.

We estimated the variance of RCAL estimates using both bootstrap and sandwich methods. In small samples, we see a dramatic difference between bootstrap and sandwich estimates, as bootstrap typically yields larger estimates of variance. Consequently, we notice over- and under- coverage for bootstrap and sandwich methods, respectively. The difference becomes minimal in large samples. Lastly, Bayesian and likelihood-based inference for PSPP yielded similar estimates in this study, with the Bayesian method achieving better precision particularly at small sample sizes.

Although we designed our simulations to cover a wide range of possibilities involving the degree of misspecification of propensity and mean modes, conclusions of this study should not be extrapolated to conditions outside of our study. In our simulations we focused on normally distributed outcomes with a constant variance. Alternative variance structures and missing data mechanisms may be explored. The underlining assumption behind the DR methods is that the data are missing at random, and all the methods are subject to bias when the missing data mechanism is not MAR.

We have confined attention here to estimates of the overall mean. Extensions to inferences about other parameters, such as subclass means or regression coefficients, are also of interest. Weighting methods apply straightforwardly to inference about subclass means, whereas PSPP requires incorporating the subclass mean indicators in the robust model (Zhang and Little, 2008).

Figure 2.1. % increase in RMSE by method and sample size for simulation 2.1.

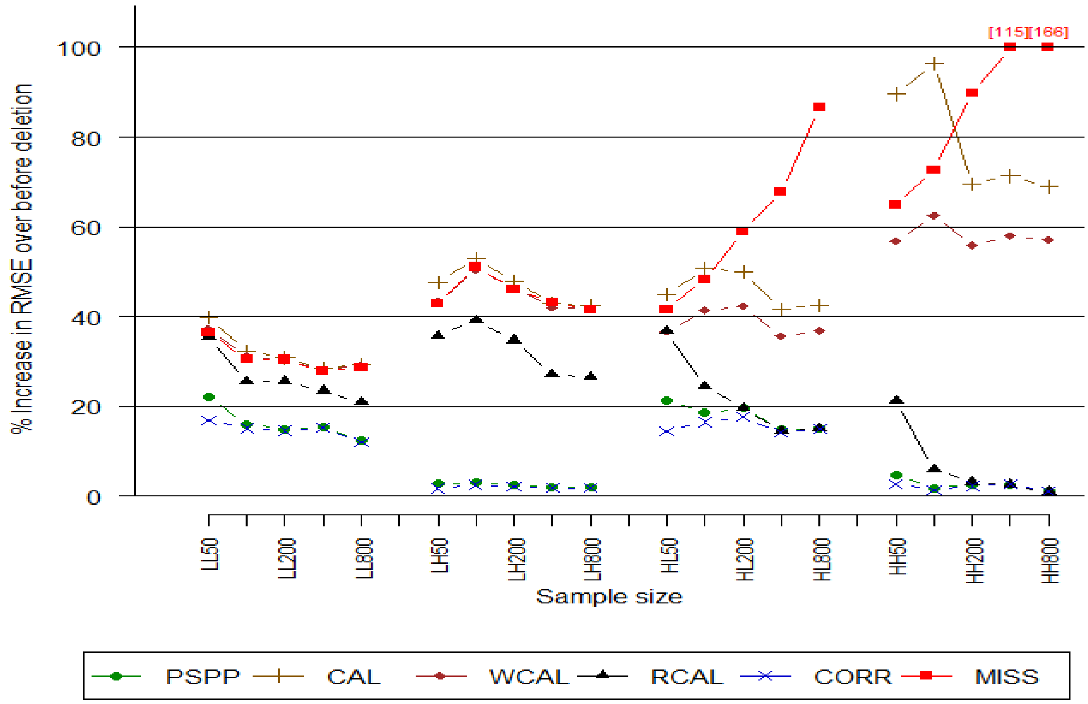


Figure 2.2. % increase in confidence interval width for simulation 2.1.

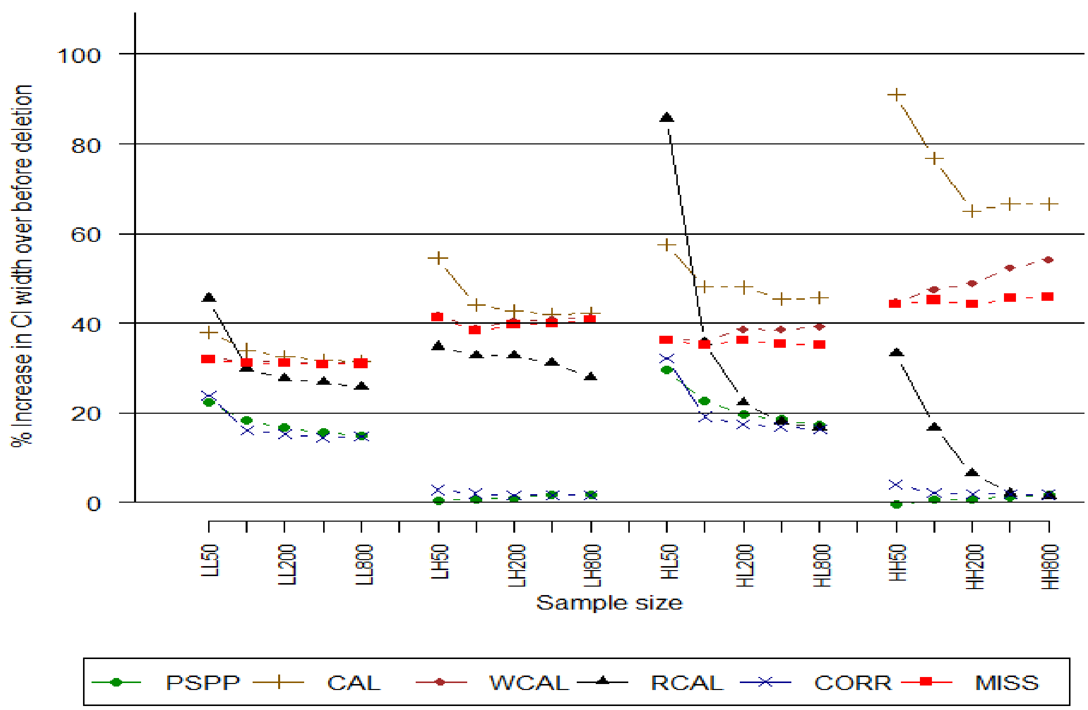


Figure 2.3. Coverage rates for simulation 2.1.

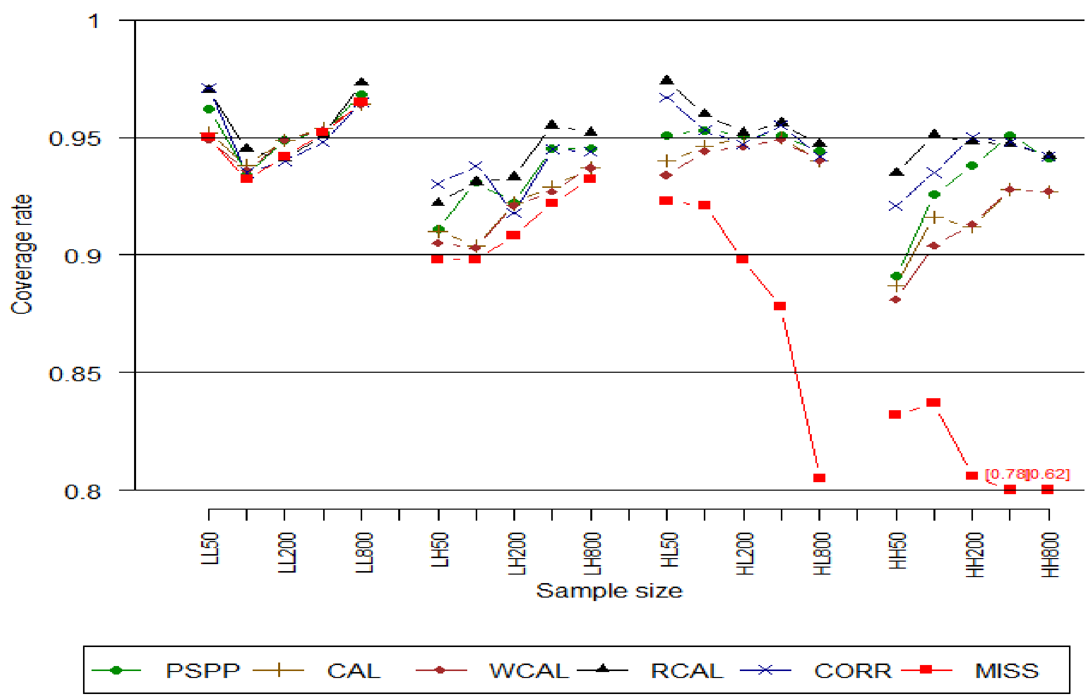


Figure 2.4. % increase in RMSE by method and sample size for simulation 2.2.

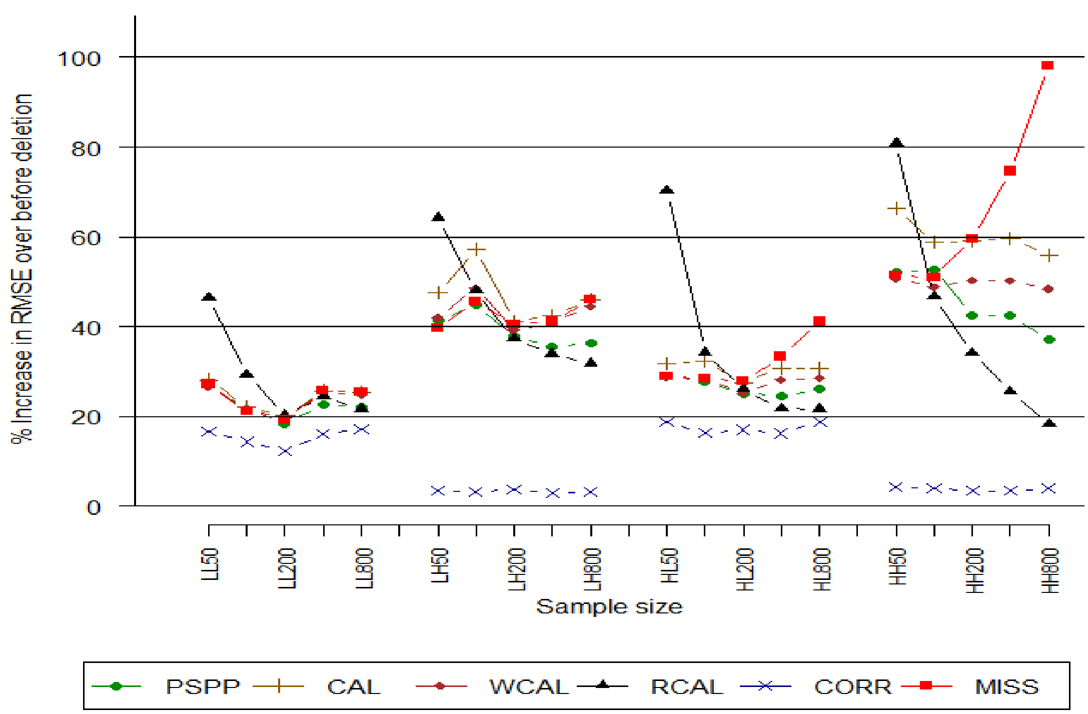


Figure 2.5. % increase in confidence interval width for simulation 2.2.

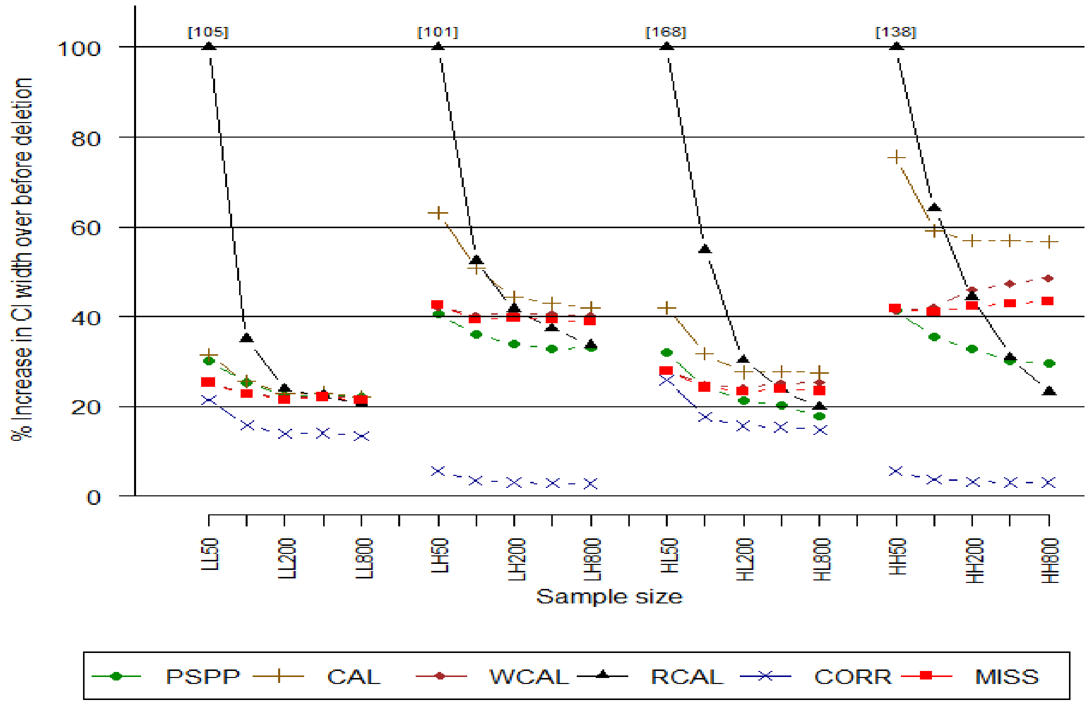


Figure 2.6. Coverage rates for simulation 2.2.

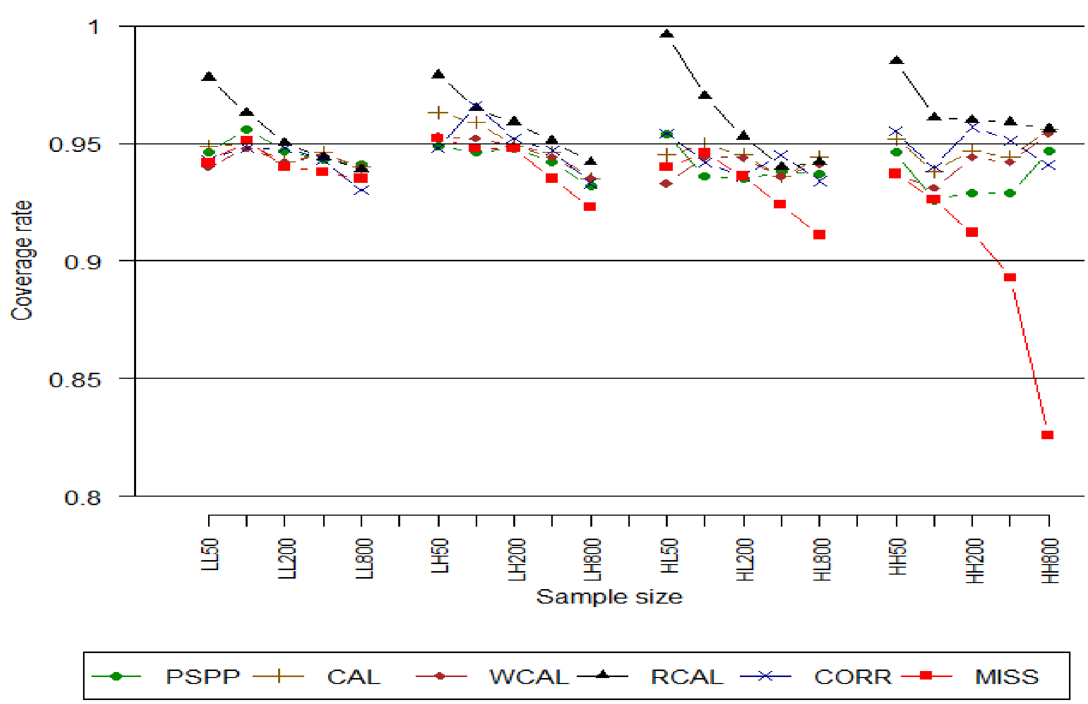


Figure 2.7. % increase in RMSE by method and sample size for simulation 2.3.

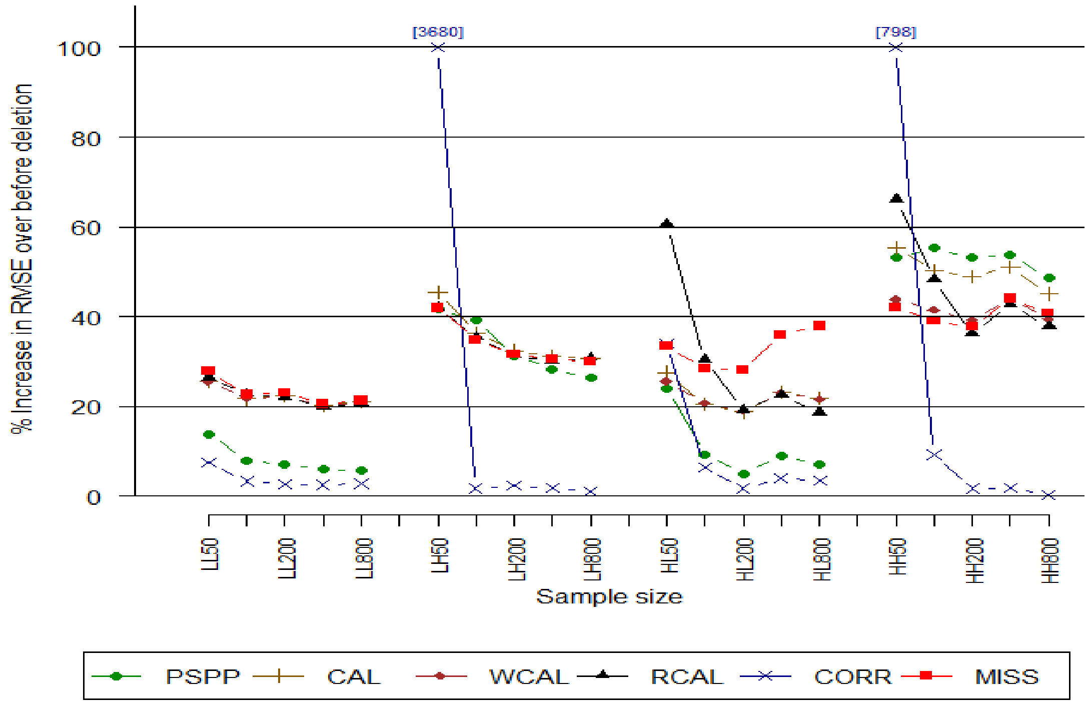


Figure 2.8. % increase in confidence interval width for simulation 2.3.

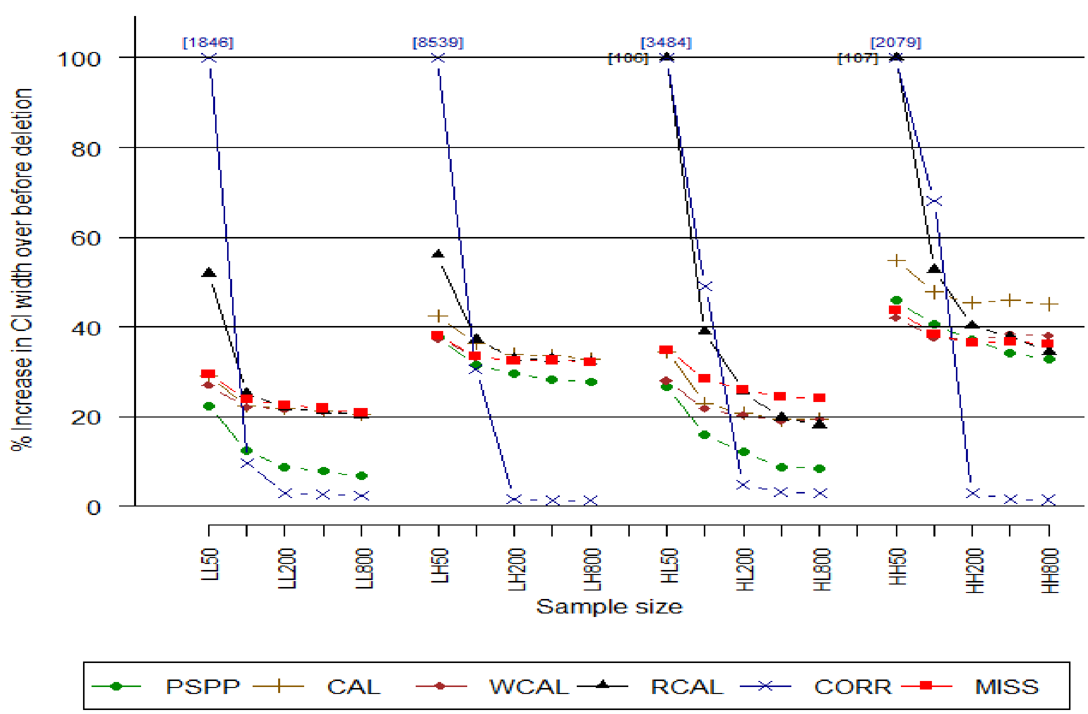


Figure 2.9. Coverage rates for simulation 2.3.

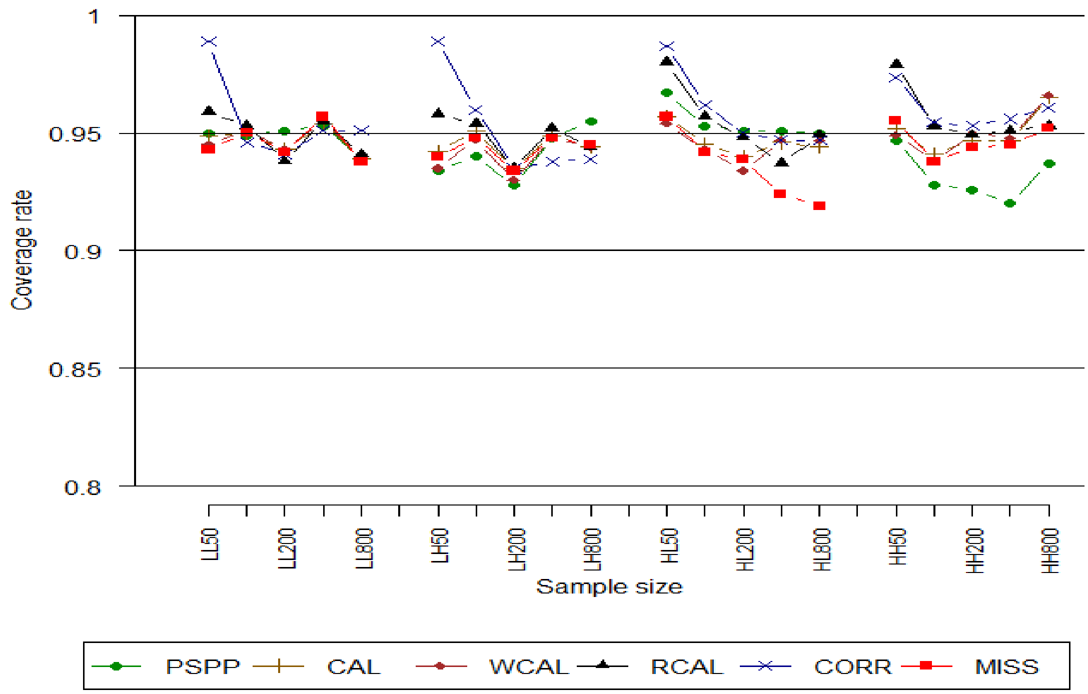


Figure 2.10. % increase in RMSE by method and sample size for simulation 2.4.

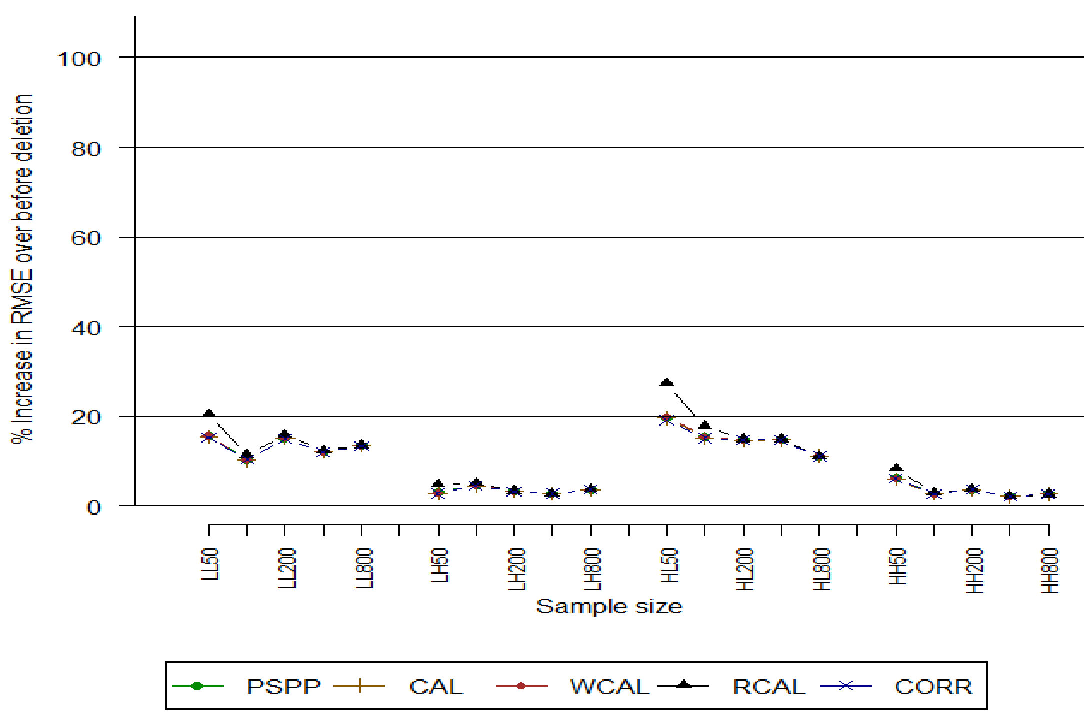


Figure 2.11. % increase in confidence interval width for simulation 2.4.

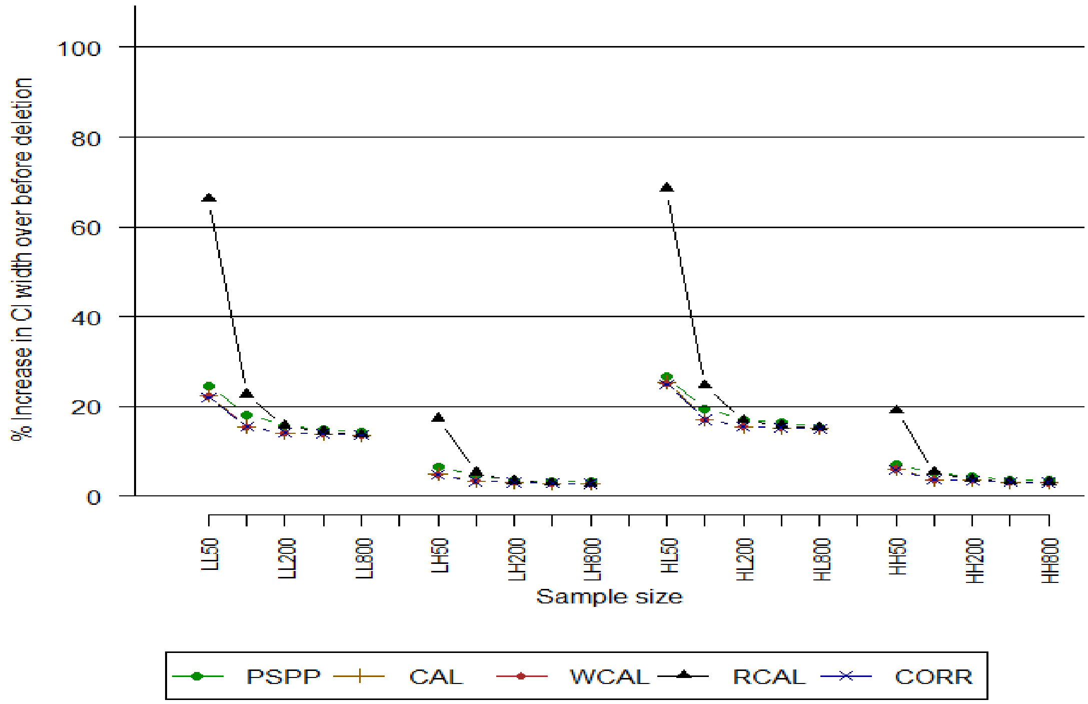


Figure 2.12. Coverage rates for simulation 2.4.

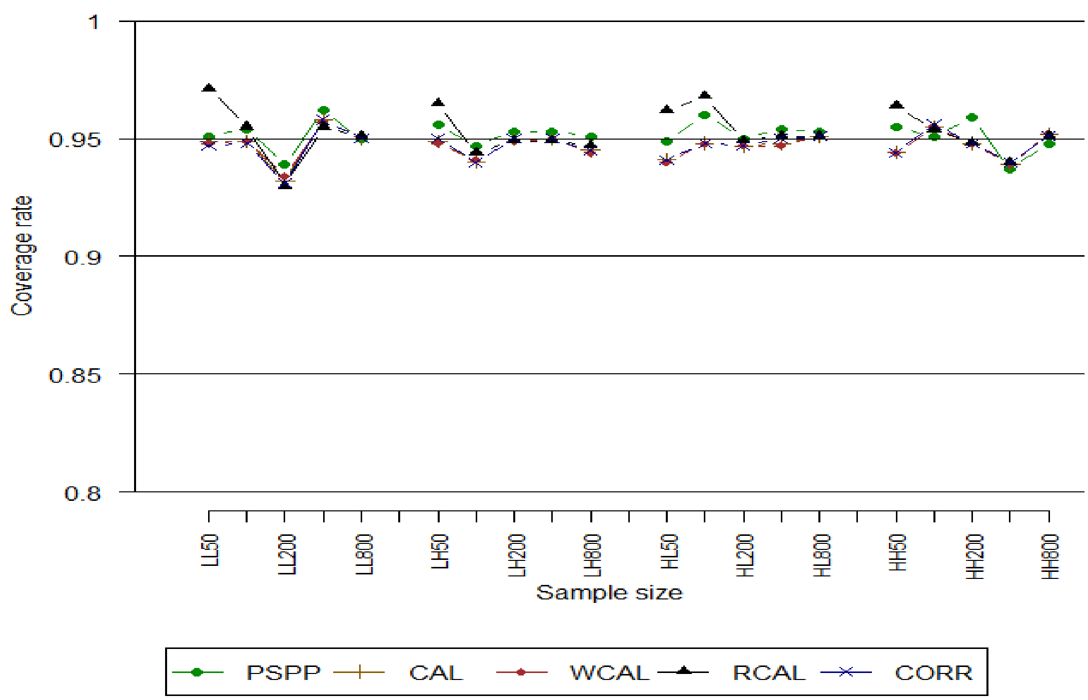


Figure 2.13. % increase in RMSE by method and sample size for simulation 2.5-8.

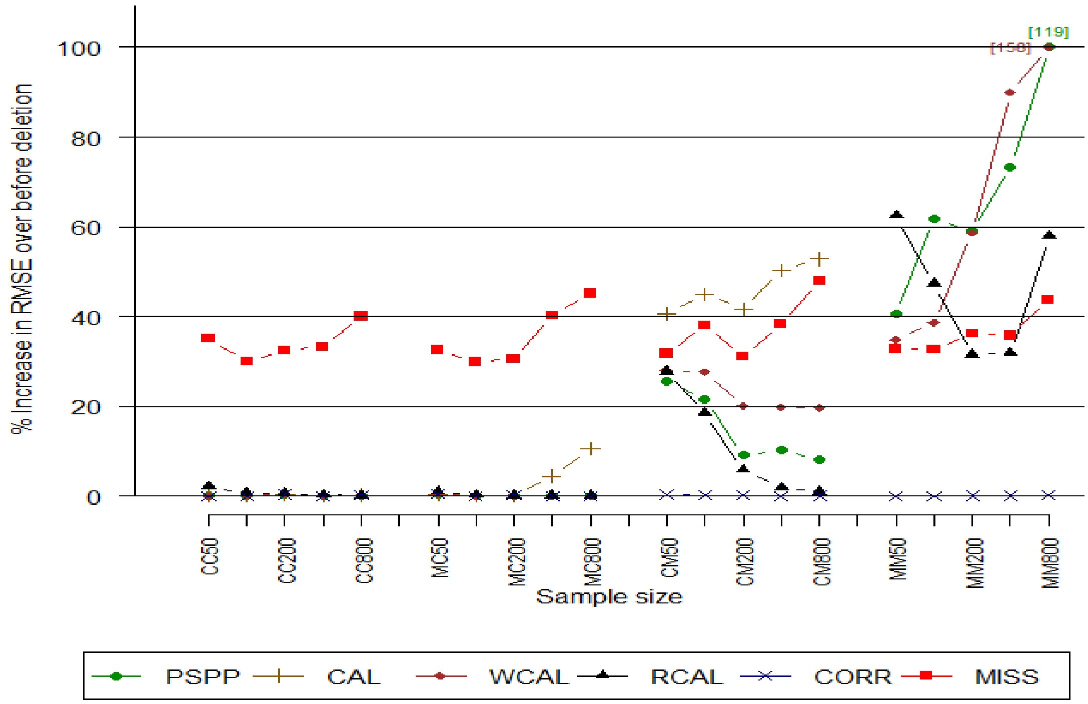


Figure 2.14. % increase in confidence interval width for simulation 2.5-8.

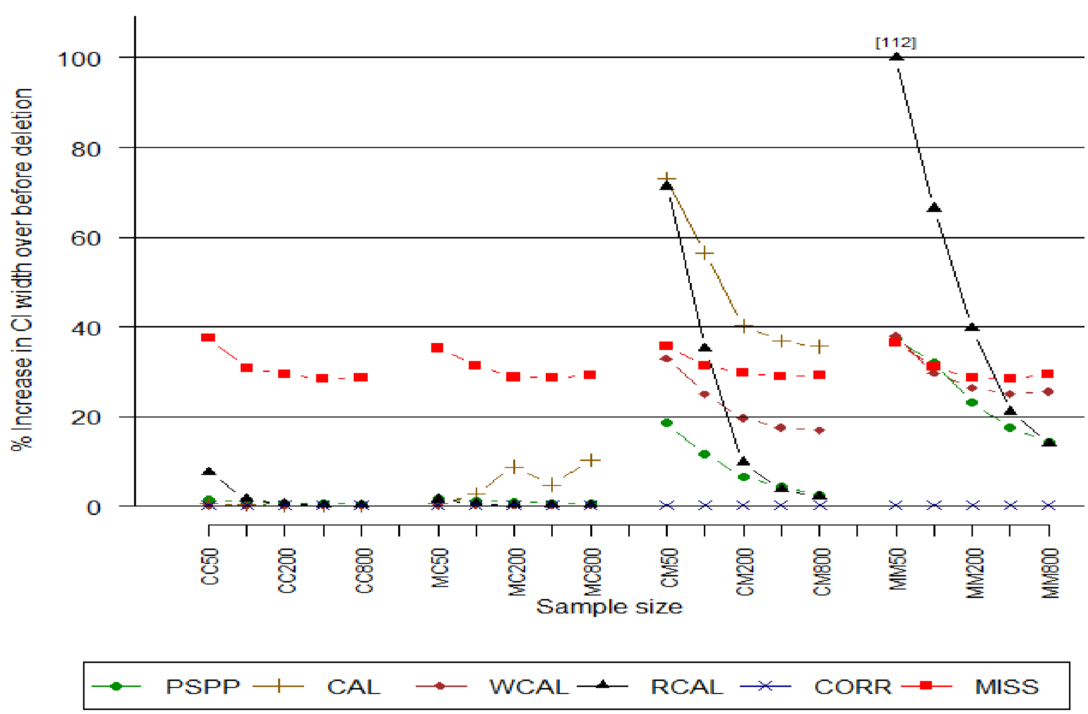


Figure 2.15. Coverage rates for simulation 2.5-8.

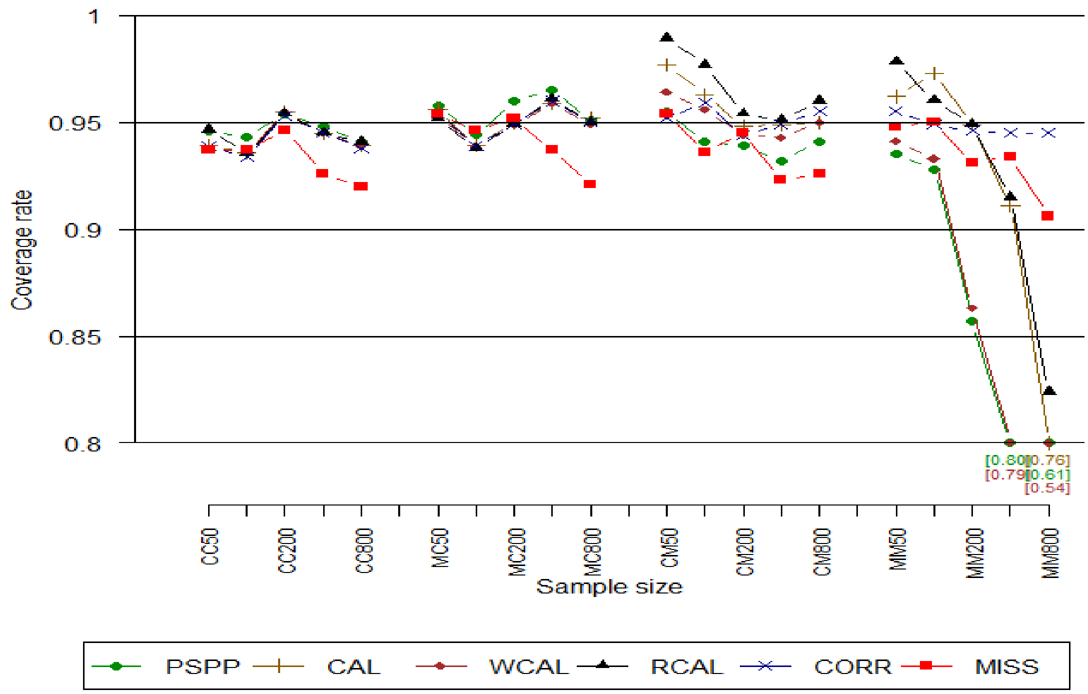


Table 2.1. One-year change in days and nights of symptoms per month

Method	Days per month		Nights per month	
	Mean	95% CI	Mean	95% CI
CC	-0.87	(-1.70, -0.05)	-0.44	(-1.09, 0.21)
PSPP	-0.79	(-1.52, -0.06)	-0.41	(-1.04, 0.23)
CAL	-0.77	(-1.45, -0.10)	-0.37	(-0.93, 0.18)
MCAL	-0.77	(-1.45, -0.10)	-0.37	(-0.93, 0.18)
WCAL	-0.77	(-1.45, -0.10)	-0.37	(-0.93, 0.18)
RCAL	-0.81	(-1.50, -0.13)	-0.43	(-0.99, 0.12)

CHAPTER III

Spline Pattern Mixture Models for Missing Data

3.1. Introduction

In this chapter we consider data where our goal is to estimate the mean of a variable Y with n_0 observed values ($\{Y_i\}$, i = 1, …, n_0), n_1 missing values ($\{Y_i\}$, $i = n_0+1, \ldots, n_0+ n_1$), when there is a set of p auxiliary variables Z_1, …, Z_p that are fully observed ($\{Z_{i1}, \ldots, Z_{ip}\}$, i = 1, …, n, $n = n_0 + n_1$). Define the response indicator R taking values 1 if Y is observed and 0 if Y is missing. It is common to use methods that assume Y is missing at random (MAR) in the sense that R is independent of Y given the observed covariates Z_1, …, Z_p. Such methods include weighting class adjustments and imputation. Our methods build on a robust MAR imputation method called penalized spline of propensity prediction (PSPP, Zhang and Little, 2009). This method (a) estimates the propensity that R = 1 given Z_1, …, Z_p based on a logistic regression of R on Z_1, …, Z_p , using all the data, and (b) imputes Y based on the regression of Y on a penalized spline of the estimated propensity, with other covariates being included parametrically if they improve the predictions.

MAR-based methods are generally biased in cases where the missingness is missing not at random (MNAR), meaning that missingness of Y depends not only on covariates Z but also on the value of Y itself. Schouten (2007) proposes a selection strategy for weighting variables that relaxes the MAR assumption. The method uses a generalized regression estimator to estimate the mean with auxiliary variables selected to minimize the maximal absolute bias under MNAR. The selection strategy, however, is based on parameters estimated under the MAR assumption and thus may be invalid if the missing data mechanism deviates heavily from MAR. Pfeffermann and Sikov (2011) propose a method for estimating the mean under MNAR by specifying models for the outcome and propensity, which is allowed depend on both the outcome and auxiliary variables. The method assumes known population totals for some or all of the auxiliary variables in the two models and estimates the model parameters in a way that takes into account the known population totals.

The bivariate normal pattern-mixture model (BNPM) of Little (1994) assumes a bivariate normal distribution for a single observed covariate X and an outcome Y within strata defined by respondents and nonrespondents, with a different mean and covariance matrix in each stratum. Parameters of BNPM are identified by assumptions about the missing data mechanism. For instance, under MAR, where missingness is assumed to depend on X but not Y, the parameters of the regression of Y on X are the same for respondents and nonrespondents; as a result, the maximum likelihood (ML) estimate for the mean of Y is the regression estimate $\hat{\mu}_Y = \bar{Y}^{(1)} + \frac{s_{XY}}{s_{XX}}(\bar{X} - \bar{X}^{(1)})$, where $\bar{X}$ is the sample mean of X, $\bar{X}^{(1)}$ is the respondent mean of X, $\bar{Y}^{(1)}$ is the respondent mean of Y, s_{XY} is the respondent covariance of X and Y, and s_{XX} is the respondent variance of X. When missingness is MNAR and is assumed to depend on Y but not X, the parameters of the regression of X on Y are the same for respondents and nonrespondents; Little (1994) shows that the resulting ML estimate of the mean of Y $\hat{\mu}_Y$ = $\bar{Y}^{(1)} + \frac{s_{YY}}{s_{XY}}(\bar{X} - \bar{X}^{(1)})$, where s_{YY} is the respondent variance of Y. The approach is easily extended to allow missingness of Y to depend on $Y^* = X + \lambda Y$ for known λ, a parameter that can then be varied in a sensitivity analysis. ML, Bayesian and multiple imputation (MI) approaches to inference for this BNPM model are described in Little (1994).

An advantage of the BNPM model is that it does not need to specify an explicit functional form for the missing data mechanism, the mechanism entering in the form of restrictions on the model parameters. The modification of MAR regression estimation to MNAR models is straightforward, as seen in the estimate of the mean of *Y* above. However, validity of the estimates depends on bivariate normality of *X* and *Y,* which is a strong assumption. For example, if *X* is normal and *Y* given *X* is normal with conditional mean a quadratic function of *X*, then the regression of *X* on *Y* is no longer linear, and ML estimates under the BNPM model are biased. In this chapter we study the impact of such forms of misspecification on inferences for the mean of *Y*.

We also propose a modification of the BNPM model, spline-BPNM (S-BPNM), which replaces a parametric linear regression by a penalized spline, extending the PSPP method (which assumes MAR) to MNAR situations; in the case where missingness depends on *Y*, we model the regression of *X* on *Y* using a flexible penalized spline, rather than assuming a linear relationship. The resulting estimate of the mean of *Y* is shown in simulations to be more robust than BNPM to the distributional relationship between *X* and *Y*. The approach can also be generalized to the case where missingness depends on $Y^* = X + \lambda Y$ for some known value of λ.

We also consider cases with more than one covariate. In that context, proxy pattern-mixture model analysis (Andridge and Little, 2011) extends the BNPM model to data with an outcome *Y* and a set of *p* observed covariates $Z_1, \ldots, Z_p$. The PPMA method replaces the set of covariates by a proxy *X*, the single best predictor of *Y* given the covariates, estimated by regressing *Y* on $Z_1, \ldots, Z_p$ for the respondents. The method then fits the pattern-mixture model in Little (1994) to *Y* and *X*. Bayesian forms of PPMA take into account the estimation of the coefficients of *Z* in the proxy variable *X*. This analysis relies on the bivariate normality assumption between the proxy *X* and *Y*, which is violated when some or all of the covariates $Z_1, \ldots, Z_p$ used to estimate *X* are not normally distributed. We propose a more flexible version of PPMA, which we call spline-PPMA (S-PPMA), which relaxes the bivariate normality assumption between the proxy and *Y* by replacing the linear regression of of *X* on *Y** implied by the bivariate normality with a penalized spline, allowing for a non-linear relationship between the variables.

We conduct simulations to examine the performance of the new S-PPMA model, and in particular to address the following questions:

1. How do inferences under S-BNPM and S-PPMA models compare with the original BNPM and PPMA methods in terms of bias, root mean squared error (RMSE) and coverage, for data sets generated under a variety of distributional assumptions?
2. How sensitive are S-BPNM and S-PPMA models to alternative assumptions about the missing data mechanism?

In the next section, we present the S-BNPM and S-PPMA models in detail. We then assess their performance in simulation studies under a variety of distributional assumptions for the auxiliary variables and missing data mechanisms.

3.2 Pattern-mixture model analysis

We consider first bivariate data on *X* and *Y*, with *X* observed for the entire sample and *Y* subject to missing data, and let *R* = 1 if *Y* is observed and *R* = 0 if *Y* is missing. Little (1994) assumes the BNPM model

$$(Y, X|\phi^{(r)}, R = r) \sim N_2\left(\begin{matrix}\mu_Y^{(r)} \\ \mu_x^{(r)}\end{matrix}, \begin{bmatrix}\sigma_{YY}^{(r)} & \sigma_{XY}^{(r)} \\ \sigma_{XY}^{(r)} & \sigma_{XX}^{(r)}\end{bmatrix}\right) \tag{3.1}$$

$$R \sim Bernoulli(\pi)$$

where $N_2(\mu, \Sigma)$ denotes the bivariate normal distribution with mean μ and covariance matrix Σ. Since we have no data on *Y* for the nonrespondents (*R* = 0), we cannot estimate all of the parameters in (4.1) for *R* = 0 without further assumptions. If assume that the missingness of *Y* depends only on *X*, we can factor the joint distribution of (*X*, *Y*, *R*) into

$$p(X, Y, R|\phi, \pi) = p(Y|X, R, \phi)p(X|R, \phi)p(R|\pi)$$

Under the bivariate normality assumption and the property that the distribution of *Y* given *X* is independent of *R*, the parameters of the regression of *Y* on *X* are the same for *R* = 1 and *R* = 0, leading to a just-identified model. Little (1994) derives the ML estimates; in particular the ML estimate for $\hat{\mu}_Y$, the mean of *Y* averaging over *R*, is

$$\hat{\mu}_Y = \bar{Y}^{(1)} + \frac{s_{XY}}{s_{XX}}(\bar{X} - \bar{X}^{(1)}) \tag{3.2}$$

Suppose now that missingness of Y depends on Y but not X. This implies that the parameters of the regression of X on Y are the same for R = 1 and R = 0, again leading to a just-identified model. The resulting ML for $\hat{\mu}_Y$ averaging over R is

$$\hat{\mu}_Y = \bar{Y}^{(1)} + \frac{s_{YY}}{s_{XY}}(\bar{X} - \bar{X}^{(1)}) \tag{3.3}$$

(Little, 1994), where s_{YY} is the respondent variance of Y.

More generally, suppose that missingness of Y depends on the value of $Y^* = X + \lambda Y$ for a given λ. Little (1994) shows that the ML estimate for $\hat{\mu}_Y$ averaging over R is then

$$\hat{\mu}_Y = \bar{Y}^{(1)} + \frac{\lambda s_{YY} + s_{XY}}{\lambda s_{XY} + s_{XX}}(\bar{X} - \bar{X}^{(1)}) \tag{3.4}$$

It is easy to see that (3.4) reduces to (3.2) when the data is MAR (λ = 0), and to (3.3) when missingness depends only on Y (λ = ∞). In practice, the data often provide no information about the value of λ. Little (1994) suggests a sensitivity analysis to capture the uncertainty about λ by estimating $\hat{\mu}_Y$ over a range of λ. Large differences in $\hat{\mu}_Y$ over λ suggest that inferences on $\hat{\mu}_Y$ are sensitive to assumptions about the missing data mechanism. Alternatively, we can specify a prior distribution that reflects the uncertainty about the choice of λ.

3.2.1 Spline pattern-mixture model

The BNPM model estimates rely heavily on the bivariate normality assumption between X and Y. For example, (X, Y) is not bivariate normal if (a) the conditional distribution of $Y|X$ is normal with $E(Y|X) = 10 + X$ and the marginal distribution of X is gamma, or (b) X is normal but the regression of Y on X is quadratic in X; in such cases the estimates from the BNPM model are potentially biased even under the correct value of λ. We propose a penalized spline regression (S-BNPM) model for *X* and *Y* that relaxes the bivariate normality assumption.

Suppose that missingness depends on the value of $Y^* = X + \lambda Y$ for some known $\lambda > 0$. The conditional distribution of $X|Y^*$ is then the same for respondents and nonrespondents, The S-BNPM method creates multiple imputations of the missing values of Y^* (and hence $Y = (Y^* - X) / \lambda$) so that the regression of *X* of Y^* for respondents (where Y^* is observed) and nonrespondents (where Y^* is imputed) follows the same spline regression model:

$$X = \beta_0 + \beta_1 Y^* + \sum_{k=1}^{K} \gamma_k (Y^* - \kappa_k)_+ + \varepsilon \quad (3.5)$$

$$\varepsilon \sim N(0, \sigma^2)$$

$$\gamma_k \sim N(0, \tau^2)$$

where a_+ = a if $a>0$ and $a_+=0$ otherwise, and $\kappa_1 < \ldots < \kappa_K$ are *K* equally spaced knots. The model may be fitted to the respondent data using a linear mixed model, treating the splines as random effects. Here, we adopt a Bayesian approach by assigning a uniform prior for β and inverse gamma (10^{-5}, 10^{-5}) priors for σ^2 and τ^2, and obtain draws from their posterior distributions using a Gibbs sampler (See Appendix for details of the algorithm).

We then adopt a hot deck procedure (Andridge and Little, 2010) to impute the missing values of Y^*, where the missing value of Y^* is imputed with the observed value of a matched donor with *X* and Y^* observed. The method involves the following steps:

1. Draw *B* values of Y^* for each nonrespondent from the distribution of $Y^*|X, R = 0$, estimated under the BNPM model. This results in a pool of $n_1 * B$ values of Y^*

($\{Y_p^*\}$, p = 1, …, n_1*B). In the simulations in Section 3.3 a value of B = 100 is sufficient.

2. Given each Y_p^* in the pool, draw a value X_p from the posterior predictive distribution of $X|Y^*$ in (3.5), with parameters estimated from respondents. This results in a set of pairs of ($\{X_p, Y_p^*\}$, p = 1, …, n_1*B) that form our donor pool.
3. For each nonrespondent j, choose a pair (X_k, Y_k^*) from the donor pool ($\{X_p, Y_p^*\}$, p = 1, …, n_1*B) with the closest value X_k to X_j, and impute Y_j^* = Y_k^* (hence Y_j = (Y_k^* - X_j) / λ) from that pair.
4. Repeat steps 2-3 above for 2000 iterations, deleting the first 1000 as burn-in and using every other 10 iterations to create D = 100 multiply-imputed data sets with values of Y imputed.

Using multiple imputation combining rules (Rubin, 1987) we obtain $\hat{\mu}_Y$ and its variance

$$\hat{\mu}_Y = \hat{\mu}_D = \frac{1}{D}\sum_{d=1}^{D} \hat{\mu}_d \tag{3.6}$$

$$Var(\hat{\mu}_Y) = \frac{1}{D}\sum_{d=1}^{D} W_d + \frac{D+1}{D(D-1)}\sum_{d=1}^{D}(\hat{\mu}_d - \bar{\mu}_D)^2 \tag{3.7}$$

where $\hat{\mu}_d$ and W_d are the estimated marginal mean and variance in the d^{th} imputed data set, respectively. For the MAR assumption of λ = 0, we apply a Bayesian form of the PSPP method (Zhang and Little, 2009). Specifically, we regress Y on a spline of X using the complete cases and impute Y by drawing directly from its predictive posterior distribution in (3.5) given the observed $X^{(1)}$ for each iteration of the Gibbs algorithm.

The underlying rationale of the procedures is as follows. Since the unobserved *Y** is a covariate in our spline model (3.5), we cannot impute *Y** by drawing directly from a model. Thus we first create a donor pool of values ($\{Y^*_{ib}\}$, b = 1, …, B, $i = n_0 + 1, \ldots, n_0 + n_1$) as draws from the BNPM model. For each donor in the pool, we create a corresponding value of *X* as a prediction from the spline model (3.5). We then match each incomplete case to a member of the donor pool with a similar value of *X*, and impute for that case the corresponding value of *Y** from the donor. When the data are normal, the "hot-deck" matching step has little effect on the final imputations of *Y**. However, when data deviates from normality, the pairs (*X*, *Y**) resulting from the hot-deck respect the spline model (3.5) and hence should improve on the imputations from the BNPM model, which incorrectly assume a linear relationship between *X* and *Y**. In practice, we create multiple initial draws of *Y** for each nonrespondent, as a large value of *B* allows flexibility in the nonlinearity adjustment by S-BNPM and ensures a close match with the donors for every observed *X*. In the following examples we find a value of B = 100 to be sufficient to ensure a near-identical match in *X*.

As in the original BNPM model, the S-BNPM model utilizes the fact that, conditional on the variables contributing to missingness, the regression model parameters are the same for both respondents and nonrespondents. However, the penalized spline improves robustness of the pattern mixture model by allowing us to model nonlinearity in the relationship between *X* and *Y*. As suggested in Little (1994), inferences for $\hat{\mu}_Y$ should be displayed for a range of potential values of λ to account for uncertainty about the true value of λ and to assess sensitivity of inferences to the choice of λ.

3.2.2 More than one covariate: extensions of proxy pattern-mixture model analysis

There may be multiple observed covariates Z_1, …, Z_p that are predictive of $\hat{\mu}_Y$. Andridge and Little (2011) proposed an extension of the pattern-mixture model analysis by taking *X* as a proxy obtained by regressing *Y* on the set of Z_1, …, Z_p and replacing the set of covariates by *X*, the estimated best predictor of *Y* given Z_1, …, Z_p. Proxy pattern-mixture model analysis (PPMA) then estimates $\hat{\mu}_Y$ by applying the pattern-mixture model in Little (1994) to *X* and *Y*. The advantage of reducing Z_1, …, Z_p to *X* is simplicity: modelling departures from MAR under one sensitivity parameter λ is much simpler than specifying a model with *p* sensitivity parameters for each of Z_1, …, Z_p. Moreover, should missingness depend on some other combination of *Z* (e.g. $W = \alpha Z$), estimates for the mean of *Y* are still approximately unbiased since *Y* is independent of *W* given *X*.

Andridge and Little (2011) showed that the uncertainty of the estimates of $\hat{\mu}_Y$ depends largely on the degree of correlation between the proxy *X* and *Y* as well as the degree of similarity between respondents and nonrespondents with respect to the value of *X*. When *X* and *Y* are highly correlated and the values of *X* are similar for respondents and nonrespondents, information on missing values of *Y* and evidence on the lack of response bias are both strong, resulting in estimates of $\hat{\mu}_Y$ with high precision. However, if *X* and *Y* are weakly correlated and the values of *X* are much different for respondents and nonrespondents, we have strong evidence for response bias with little information on the missing values of *Y*, resulting in estimates of $\hat{\mu}_Y$ with high uncertainty.

3.2.3 *Spline-proxy pattern-mixture model*

As in the bivariate case, validity of the proxy pattern-mixture model proposed by Andridge and Little (2001) when data are MNAR relies on the assumption of bivariate normality between the proxy *X* and *Y*, which is violated when some or all of the Z_1, …, Z_p used to obtain *X* are not normally distributed. Suppose, for example, *Z* is a fully observed standard normal variable and *Y* given *Z* is normal with mean $Z + Z^2$. Let *X* be a proxy from the regression of *Y* on *Z* and Z^2. When the data is MAR, *X* is an unbiased predictor of *Y*, hence estimates from the pattern-mixture model under λ = 0 are unbiased. However, when missingness depends on *Y*, the resulting proxy *X* is no longer an unbiased predictor of *Y* since the regression coefficients in the regression of *Y* on *Z* and Z^2 based on the respondents are biased for the nonrespondents. Since *X* is some function of *Z* and Z^2 which is not normally distributed, the assumption of bivariate normality, hence linearity, with *Y* fails, resulting in biased estimates for all values of λ.

We propose a modification of the proxy pattern-mixture model that relaxes the assumption of bivariate normality between *X* and *Y*. Suppose, as before, *X* is the predicted value of *Y* based on regression of *Y* on Z_1, …, Z_p for the complete cases, and that missingness depends on the value of *Y**. The conditional distribution of *X* given *Y** is independent of *R* and the regression coefficients of *X* on *Y** are the same for both respondents and nonrespondents. The model proposed in Andridge and Little (2011) assumes linearity between X and Y, and hence Y^*, which as discussed may not be appropriate when X and Y are not bivariate normal. Thus, we propose a spline-proxy pattern mixture model analysis (S-PPMA) to describe the relationship between X and Y. Under S-PPMA, we first estimate the proxy based on a complete-case regression of Y on Z_1, …, Z_p as in Andridge and Little (2011), and set X as the predicted value of Y from this regression. Then, we apply a penalized spline model to X and Y and estimate $\hat{\mu}_Y$ as discussed in section 2.1. As in the bivariate model, we believe S-PPMA will further enhance the robustness of PPMA by relaxing the bivariate normality assumption.

In the next section, we describe simulation studies to assess the performance of S-PPMA under various distributions of Z_1, …, Z_p, Y, and missing data mechanisms. For comparison we include estimates from the proxy pattern-mixture model proposed in Andridge and Little (2011).

3.3. Simulation studies

We assess the performance of S-PPMA for inferences about the mean of Y with respect to average bias, root mean square error, 95% confidence interval width, and rate of confidence interval non-coverage over 1000 replications and six scenarios. For each replication, we construct 95% confidence intervals and estimate the non-coverage rate as the proportion of the 1000 confidence intervals that do not cover the true value, where 95% CI = ($\hat{\mu}_Y - t_{n\text{-}1,0.975}\sqrt{Var(\hat{\mu}_Y)}$, $\hat{\mu}_Y + t_{n\text{-}1,0.975}\sqrt{Var(\hat{\mu}_Y)}$), $t_{n\text{-}1,0.975}$ is the 97.5th percentile of the t-distribution with n-1 degrees of freedom, and $Var(\hat{\mu}_Y)$ is the estimated variance of the mean from (3.7). Confidence interval widths (CIW) are computed as CIW = $2^*t_{n\text{-}1,0.975}\sqrt{Var(\hat{\mu}_Y)}$. For all simulations, we set sample sizes of n = 100 and n = 400.

For the first scenario, we assume bivariate normal data of X and Y and compare estimates of the mean of Y under BNPM and S-BNPM models. For scenarios 2-5, we assume a set of fully observed covariates Z_1, …, Z_p. Here, we first obtain the proxy X from a correctly specified regression of Y on Z_1, …, Z_p using the respondent sample. Then, we estimate the mean of Y using three methods

1. We estimate apply the S-PPMA model to X and Y using a penalized spline in (3.5). (S-PPMA)
2. We assume bivariate normality between X and Y and estimate $\hat{\mu}_Y$ via maximum likelihood in (3.4) as originally proposed in Andridge and Little (2011). Variance is estimated using 200 bootstrap samples. (PPMA-ML).
3. We assume bivariate normality between X and Y and draw $\hat{\mu}_Y$ from its posterior distribution as described in Little (1994). 95% credibility intervals and coverage are based on draws from the posterior distribution. (PPMA-BAYES)

Let λ_T be the true, unobservable value of λ generating missing data, and let λ_A be the assumed value of λ in our models. For each scenario, we simulate nonresponse using λ_T = 0, 1 and ∞. To assess sensitivity of inferences to λ_A, we produce estimates under λ_A = 0, 1 and ∞ for each value of λ_T, one of which corresponds to the true underlying value of λ_T. While inferences under additional values of λ_A may be explored, we chose these three values to capture a range of potential missing data mechanisms. In the following section, only results for which $\lambda_A = \lambda_T$ are shown (for rest, see Appendix).

3.3.1. *Scenario 1: bivariate normal data*

We assume a fully observed covariate X and a Y that is bivariate normal with X and subject to missingness. The data is generated under the following pattern-mixture model with a sample size of $n = 400$:

$$R \sim Bernoulli(0.5)$$

$$X, Y | R = 1 \sim N_2\left(\begin{matrix}0\\0\end{matrix}, \begin{bmatrix}1 & 0.5\\0.5 & 1\end{bmatrix}\right)$$

$$X | R = 0 \sim N(1, 1)$$

In this and all subsequent scenarios, nonresponse rates are approximately 50%. For simplicity we only display results at $n = 400$, as results for $n = 100$ are generally similar. Figure 3.1 displays performances of each estimator in terms of average bias, root mean squared error (RMSE), 95% confidence interval width (CIW), and its corresponding non-coverage rate out of 1000 replications when $\lambda_A = \lambda_T$. In the figure, the true missingness of Y depends on $X + \lambda_T Y$ for $\lambda_T = 0$, 1, and ∞. Results show little differences between the methods in bias, RMSE, and CIW regardless of λ_A in all values of λ_T (results for $\lambda_A \neq \lambda_T$ in Appendix). As expected, when $\lambda_A = \lambda_T$, all estimates are approximately unbiased and non-coverages are near the nominal 5%, as BNPM is the correct model for the data. Moreover, CIW increases as λ_T increases, reflecting a rise in uncertainty as a result of nonresponse due to Y. We notice that the CIW for S-BNPM at $\lambda_A = \infty$ is narrower than that for BNPM under both ML and Bayes for all values of λ_T. This may be due to a small correlation of 0.5 between *X* and *Y*, which may lead to a large value of $\frac{s_{YY}}{s_{XY}}$ in (3.3) and consequently an extreme $\hat{\mu}_Y$. In S-BNPM, the process of generating multiple initial draws of the missing *Y* and matching on the donor pool based on predictions from the spline model helps to alleviate this problem as draws of *X* from extreme values of *Y* are less likely to be matched to observed values of *X*, leading to less extreme imputations in this particular scenario.

3.3.2. Scenario 2: bivariate non-normal data

Suppose *X* is a fully observed, gamma-distributed covariate and *Y* is normal conditional on *X* and is subject to missingness. We generate the data under a selection model with a sample size of $n = 400$:

$$X \sim Gamma(1, 1/4)$$

$$Y|X \sim N(10 + X, 1)$$

We generate missing value of *Y* under the following models to reflect both MAR and MNAR scenarios, assuming an unobserved latent variable *U*

A. $U|X, Y \sim N(-1.5 + 0.5X, 1)$ $(\lambda_T = 0)$

B. $U|X, Y \sim N(-2.5 + 0.15(X + Y), 1)$ $(\lambda_T = 1)$

C. $U|X, Y \sim N(-3.5 + 0.25Y, 1)$ $(\lambda_T = \infty)$

where *Y* is missing if $U > 0$ and observed otherwise.

In this scenario we include estimates from the true model, which models *Y** on *U* and *X* for $\lambda > 0$, since *Y** and *U* are bivariate normal conditional on *X*. Since *U* is unobserved, we estimate *U* and produce posterior draws of the missing *Y** iteratively by the following steps, where $\hat{Y}*^{(R=0,i=1)}$ are imputations of *Y** at the i^{th} iteration and $\hat{Y}*^{(i=1)}$ are the observed and imputed values for the whole sample:

1. Initialize values of $\hat{Y}*^{(R=0,i=1)}$ and $\hat{U}^{(i=1)}$ by setting $\hat{Y}*^{(R=0,i=1)}$ as predictions from the regression of $\hat{Y}*^{(R=1)}|X^{(R=1)}$, and draw $\hat{U}^{(i=1)}$ from a normal distribution with variance 1 and mean $Z_{\hat{\pi}}$ - $\bar{Y}*^{(i=1)}$ + $\hat{Y}*^{(i=1)}$, where $\hat{\pi}$ is the nonresponse rate, Z_{α} is the α^{th} percentile of the standard normal distribution, and $\bar{Y}*^{(i=1)}$ is the mean combining the observed $Y*^{(R=1)}$ and the initialized $\hat{Y}*^{(R=0,i=1)}$. For respondents, positive values of $\hat{U}^{(i=1)}$ are discarded and redrawn until all values are negative. Likewise for nonrespondents, we discard and redraw negative values of $\hat{U}^{(i=1)}$.
2. At the i^{th} iteration, draw $\hat{Y}*^{(R=0,\ i)}\,|\,\hat{U}^{(R=0,\ i-1)}$, $X^{(R=0)}$ from the posterior predictive distribution based on a linear regression of $\hat{Y}*^{(i-1)}|\hat{U}^{(i-1)}$, *X* on the entire imputed sample with values $\hat{Y}*^{(R=0,\ i-1)}$ and $\hat{U}^{(i-1)}$ drawn from the previous iteration.
3. Obtain posterior predictive draws of $\hat{U}^{(i)}|\hat{Y}*^{(i)}$ based on a linear regression model of $\hat{U}^{(i-1)}|\hat{Y}*^{(i)}$ for the entire sample. We again discard and redraw all positive values of $\hat{U}^{(i)}$ for respondents and negative values of $\hat{U}^{(i)}$ for nonrespondents.

4. Repeat steps 2 and 3 over 1000 iterations, discarding the first 100 as burn in. We then apply (3.6) and (3.7) over the 900 imputed sets of $\hat{Y}^*$ to estimate the mean and variance.

For λ = 0, we impute the missing *Y* based on posterior predictive draws from the regression of *Y* on *X* on the complete cases.

Figure 3.2 displays results under $\lambda_A = \lambda_T$ for $\lambda_T = 0$, 1, and ∞. When $\lambda_T = 0$, all methods are unbiased, with S-BNPM having slightly higher RMSE and more conservative 95% confidence intervals. Since data is MAR and $Y|X$ is normal with a mean that is linear on *X*, the BNPM model is correctly specified and thus it is not surprising that its estimates are unbiased and have better precision than S-BNPM. However, when $\lambda_T = 1$, linearity assumptions for $X|Y^*$ are violated, and consequently we see bias and under-coverage by BNPM. Here, S-BNPM shows reductions in bias and to a lesser extent RMSE, and achieves near nominal 5% non-coverage with a minor penalty in RMSE and precision compared to the true model. The heavier the data deviates from MAR, the higher the gains in bias and RMSE from S-BNPM, as evident in the results under $\lambda_T = \infty$. S-BNPM shows a noticeable improvement in RMSE over BNPM and still yields close to nominal non-coverage. Robustness to normality, however, comes at the price of precision, as S-BNPM tends to yield wider intervals than both BNPM and the true model.

3.3.3. Scenario 3: set of normal Z's

In this scenario, we assume a set of covariates that are normally distributed. Let Z_1, Z_2, Z_3 be fully observed covariates with distributions

$$Z_1 \sim N(0, 1)$$

$$Z_2 \sim N(0, 1)$$

$$Z_3 \sim N(0, 1)$$

$$Y|Z_1, Z_2, Z_3 \sim N(15 + Z_1 + 2Z_2 + Z_3, 1)$$

Let *Y* be missing under the following logistic models

A. $Logit[Pr(R = 0)] = 0.5(Z_1 + 2Z_2 + Z_3)$ $(\lambda_T = 0)$

B. $Logit[Pr(R = 0)] = -3.5 + 0.25(0.98Z_1 + 1.95Z_2 + 0.98Z_3 + Y)$ $(\lambda_T = 1)$

C. $Logit[Pr(R = 0)] = -7.5 + 0.5Y$ $(\lambda_T = \infty)$

D. $Logit[Pr(R = 0)] = 2Z_2$

E. $Logit[Pr(R = 0)] = -7.5 + 0.5(2Z_2 + Y)$

For each missing data mechanism, we obtain the proxy *X* by regressing *Y* on Z_1, Z_2, and Z_3 apply the estimators to *X* and *Y*. Figure 3.3 shows results for $\lambda_A = \lambda_T$, with $\lambda_T = 0$, 1, and ∞ (see Appendix for rest of results). In addition there are two nonresponse mechanisms, D and E, that do not correspond to any λ_T. When $\lambda_T = 0$, *Y* is MAR, $\lambda_A = 0$ is the correct assumption about nonresponse and as a result all estimators are approximately unbiased and yield similar RMSE, confidence interval widths, and near-nominal non-coverage of 5%. For values of $\lambda_A = 1$ and ∞ when $\lambda_T = 0$, all three methods exhibit bias, with negligible differences in RMSE, CIW, and non-coverage. Similarly when $\lambda_T = 1$ and ∞, values of λ_A such that $\lambda_A = \lambda_T$ result in negligible bias and near nominal non-coverage for all estimators. For values of λ_A such that $\lambda_A \neq \lambda_T$, all methods are biased with higher than nominal non-coverage, as expected given that the assumptions about nonresponse are wrong. Results for mechanism D (see Appendix) are generally similar to those of A, where $\lambda_T = 0$. Here, all methods have negligible bias and nominal non-coverage at $\lambda_A = 0$ and yield similar RMSE and CIW at all values of λ_A. In mechanism E, all methods have minor bias at $\lambda_A = 1$ and cover the true mean at a rate close to 95%, with minor differences in RMSE and CIW regardless of λ_A. In this scenario, nonresponse mechanisms D and E do not deviate much from mechanisms A and B, which explains the similarity of results.

This scenario assumes that all auxiliary variables are normally distributed, resulting in a proxy X that is normal and linear with Y regardless of the nonresponse mechanism. As such the methods in Andridge and Little (2011) produce valid estimates under the correct value of λ_A. We again notice that S-PPMA tends to yield slightly more conservative confidence intervals than PPMA, which suggests there is some penalty in precision from fitting a more robust model when normality assumptions are met.

3.3.4. Scenario 4: varying distributions of Z

Let Z_1, Z_2, Z_3 be fully observed covariates with the following distributions

$$Z_1 \sim N(0, 1)$$

$$Z_2 \sim GAMMA(1, 1)$$

$$Z_3 \sim BERNOULLI(0.5)$$

$$Y | Z_1, Z_2, Z_3 \sim N(10 + Z_1 + 4Z_2 + Z_3, 1)$$

Let Y be missing under the following logistic models simulating different response mechanisms

A. $Logit[Pr(R = 0)] = -2 + 0.5(Z_1 + 4Z_2 + Z_3)$ ($\lambda_T = 0$)

B. $Logit[Pr(R = 0)] = -4.5 + 0.25(0.98Z_1 + 3.9Z_2 + 0.98Z_3 + Y)$ ($\lambda_T = 1$)

C. $Logit[Pr(R = 0)] = -7 + 0.5Y$ ($\lambda_T = \infty$)

D. $Logit[Pr(R = 0)] = -1 + Z_2$

E. $Logit[Pr(R = 0)] = -4 + 0.25(2Z_2 + Y)$

We obtain the proxy by regressing Y on Z_1, Z_2, and Z_3 using respondent data and apply the estimators under $\lambda_A = 0$, 1 and ∞. Results for which $\lambda_A = \lambda_T$ are shown in Figure 3.4 (see Appendix for rest of results). Mechanisms D and E do not correspond to any value of λ_T. In this scenario we vary the distributions of the auxiliary variables and the conditional mean of Y given Z_1, Z_2, and Z_3 is dominated by a gamma distributed Z_2. For $\lambda_T = 0$ where Y is MAR, all three methods yield approximately unbiased means with close to nominal non-coverage when the correct value of $\lambda_A = 0$ is used. Under the incorrect values of $\lambda_A = 1$ and ∞, however, the S-PPMA has lower bias, lower RMSE, and lower non-coverage rate than the linear models albeit with more conservative confidence intervals.

For $\lambda_T = 1$ and ∞, the PPMA estimates exhibit small bias even when $\lambda_A = \lambda_T$, most likely as a result of lack of linearity between X and Y due to MNAR and some of the auxiliary variables being non-normal. The S-PPMA estimates at the correct λ_A show low bias and non-coverages close to 5%, which may be explained by the spline's ability to model nonlinearity between X and Y. It is worth noting, however, that despite the bias PPMA still achieves good coverage at $\lambda_A = \lambda_T = 1$. In terms of RMSE, S-PPMA has no noticeable gains over PPMA under $\lambda_A = \lambda_T = 1$, and larger gains when $\lambda_A = \lambda_T = \infty$. This suggests that as dependence of nonresponse on Y increases, the degree of nonlinearity adjustment by the penalized spline increases. Robustness to λ_T comes at the expense of precision, as the penalized spline yields wider intervals under all values of λ_A for any λ_T. However, it is important to note that values of Y tend to be much lower for respondents than nonrespondents as a result of the nonresponse mechanism, which leads to sparse data and extrapolation at higher values of Y. Thus, wider interval widths by the spline may be a reflection of uncertainty in imputing the missing values by extrapolating a nonlinear model. For mechanism D (see Appendix), there are no significant differences in RMSE and CIW regardless of λ_A, with negligible bias at $\lambda_A = 0$ and close to nominal coverage at both $\lambda_A = 0$ and 1 for all methods. In mechanism E, both S-BNPM and BNPM yield similar estimates with nominal non-coverage at $\lambda_A = 1$.

3.3.5. Scenario 5: quadratic term in mean of Y

Let Z_1 and Z_2 be fully observed covariates with the following distributions

$$Z_1 \sim N(0, 1)$$

$$Z_2 \sim N(0, 1)$$

$$Y|Z_1, Z_2 \sim N(10 + Z_1 + Z_2 + 2Z_2^2, 1)$$

Let Y be missing under the following mechanisms

A. $Logit[Pr(R = 0)] = -1 + 0.5(Z_1 + Z_2 + 2Z_2^2)$ ($\lambda_T = 0$)

B. $Logit[Pr(R = 0)] = -3 + 0.25(0.97Z_1 + 0.97Z_2 + 1.95Z_2^2 + Y)$ ($\lambda_T = 1$)

C. $Logit[Pr(R = 0)] = -6 + 0.5Y$ ($\lambda_T = \infty$)

D. $Logit[Pr(R = 0)] = 4Z_2$

E. $Logit[Pr(R = 0)] = -5.5 + 0.5(4Z_2 + Y)$

We estimate the proxy X by regressing Y on Z_1, Z_2, and Z_2^2 using the complete cases and apply the estimators under the different values of λ_A. Here we introduce a quadratic term in the conditional mean of Y. For $\lambda_T = 0$, when data is MAR, the estimated proxies are unbiased estimates of Y since they are based on a correctly specified regression model. As a result all methods are unbiased with close to nominal 5% non-coverage when we assume the correct value of $\lambda_A = 0$, with the spline having slightly wider interval widths (Figure 3.5). For other values of λ_A, the S-PPMA shows smaller bias, lower RMSE, and much higher coverage rate than their linear counterparts, and still achieves near nominal non-coverage under the incorrect assumption of $\lambda_A = 1$ (see Appendix).

For $\lambda_A = \lambda_T = 1$, where missingness depends equally on both Y and the auxiliary variables, estimates under $\lambda_A = 0$ (see Appendix) are similarly biased and intervals undercover the true value for all methods, which is not surprising since the assumption of λ_T is incorrect. However, S-PPMA has minor bias under $\lambda_A = 1$, which is the correct assumption in this case shown in Figure 3.5, and near nominal non-coverage rates under both assumptions of $\lambda_A = 1$ and $\lambda_A = \infty$, where the PPMA estimates are biased and undercover the true value. With respect to RMSE, S-PPMA shows increasing gains over PPMA as λ_A increases.

When $\lambda_T = \infty$, where missingness depends only on Y, the penalized spline is again approximately unbiased with nominal non-coverage under the correct assumption of $\lambda_A = \infty$, while the linear models are heavily biased. This is due to nonlinearity between X

and Y caused by the quadratic Z_2^2 term in the mean of Y, violating the bivariate normality assumption required in PPMA. Although the spline yields more conservative intervals, possibly from extrapolating nonlinearity, its ability to model nonlinearity results in estimates that are unbiased and have good coverage rates. This is especially important when the missing data mechanism is MNAR, where the proxy X is no longer unbiased and has a nonlinear relationship with Y. It is interesting to note, however, that in this and the previous scenario, PPMA shows slightly lower RMSE at the wrong assumption of $\lambda_A = 1$ when the true value is $\lambda_T = \infty$ (see Appendix).

In mechanism D, the methods show low bias and similar RMSE at all values of λ_A, with the ML estimate of BNPM having significantly wider intervals than S-BNPM and the Bayesian estimate of BNPM, resulting in better coverage. In mechanism E, all methods are generally biased and fail to achieve nominal non-coverage regardless of λ_A, with small differences in RMSE. Again the ML estimate of BNPM tends to yield much wider intervals that result in better coverage.

3.3.6. Scenario 6: interaction term in mean of Y

Let Z_1 and Z_2 be fully observed covariates with the following distributions

$$Z_1 \sim N(0, 1)$$

$$Z_2 \sim N(0, 1)$$

$$Y|Z_1, Z_2 \sim N(20 + Z_1 + Z_2 + 2Z_1Z_2, 1)$$

Let Y be missing under the following mechanisms

A. $Logit[Pr(R = 0)] = Z_1 + Z_2 + 2Z_1Z_2$ $\qquad (\lambda_T = 0)$

B. $Logit[Pr(R = 0)] = -20 + 0.98Z_1 + 0.98Z_2 + 1.96Z_1Z_2 + Y$ $(\lambda_T = 1)$

C. $Logit[Pr(R = 0)] = -20 + Y$ $(\lambda_T = \infty)$

D. $Logit[Pr(R = 0)] = 5Z_2$.

E. $Logit[Pr(R = 0)] = -10 + 0.5(5Z_2 + Y)$.

In this last scenario, we let the conditional mean of *Y* be a function of two normally distributed variables and their interaction. We then model *Y* using a correctly specified regression on Z_1, Z_2, and $Z_1 * Z_2$ for the respondents, and obtain the predicted values of *Y* as our proxy *X*. As in all scenarios, Figure 3.6 shows that when missingness is at random, all methods are unbiased, yield similar RMSE, and achieve nominal non-coverage under $\lambda_A = 0$ since the proxy *X* itself is unbiased for *Y*. However, under the incorrect values of $\lambda_A = 1$ and $\lambda_A = \infty$, the S-PPMA shows significantly larger bias, RMSE, and CIW than PPMA (see Appendix).

When $\lambda_T = 1$, all methods have negligible bias under the correct value of $\lambda_A = 1$ as shown in Figure 6, and achieve close to 5% non-coverage. There are generally minor differences in RMSE between the methods regardless of the assumption in λ_A, though S-PPMA tends to be slightly more conservative in terms of interval widths. For $\lambda_T = \infty$, all methods yield low bias with similar RMSE at $\lambda_A = \infty$ and nominal non-coverage. All methods have similar bias, RMSE, CIW, and coverage at all other values of λ_A. Although the mean of *Y* in this scenario depends on the interaction of Z_1 and Z_2, which is not normally distributed, the model assuming linearity between *X* and *Y* still yields good estimates of the mean under MNAR when $\lambda_A = \lambda_T$. This may be because the distribution of Z_1*Z_2 does not result in a drastic departure from normality in the proxy X, so the bivariate normality assumption between X and Y still approximately hold.

In the results for mechanism D (see Appendix), which does not correspond to any value of λ_T, estimates at $\lambda_A = 0$ are generally unbiased with minor differences in RMSE, and achieve close to nominal non-coverage with the exception of the Bayesian BNPM. In mechanism E, all methods show some bias at all values of λ_A with S-BNPM yielding higher RMSE than BNPM at $\lambda_A = \infty$.

3.4. Example: child asthma study

We apply S-PPMA and PPMA to an asthma study conducted by the University of Michigan Schools of Public Health and Medicine. The study consists of children with asthma from Detroit elementary and middle schools, whose aim is to evaluate the effectiveness of an educational intervention in reducing asthma symptoms. The main outcome of interest is the average number of nights the child experiences asthma symptoms per month, collected at baseline and one-year follow-up. Our goal is to estimate the mean change in nights of symptoms per month from baseline to follow-up, which is subject to dropout. However, since it is well documented that asthma severity naturally declines as the child ages, we restrict our attention to only those in the control group with symptoms at baseline.

Out of 133 children ages 6-14 with asthma symptoms at baseline in the control group, 41 (31%) dropped out before follow-up information was obtained. Since dropout may be attributed to asthma severity, we apply the S-PPMA and PPMA models to estimate the mean change in nights of symptoms per month. Only age and measurement at baseline are significantly associated with the outcome, with baseline age also being significantly associated with response. We first obtain our proxy by regressing change in nights per month on its baseline value and age using the respondent sample. We then apply the S-PPMA and PPMA models to estimate the mean change in nights of symptoms per month for the sample.

Figure 3.7 shows the distributions of baseline age and nights per month in our data. Both variables show deviations from normality, particularly nights of symptoms per month. Figure 3.8 displays scatter plots for the relationship between X and Y along with the average regression lines for PPMA and S-PPMA. For the regression of Y on X under the assumption of $\lambda = 0$, both PPMA and S-PPMA yield near identical regression lines. However, differences can be seen for the regression of X on Y under the assumption of $\lambda = \infty$, where S-PPMA seems to provide a minor improvement in fit. As such, we expect some differences between estimates from S-PPMA and PPMA, particularly at $\lambda = \infty$. Figure 3.9 shows estimates of the mean change under each method. Each line represents the mean and its 95% confidence interval for S-PPMA (PS) and PPMA, which is estimated using both maximum likelihood bootstrap (ML) and posterior draws (PD). To assess sensitivity to our assumption about λ, we display estimates under $\lambda = 0, 0.5, 1, 4$, and ∞. Results show that the mean change in symptoms per month generally decreases as we place more weight on our outcome to response, which suggests that children with higher decrease in symptoms may be less likely to participate in the follow-up survey. As expected from Figure 3.8, estimates for PPMA and S-PPMA at $\lambda = 0$ are similar, with differences between the methods being most pronounced at $\lambda = \infty$. There are minor differences between the PPMA estimates, with the posterior draws generally producing more conservative intervals than maximum likelihood. As in the simulations, interval lengths tend to widen slightly as λ increases due to increasing uncertainty when missingness depends on the outcome. S-PPMA is to a small degree less sensitive to assumptions about λ than PPMA, as estimates of mean change are within 0.1 nights of each other for values of $\lambda > 0$, whereas estimates from PPMA are generally within 0.4 nights as λ varies from 1/2 to ∞. In terms of precision, S-PPMA tends to be more conservative than ML but has slightly narrower interval widths than PD.

In practice, one might choose some intermediate value of λ (e.g. λ=1) since it represents a more conservative assumption about the missing data mechanism. However, lack of sensitivity to λ allows for more robustness of estimates to the assumptions about missingness, which is important since any belief regarding λ cannot be tested.

3.5. Discussion

Most nonresponse adjustment methods assume MAR, which can be a strong and untestable assumption. An advantage of the PPMA model is it allows us to make inferences about the mean of an outcome variable without assuming MAR. Moreover, the model does not require us to specify a propensity model, since it assumes that missingness depends only on the value of $X + \lambda Y$. The method simplifies nonresponse adjustment by combining a set of auxiliary variables into a single measure *X* and models departures from MAR using a single sensitivity parameter λ. In our proposed extension to the PPMA model, we model the relationship between *X* and *Y* through a spline. An advantage of this approach is that it does not require *X* and *Y* to be bivariate normal, which is assumed in PPMA, since splines allow us to model nonlinearity between the variables. As a result, we do not require the auxiliary variables to be normally distributed, as the model is robust to non-normal distributions of the auxiliary variables. It is important to note, however, that we do not specify a joint distribution between *X* and *Y*. Thus S-PPMA is more appropriately a method than a true model.

While S-PPMA utilizes initial values of *Y* generated from the potentially incorrect PPMA model, the additional steps of spline modelling and hot deck imputation helps to adjust for this nonlinearity. Our simulations show that the proposed S-PPMA model with penalized spline consistently yields approximately unbiased estimates with near nominal non-coverage regardless of the distributions of the auxiliary variables when the correct value of λ is used. Compared to the original PPMA proposed in Andridge and Little (2011), S-PPMA has shown to yield estimates that are more robust to covariate distributions, though with a slight penalty in precision when the PPMA model is correct. The gains in bias and RMSE are particularly noticeable the more the auxiliary variables deviate from normality. Results for a smaller sample size of n = 100 (see Appendix) show similar trend, where S-PPMA provide some gains in bias and RMSE when covariates are not normal and missingness is not at random, though differences in bias and RMSE tend to be less pronounced than in larger sample sizes. Moreover, the bootstrap variance estimates of PPMA tend to be more conservative than their Bayesian counterpart, leading to better coverages.

It may be tempting to estimate the value of λ by specifying a prior distribution. However, any inference about λ would be driven entirely by the prior since the data contains no information about λ. Thus we recommend conducting a sensitivity analysis by applying the S-PPMA model over a range of λ. The sensitivity analysis reflects our uncertainty about the nonresponse mechanism by displaying estimates of the mean over different values of λ, ranging from MAR ($\lambda = 0$) to the more extreme MNAR that assumes missingness depends only on the outcome itself ($\lambda = \infty$). Comparing estimates over a range of λ helps provide us an idea of how sensitive our inferences are to the missing data mechanism.

Our examples assume that the variables used to predict the outcome are fully observed, which may not be the case since often both outcome and covariates are missing at the same time, as is the case in unit nonresponse. Extension to the S-PPMA model incorporating additional assumptions about missingness of the covariates may be explored. In our simulations, S-PPMA tends to yield wider confidence intervals than the bivariate normal model particularly for $\lambda > 0$. This may be attributed to the fact that when the data is MNAR, values of the outcome for the nonrespondents may be drastically different than the respondents, leading to extrapolation. Estimation becomes particularly tricky when the relationship between Y and X is nonlinear. Thus, the lack of precision by the penalized spline at high values of λ may be a reflection of our uncertainty in extrapolating a nonlinear model.

The S-PPMA and PPMA models assume that missingness depends only on the value of $X + \lambda Y$, where X is a function of the covariates $Z_1, \ldots, Z_p$. In reality, there are infinite ways in which data is missing. For example, missingness of Y may depend only on some subset of $Z_1, \ldots, Z_p$, which would not be reflected by $X + \lambda Y$ for any λ. While we may place additional sensitivity parameters on the auxiliary variables, it will reduce simplicity of the model. Finally, we assume that our outcome variable, Y, is continuous and limit our inferences to the mean. Extensions to the PPM model are needed to model non-continuous outcome variables, and to estimate parameters of the regression of Y on the covariates under MNAR.

Figure 3.1. Results for scenario 1 where $\lambda_A = \lambda_T$.

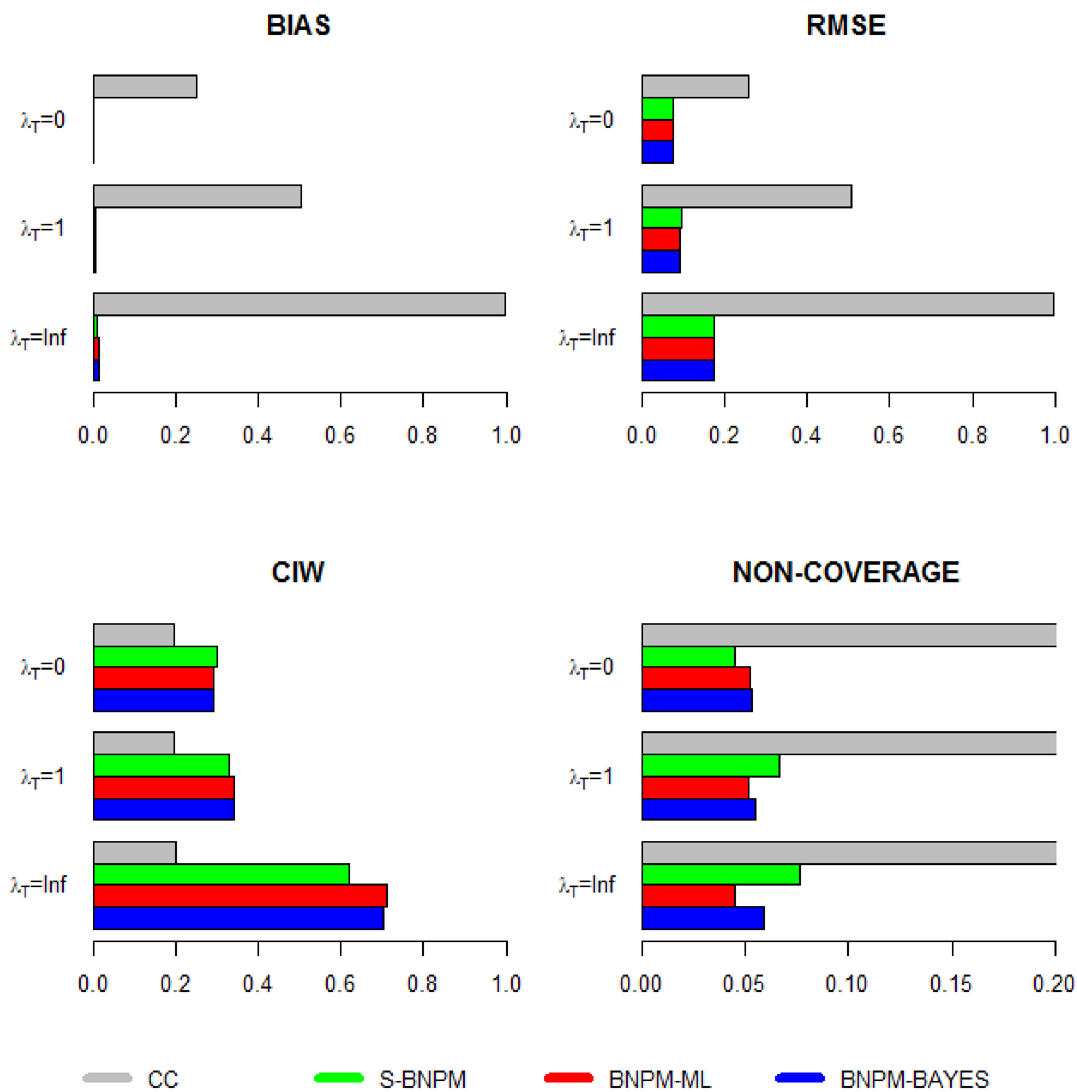


Figure 3.2. Results for scenario 2 where $\lambda_A = \lambda_T$.

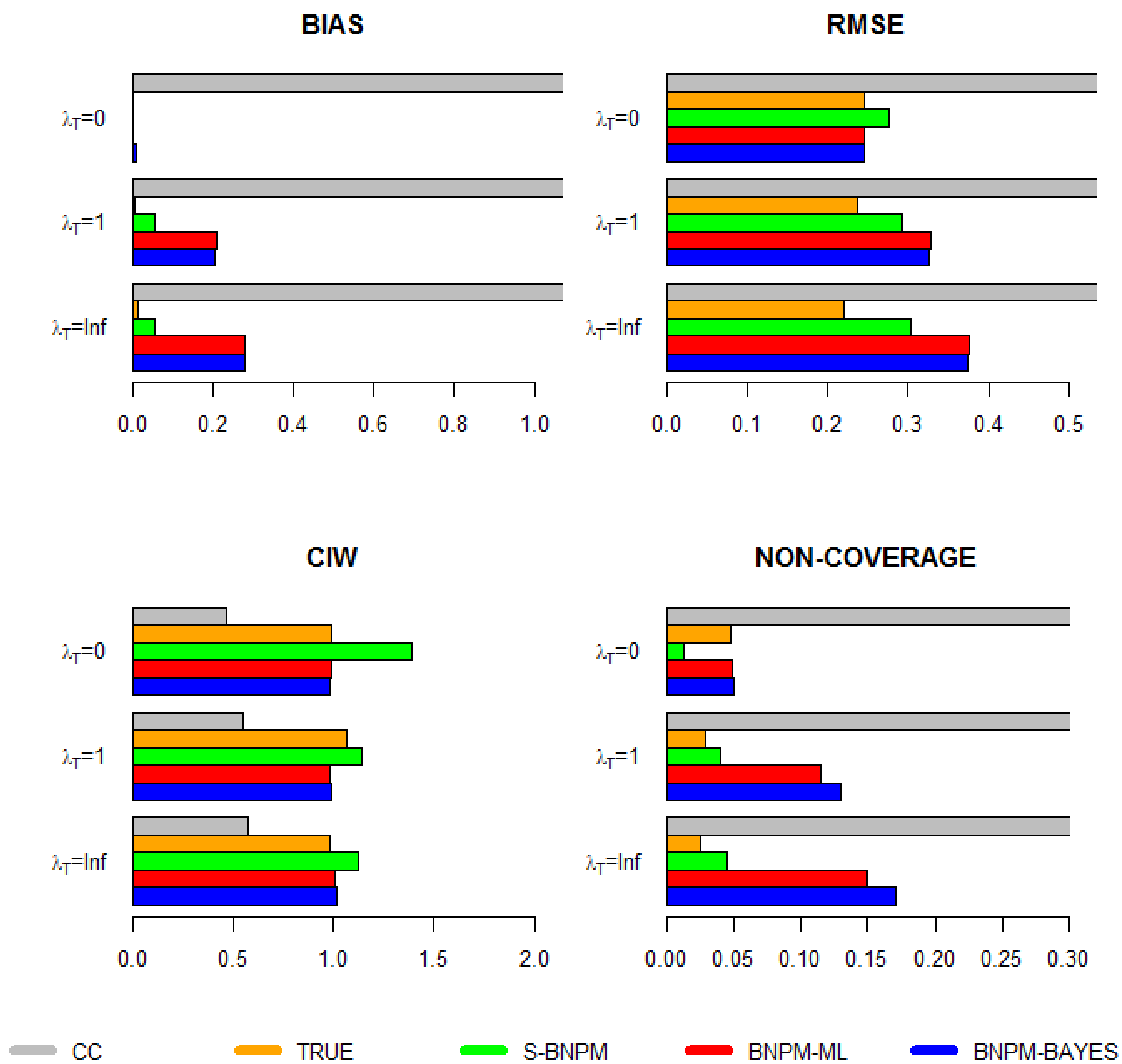


Figure 3.3. Results for scenario 3 where $\lambda_A = \lambda_T$.

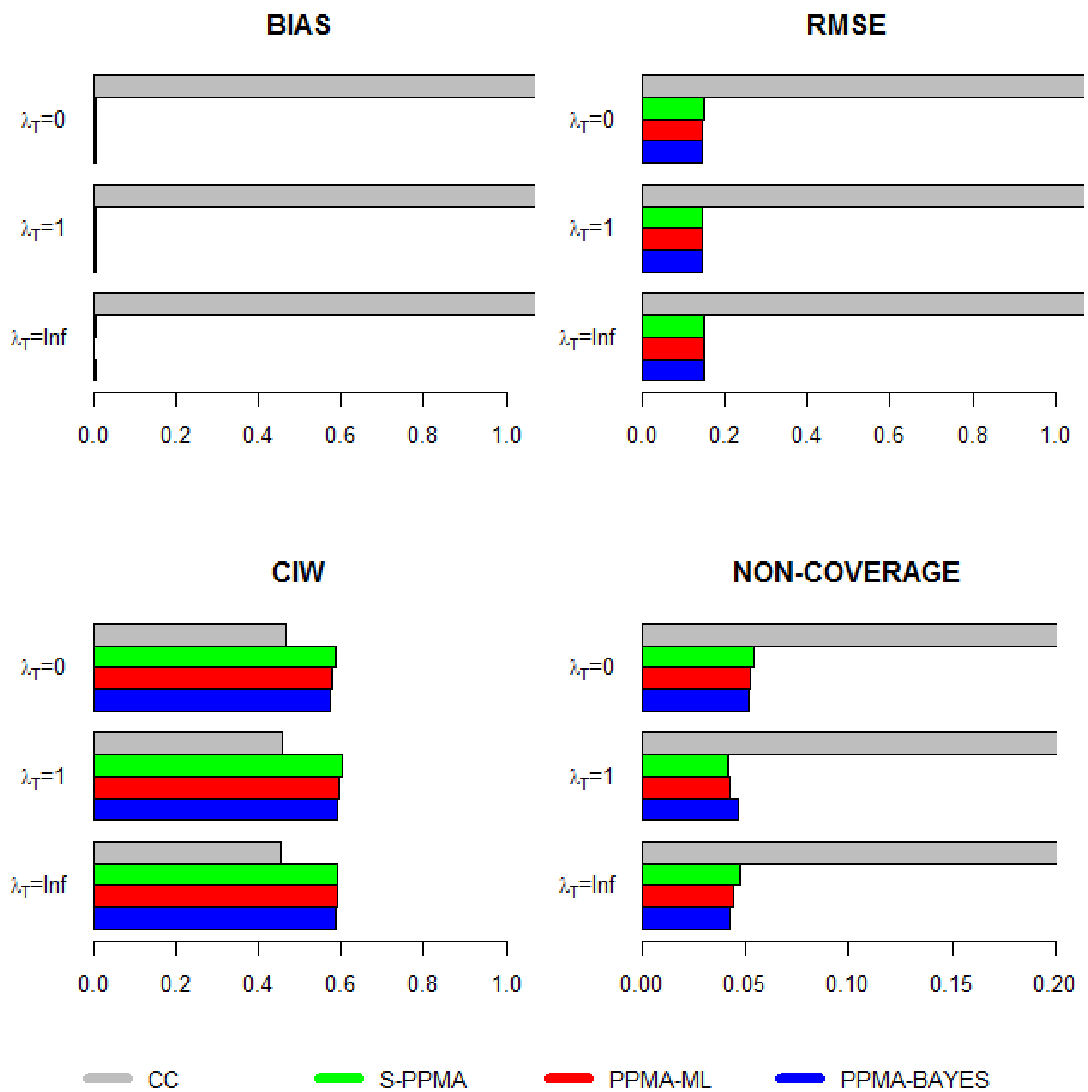


Figure 3.4. Results for scenario 4 where $\lambda_A = \lambda_T$.

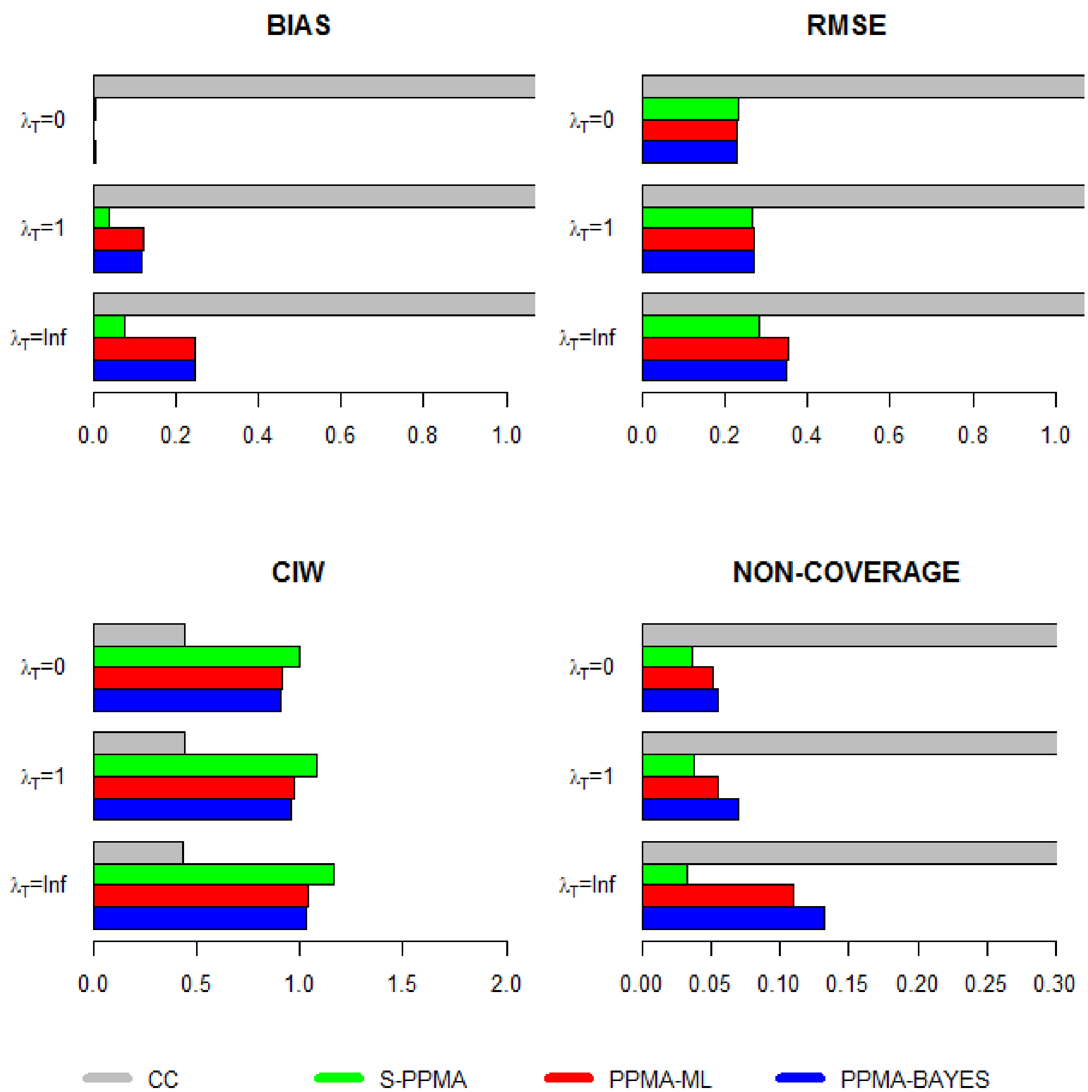


Figure 3.5. Results for scenario 5 where $\lambda_A = \lambda_T$.

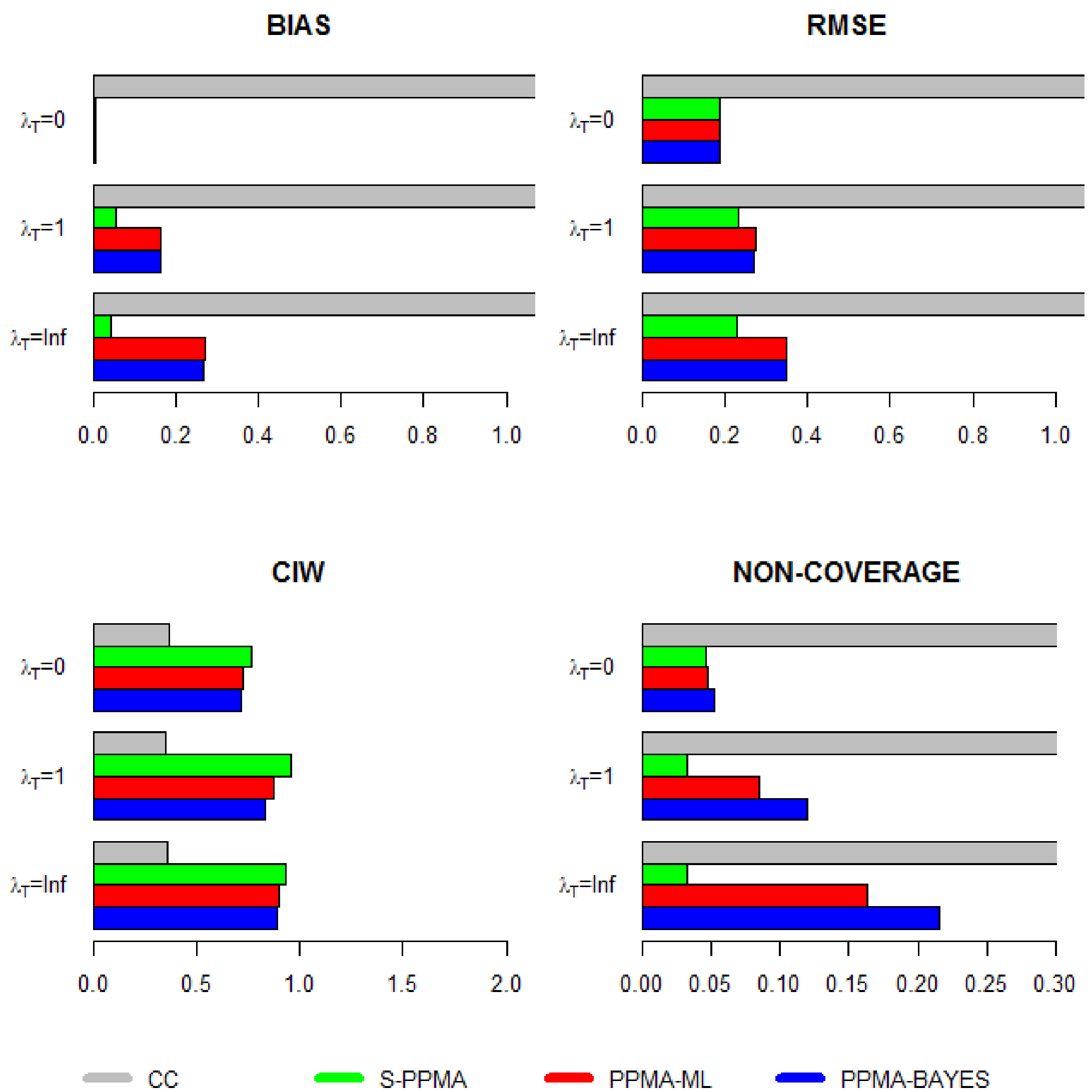


Figure 3.6. Results for scenario 6 where $\lambda_A = \lambda_T$.

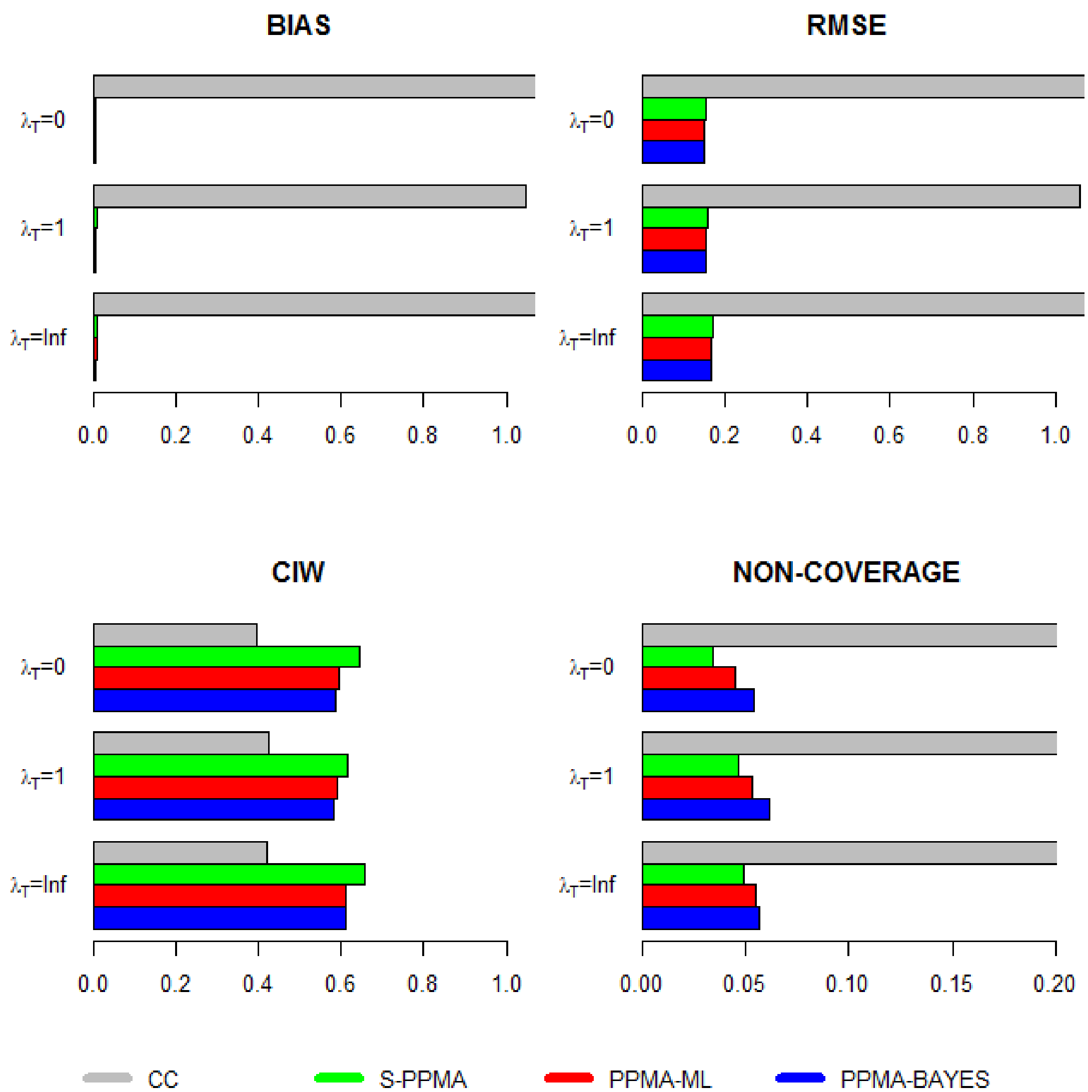


Figure 3.7. Distributions of baseline covariates.

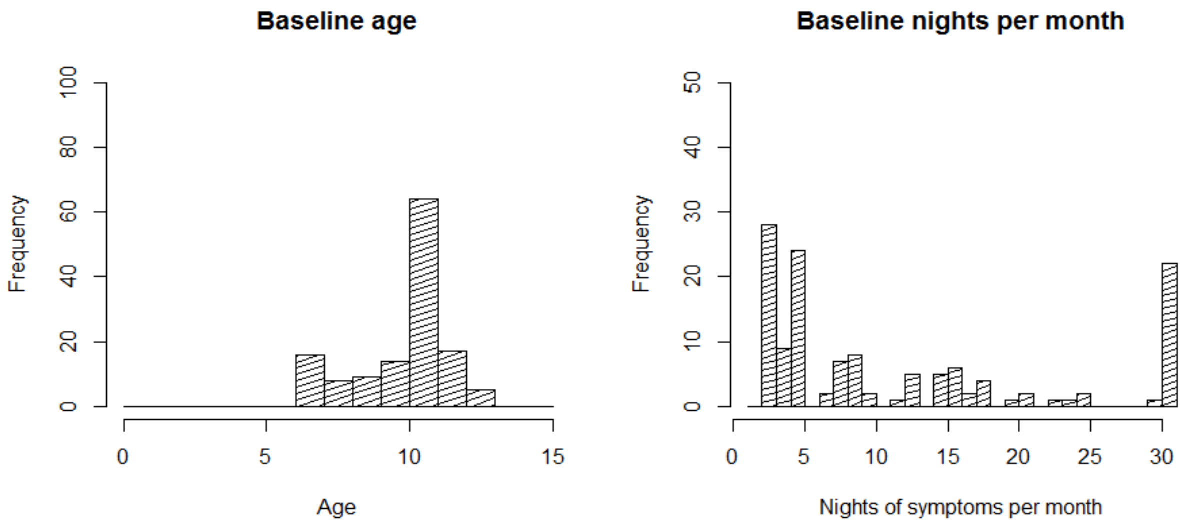


Figure 3.8. Relationship between *X* and *Y*.

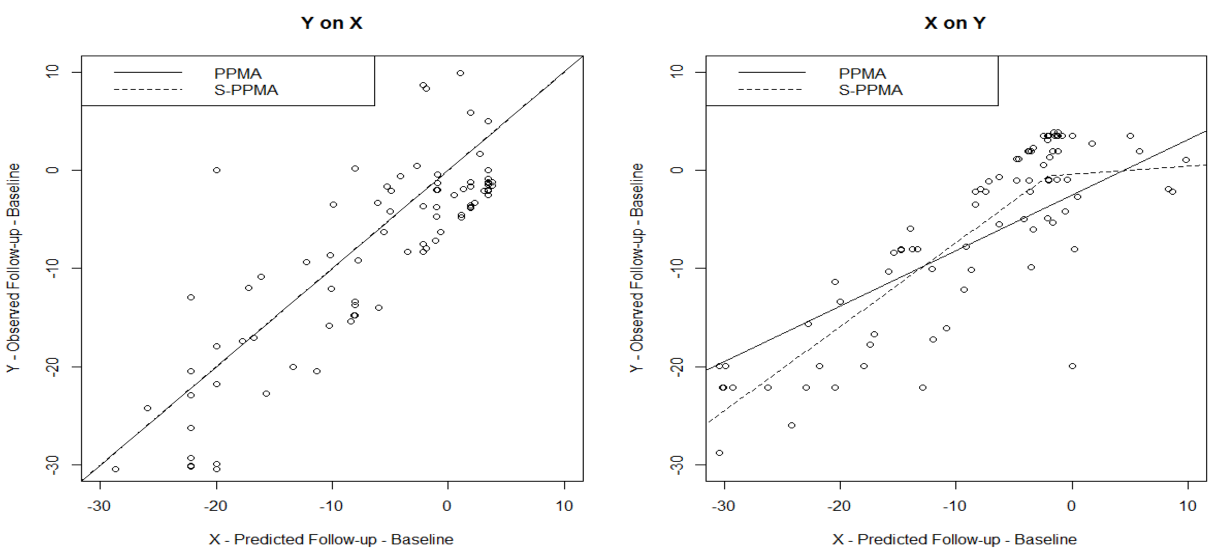
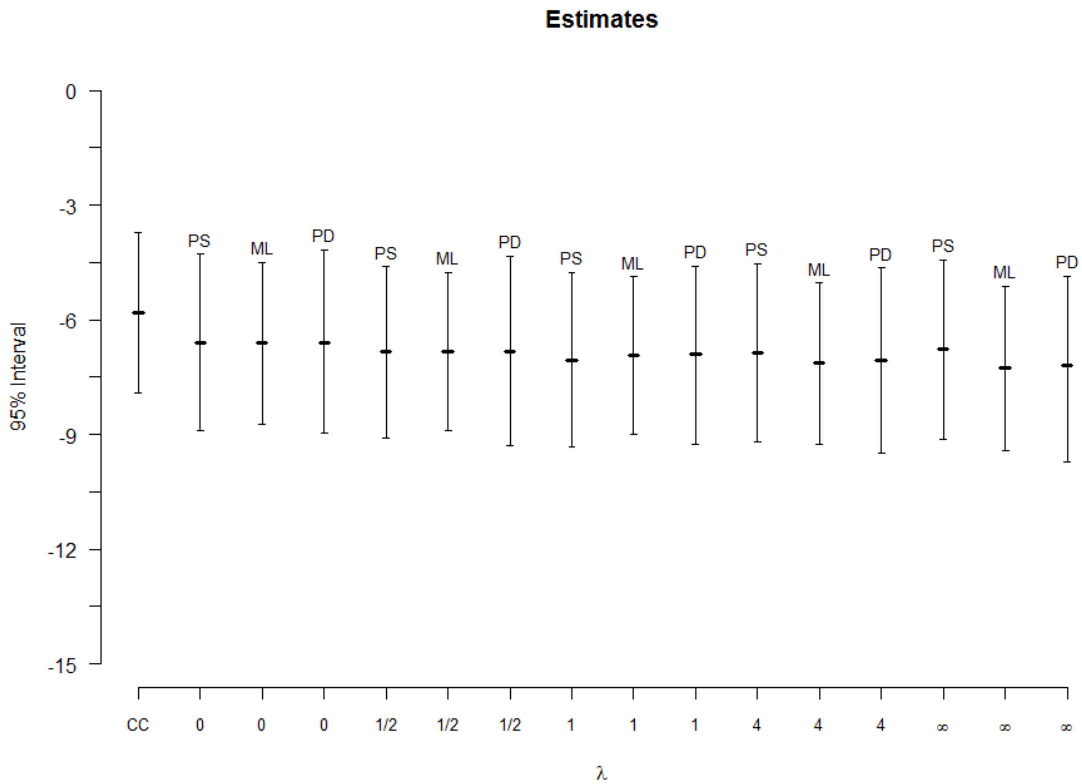


Figure 3.9. Estimates for mean change in nights of symptoms per month.



CHAPTER IV

Spline Pattern-Mixture Models for Missing Categorical Variables

4.1. Introduction

We consider the goal of estimating the mean of a categorical outcome *Y* with n_0 observed ($\{Y_i\}$, i = 1, …, n_0) and n_1 missing values ($\{Y_i\}$, $i = n_0+1,…, n_0+ n_1$). Suppose we fully observe a set of *p* auxiliary variables Z_1, …, Z_p ($\{Z_{i1}, …, Z_{ip}\}$, i = 1, …, n, $n = n_0 + n_1$), and let *R* be a response indicator such that *R* = 1 if *Y* is observed and *R* = 1 if *Y* is missing. If *Y* is missing at random (MAR, Rubin 1976) in that missingness does not depend on *Y* conditional on the observed variables Z_1, …, Z_p, methods such as regression imputation and weighting yield unbiased estimates of the mean. For example, with binary *Y*, one may specify a logistic regression model of *Y* on Z_1, …, Z_p using the complete cases, and impute the missing values of *Y* from this model. Alternatively, one may estimate a regression model for *R* on Z_1, …, Z_p and weight complete cases by the inverse of the estimated response propensity.

When *Y* is missing not at random (MNAR), in that missingness of *Y* depends on the value of *Y*, MAR-based methods are generally biased. Fay (1986) develops methods for estimating the mean of categorical variables subject to nonresponse in incomplete contingency tables. The method estimates expected frequencies via the EM algorithm and log-linear models for a set of causal models allowing for MNAR nonresponse. Nordheim (1984) proposed a method for estimating the mean of a binary outcome as a function of the ratio of nonresponse probabilities in each category. The ratio is assumed to be known and its value is varied in a sensitivity analysis. Baker and Laird (1988) further developed log-linear models for categorical responses subject to nonignorable nonresponse. All of these methods are for contingency table data with categorical response and predictors. Here we focus on estimating a binary outcome when information on continuous covariates is available. Extensions to *Y* with more than two categories are also outlined.

The starting point for our method is the bivariate normal pattern mixture model (BNPM) of Little (1994) for a continuous variable *Y* and a single observed covariate *X*. The BNPM model assumes a bivariate normal distribution for *X* and *Y*, with a different mean and covariance matrix for respondents and nonrespondents of *Y*, and identifies the parameters of the model by assumptions about the missing data mechanism. Andridge and Little (2011) extend this idea to multiple observed covariates $Z_1, \ldots, Z_p$ using a proxy pattern mixture model (PPMA), which reduces *Z* into a single proxy variable *X* obtained by regressing *Y* on *Z* for the respondents and setting *X* as predictions of *Y* for the sample. The method then applies the BNPM model to *X* and *Y*.

Both BNPM and PPMA assume a bivariate normal relationship between *X* and *Y*, and estimate the mean via a linear regression model with the independent variable determined by assumptions about the missing data mechanism. Estimates may be biased when the normality assumption, and hence linearity between *X* and *Y*, fails. Yang and Little (2014) propose a modification of the BNPM, S-BNPM, which replaces the parametric linear regression between *X* and *Y* by a penalized spline. S-BNPM builds on a robust MAR imputation method called penalized spline of propensity prediction (PSPP, Zhang and Little, 2009), which estimates the propensity that $R = 1$ given $Z_1, \ldots, Z_p$ based on a logistic regression of *R* on $Z_1, \ldots, Z_p$ using the sample, and imputes *Y* based on the regression of *Y* on a penalized spline of the estimated propensity. S-BNPM allows us to model nonlinear relationships between *X* and *Y* for a given assumption about the nonresponse mechanism. Simulations show that S-BNPM is more robust to normality assumptions than BNPM, at the expense of some precision. The idea is easily extended to data with more than one covariate (S-PPMA).

These methods are suitable for continuous outcomes. Andridge and Little (2009) extend PPMA to binary responses by a latent variable approach, where the value of a binary outcome *Y* is determined by a continuous, unobservable *U* such that $Y = 1$ when $U > 0$ and $Y = 0$ otherwise. This approach, which we label bin-PPMA, obtains a proxy *X* via a probit regression of *Y* on *Z* over the respondents, setting *X* as predicted values from the probit model for the whole sample. Respondent values of *U* are then drawn from a normal distribution with mean *X* and variance 1, from which *X* is recreated by regressing *U* on *Z*. Values of *X* and *U* are drawn iteratively and BNPM is applied to *X* and *U* at each iteration to estimate nonrespondent values of *U* given an assumption about the nonresponse mechanism, imputing *Y* such that $Y = 1$ if $U > 0$ and $Y = 0$ otherwise.

This latent variable model is sensitive to the normality assumptions for *X* and *U*. For example, if the covariates *Z* used to estimate *X* are non-normal, then the resulting proxy X is non-normal and the bivariate normality assumption between X and U fails, resulting in biased estimates. Thus we propose a S-PPMA method for binary Y (binS-PPMA), where we replace the linear regression model in bin-PPMA by a penalized spline, which we use to impute nonrespondent values of U given an assumption about the missing data mechanism. Imputations for U can then be translated directly to Y. More specifics on the method are given in the next section.

We study the performance of binS-PPMA by simulation, for data generated under various distributions of Z and missing data mechanisms. Specifically, we attempt to answer the following questions:

a. How does binS-PPMA compare with bin-PPMA with respect to bias, root mean squared error (RMSE), and coverage?
b. How robust is binS-PPMA to distributional assumptions and nonresponse mechanisms compared to bin-PPMA?

We now provide more details on BNPM, PPMA, S-PPMA, and their extensions to a binary responses.

4.2. Pattern mixture models for continuous outcomes

4.2.1 Review of bivariate normal pattern mixture model

Suppose X and Y are continuous variables, where X is a fully observed covariate and Y is the outcome for which we want to estimate the mean, but which may be missing not at random. Let R be an indicator of response. Little (1994) assumes a bivariate normal relationship between X and Y, specifically

$$(Y, X|\phi^{(r)}, R = r) \sim N_2\begin{pmatrix}\mu_Y^{(r)}, & \begin{bmatrix}\sigma_{YY}^{(r)} & \sigma_{XY}^{(r)}\\ \sigma_{XY}^{(r)} & \sigma_{XX}^{(r)}\end{bmatrix}\\ \mu_x^{(r)}, & \end{pmatrix} \tag{4.1}$$

$$R \sim Bernoulli(\pi)$$

where $N_2(\mu,\Sigma)$ denotes a bivariate normal distribution with mean μ and covariance matrix Σ. BNPM assumes that missingness depends on the value of $Y^* = X + \lambda Y$ for a given λ. Little (1994) shows that under this assumption the model is just-identified and the ML estimate for μ_Y is

$$\hat{\mu}_Y = \bar{Y}^{(1)} + \frac{\lambda s_{YY}+s_{XY}}{\lambda s_{XY}+s_{XX}}(\bar{X} - \bar{X}^{(1)}) \tag{4.2}$$

It is easy to see that a value of λ = 0 corresponds to an assumption of MAR, and $\hat{\mu}_Y$ is derived from a linear regression of *Y* on *X*. On the other end of the extreme, $\lambda = \infty$ implies missingness depends solely on *Y*, resulting in an estimate derived from the regression of *X* on *Y*. The advantage of the BNPM model is that we do not need to specify a model for the propensity, and departures from MAR can be represented by a single sensitivity parameter λ. Since we do not know the true value of λ, it is advisable to conduct a sensitivity analysis by displaying estimates over a range of potential values of λ to capture its uncertainties.

4.2.2 Extension of PPMA to more than one covariate for a continuous outcome

In many studies we have information on multiple observed covariates Z_1, …, Z_p that can be predictive of *Y*. Andridge and Little (2011) extends the idea of BNPM by reducing Z_1, …, Z_p into a single best predictor, or proxy, *X* of *Y* (PPMA). The method estimates X by regressing Y on Z_1, …, Z_p using the respondents and setting X to be predictions of Y for the sample. With a fully observed set of X, we can then apply BNPM to estimate μ_Y for a given assumption of λ. Both BNPM and PPMA are dependent on the bivariate normality assumption of X and Y, which can be violated when some or all of the Z_1, …, Z_p used to estimate X are non-normal. In the next section, we extend PPMA to categorical outcomes, and propose a penalized spline method that relaxes the normality assumptions.

4.3. Extensions to categorical outcomes

Thus far we have discussed methods for assessing nonresponse bias for continuous outcomes. We can extend the ideas of PPMA to categorical outcomes using a latent variable approach (Andridge and Little 2009). Here we consider the case of a binary Y, though methods presented can be generalized to an ordinal Y with multiple categories. Suppose Y is a binary response with n_0 observed ($\{Y_i\}$, $I = 1, …, n_0$) and n_1 missing values ($\{Y_i\}$, $I = n_0+1,…, n_0+ n_1$), and Z_1, …, Z_p ($\{Z_{i1}, …, Z_{ip}\}$, $I = 1, …, n$, $n = n_0 + n_1$) are p fully observed covariates. In addition, suppose there exists a continuous, latent U such that

a. $U|Z \sim N(\alpha Z, 1)$ (4.3)

b. $Y = \begin{cases} 0 \text{ if } U < 0 \\ 1 \text{ if } U \geq 0 \end{cases}$

The latent variable approach allows for a straightforward application of PPMA to assess for nonresponse bias. Borrowing from the idea in PPMA, we can obtain a proxy X that is the single best predictor of U, estimated by a linear regression of U on Z for the respondents and setting X to be the predicted values from the regression model for the whole sample.

Since we do not observe the values of U, we must estimate both U and X simultaneously through data augmentation. In this chapter we use Gibbs sampling to iteratively draw U and X from their respective distributions (Albert and Chib, 1993). We summarize the procedure as follows:

1. Obtain an initial estimate of X via a probit regression of Y on Z for the respondents, letting X be the predicted values from the model for the sample. Specifically, we fit the model via maximum likelihood:

   $$\widehat{Pr}(Y = 1 | Z, R = 1) = \phi(\hat{\alpha}Z)$$

   and initialize $X = \hat{\alpha}Z$ for both respondents and nonrespondents given their information on Z.

2. At the d^{th} iteration, draw respondent values for U under a truncated distribution

   $$(U^{(d)} | Y, X^{(d\text{-}1)}, R = 1) \sim N(X^{(d\text{-}1)}, 1)$$

   where drawn values of $U < 0$ for which $Y = 1$ or $U > 0$ for which $Y = 0$ are discarded and redrawn.

3. Draw $(\hat{\alpha}^{(d)} | Y, U^{(d)}, R = 1) \sim N((Z^{T}Z)^{-1}Z^{T}U^{(d)}, (Z^{T}Z)^{-1})$ and set $X^{(d)} = \hat{\alpha}^{(d)}Z$ for the sample.

4. Repeat 2 – 3 over 1000 iterations to create 1000 sets of fully observed X and partially observed (for respondents) U.

Since X are predicted values of U, X is unbiased for U for the respondents, hence $(U|Y, X, R = 1) \sim N(X, 1)$. We then recreate X at the end of each iteration to account for uncertainties associated in estimating X. For each set of (X, U), Andridge and Little (2009) applies PPMA to obtain imputations of the missing values of U, and derive imputations of Y based on imputed values of U.

4.3.1 Spline bivariate proxy pattern mixture model for binary Y

One of the limitations of the BNPM and PPMA models is sensitivity to the assumption of bivariate normality, which is the foundation of the methods. For example, when Y is continuous, (X, Y) is not bivariate normal when the marginal distribution of X is gamma, or when X is normal but the mean of Y given X is quadratic on X. In such cases X and Y may not follow a linear relationship, leading to biased estimates from BNPM even with the correct assumption of λ. Similarly when Y is categorical, the relationship between X and U may not be linear when the variables are non-normal. Since Y is imputed based on the value of U, bias in the imputations of U leads to bias in the imputations of Y.

To account for potential nonlinearity, we propose a penalized spline regression to model U and Y (binS-PPMA). The binS-PPMA utilizes the same principle that when missingness depends on the value of $U^* = X + \lambda U$, for some known λ, the regression of X on U^* is the same over patterns of response. Specifically, binS-PPMA assumes the following model for this regression:

$$X = \beta_0 + \beta_1 U^* + \sum_{k=1}^{K} \gamma_k (U^* - \kappa_k)_+ + \varepsilon \quad (4.4)$$

$$\varepsilon \sim N(0, \sigma^2)$$

$$\gamma_k \sim N(0, \tau^2)$$

where a_+ = a if $a > 0$ and $a_+ = 0$ otherwise, and $\kappa_1 < \ldots < \kappa_K$ are K equally spaced knots. We estimate the parameters using a Bayesian approach, assigning a uniform prior for β and inverse gamma (10^{-5}, 10^{-5}) priors for σ^2 and τ^2, and draw from their posterior distribution via a Gibbs sampler.

Imputations of missing values of U^*, however, are not as straightforward, since U^* appears as a covariate in the model. Thus, as in Yang and Little (2015), we apply a hotdeck procedure where we generate a donor pool and impute U^* with the observed value of a matched donor with X and U^* observed. The procedure for values of $\lambda > 0$ is summarized as follows:

1. For the d^{th} set of draws ($X^{(d)}$, $U^{*(d)}$), draw B values of $U^{*(d)}$ for each nonrespondent from the distribution of $U^{*(d)}|X^{(d)}$, $R = 0$, estimated under the BNPM model. This results in a pool of n_1*B values of $U^{*(d)}$ ($\{U_p^{*(\mathrm{d})}\}$, $p = 1, \ldots, n_1$*B). In the simulations in Section 4.4 a value of $B = 3$ is sufficient.
2. Given each $U_p^{*(d)}$ in the pool, draw a value $X_p^{(d)}$ from the posterior predictive distribution of $X^{(d)}|U^{*(d)}$ in (4.4), with parameters estimated from respondents. This results in a set of pairs of ($\{X_p^{(d)}, U_p^{*(d)}\}$, $p = 1, \ldots, n_1$*B) that form our donor pool.

3. For each nonrespondent j, choose a pair $(X_k^{(d)}, U_k^{*(d)})$ from the donor pool $(\{X_p^{(d)}, U_p^{*(d)}\}$, p = 1, …, n_1*B) with the closest value $X_k^{(d)}$ to $X_j^{(d)}$, and impute $U_j^{*(d)} = U_k^{*(d)}$ (hence $U_j^{(d)} = (U_k^{*(d)} - X_j^{(d)}) / \lambda$ and $Y_j = I\{U_j^{(d)} > 0\}$) from that pair.
4. Repeat steps 1-3 over d = 10, 20, 30, …, 1000 to create D = 100 multiply-imputed data sets with values of Y imputed.

We then obtain $\hat{\mu}_Y$ and its variance via multiple imputation combining rules (Rubin, 1987)

$$\hat{\mu}_Y = \hat{\mu}_D = \frac{1}{D}\sum_{d=1}^{D} \hat{\mu}_d \tag{4.5}$$

$$Var(\hat{\mu}_Y) = \frac{1}{D}\sum_{d=1}^{D} W_d + \frac{D+1}{D(D-1)}\sum_{d=1}^{D}(\hat{\mu}_d - \bar{\mu}_D)^2 \tag{4.6}$$

where $\hat{\mu}_d$ and W_d are the estimated marginal mean and variance in the d^{th} imputed data set, respectively. For the assumption of λ = 0, we reverse the regression in (4.4) to model $U^{(d)}$ on X, drawing imputations of $U^{(d)}$ (and hence $Y = I\{U^{(d)} > 0\}$) directly from its posterior predictive distribution given observed values of X.

When the regression of X on U^* is linear, the initial draws from BNPM are unbiased estimates of U^*, and the hotdeck procedure has little impact on the imputation of U^*. However, when the relationship is nonlinear, thereby violating the linearity assumption in BNPM, the spline mimics the true regression line of X given U^*, resulting in improvements in the imputations of U^*. Hence the hotdeck procedure serves as an adjustment for nonlinearity between X and U^*. A value of B is chosen to ensure that there exists a close match for X in the donor pool for every nonrespondent. Since (X, U) are estimated iteratively, the procedure requires us to draw a single set of parameters from the spline regression model ($\hat{\beta}$, $\hat{\gamma}$, $\hat{\sigma}^2$, $\hat{\tau}^2$) of X on U^* separately for each set of ($X^{(d)}$, $U^{(d)}$). To reduce autocorrelation we use every other 10 sets of drawn ($X^{(d)}$, $U^{(d)}$).

The binS-PPMA method allows us to model nonlinearity in the regressions between X and U^*, improving robustness to bivariate normality. Unbiased estimates of U are particularly important near the threshold of $U = 0$, where the value of Y is determined. Misspecification of the pattern mixture model near the threshold will result in biased estimates of $\hat{\mu}_Y$. As in all pattern mixture model analyses for continuous and binary outcomes, we should display estimates over a range of λ to capture our uncertainties about its true value. In the following section, we conduct simulation studies to compare performances of binS-PPMA and bin-PPMA in their extensions to binary outcomes.

4.4. Simulation studies

We now conduct simulations to assess the performance of binS-PPMA for binary outcomes over a range of distributional assumptions and missing data mechanisms. For comparison we include two estimates from bin-PPMA: a Bayesian approach (BA) and multiple imputation (MI).

In bin-PPMA (BA), we apply BNPM to ($X^{(d)}$, $U^{(d)}$) for d = 10, 20, 30, …, 1000 to obtain posterior draws for the parameters in (4.1), assuming ($X^{(d)}$, $U^{(d)}$) are bivariate normal. We then obtain the posterior distribution of μ_Y by computing

$$\hat{\mu}_Y = \widehat{Pr}(Y = 1) = \widehat{Pr}\,(U^{(d)} > 0) = \hat{\pi}\phi(\hat{\mu}^{(1)}_{U^{(d)}} \,/\, \sqrt{\hat{\sigma}^{(1)}_{U^{(d)}U^{(d)}}}) + (1-\hat{\pi})\phi(\hat{\mu}^{(0)}_{U^{(d)}} \,/\, \sqrt{\hat{\sigma}^{(0)}_{U^{(d)}U^{(d)}}})$$

where $\hat{\mu}^{(1)}_{U^{(d)}}$ and $\hat{\sigma}^{(1)}_{U^{(d)}U^{(d)}}$ are posterior draws for the mean and variance of $U^{(d)}$ for respondents, and $\hat{\mu}^{(0)}_{U^{(d)}}$ and $\hat{\sigma}^{(0)}_{U^{(d)}U^{(d)}}$ are posterior draws for the mean and variance of $U^{(d)}$ for nonrespondents, respectively. Finally, we obtain the median and its 95% credibility interval for μ_Y.

In bin-PPMA (MI), we impute the nonrespondent values of $U^{(d)}$ from

$$U^{(d)} | X^{(d)}, R = 0 \sim N\left(\hat{\mu}^{(0)}_{U^{(d)}} + \frac{\hat{\sigma}^{(0)}_{U^{(d)}X^{(d)}}}{\hat{\sigma}^{(0)}_{X^{(d)}X^{(d)}}} (X^{(d)} - \hat{\mu}^{(0)}_{X^{(d)}}), \hat{\sigma}^{(0)}_{U^{(d)}U^{(d)}} - \frac{\hat{\sigma}^{(0)\,2}_{U^{(d)}X^{(d)}}}{\hat{\sigma}^{(0)}_{X^{(d)}X^{(d)}}}\right)$$

with parameters estimated by BNPM assuming bivariate normality of ($X^{(d)}$, $U^{(d)}$). We repeat for d = 10, 20, 30, …, 1000 to obtain 100 imputed data sets. This approach is less sensitive to violations of normality than bin-PPMA (BA) since the normality assumption is confined to the imputations of the missing values.

We compare performance of binS-PPMA, bin-PPMA (BA), and bin-PPMA (MI) with respect to average bias, root mean square error, 95% confidence interval width, and rate of confidence interval non-coverage over 1000 replications. For each replication we construct 95% confidence intervals as

$$95\%\ CI = (\hat{\mu}_Y - t_{n\text{-}1,0.975}\sqrt{Var(\hat{\mu}_Y)},\ \hat{\mu}_Y + t_{n\text{-}1,0.975}\sqrt{Var(\hat{\mu}_Y)})$$

where $t_{n\text{-}1,0.975}$ is the 97.5th percentile of the t-distribution with n-1 degrees of freedom, and $Var(\hat{\mu}_Y)$ is the estimated variance of the mean. The corresponding confidence interval width is then

$$CIW = 2 * t_{n\text{-}1,0.975}\sqrt{Var(\hat{\mu}_Y)}$$

Finally, we estimate the non-coverage rate as the proportion of the 1000 confidence intervals that do not cover the true value. In the first scenario, we simulate the situation where *X* and *U* are bivariate normal and compare estimates from bin-PPMA and binS-PPMA. In the second and third scenarios, we simulate data with non-normal distributions for *X* and *U*. Finally, in the last scenario we explore data from a 2x2 contingency table where both the predictor and outcome are binary. For each scenario, we estimate the proxy based on a correctly specified regression of *U* on *Z* using the respondent sample.

Let λ_T be the true, unobservable value of λ for the missing data mechanism, and let λ_A be our assumption value of λ in our estimates. For each scenario, we vary the true value λ_T from 0, 1, and ∞ to simulate situations where missingness depends on *X*, *Y*, or a combination of the two. At each λ_T we conduct a sensitivity analysis under λ_A = 0, 1, and ∞ to account for our uncertainty about λ, with one of our assumptions λ_A being the true value. We choose values of λ_A = 0, 1, and ∞ as they capture a wide range of potential nonresponse mechanisms, though other values may be explored in practice. The following results display estimates for which $\lambda_A = \lambda_T$, since our focus is to compare performances of the methods when the assumption about the nonresponse mechanism is correct. Results at other values of λ_A are discussed, with results shown in supplementary materials.

4.4.1 Scenario 1: bivariate normal sample

Suppose *Z* is a fully observed, normally distributed covariate and *Y* is a binary outcome whose value is determined by a latent *U* that is normal conditional on *Z*:

$$Z \sim N(0, 1)$$

$$U|Z \sim N(0.5 + Z, 1)$$

$$Y = \begin{cases} 0 \text{ if } U < 0 \\ 1 \text{ if } U \geq 0 \end{cases}$$

We generate missing value of *Y* under the following logistic models, which simulate MAR and MNAR

A. $Logit[Pr(R = 0)] = Z$ ($\lambda_T = 0$)

B. $Logit[Pr(R = 0)] = -0.1 + 0.25(0.98Z + U)$ ($\lambda_T = 1$)

C. $Logit[Pr(R = 0)] = -0.25 + 0.5U$ ($\lambda_T = \infty$)

D. $Logit[Pr(R = 0)] = -0.8 + Z^2$ ($\lambda_T = 0$)

E. $Logit[Pr(R = 0)] = -0.8 + 0.25(0.74Z + U)^2$ ($\lambda_T = 1$)

F. $Logit[Pr(R = 0)] = -1 + 0.5U^2$ ($\lambda_T = \infty$)

For all scenarios we generate data for $n = 100$ and $n = 400$ and a response rate of approximately 50%. Due to similarity of results only those under $n = 400$ are displayed (see supplemental materials for results with $n = 100$). For comparison we include estimates from complete cases analysis (CC) and a Bayesian logistic regression of *Y* on *Z* assuming MAR (LOGREG). Figures 4.1a-b displays average bias, RMSE, CIW, and non-coverage rates out of 1000 replications when $\lambda_A = \lambda_T$ for $\lambda_T = 0$, 1, and ∞. As expected, as assumption of bivariate normality holds in this scenario, both bin-PPMA and binS-PPMA are approximately unbiased and achieve a near 5% nominal non-coverage rate regardless of the nonresponse mechanism when the correct assumption about λ is made. Results show minor differences in RMSE and CIW regardless of λ_A and λ_T (results for $\lambda_A \neq \lambda_T$ in Appendix). CC is highly biased under all nonresponse mechanisms, while

LOGREG yields valid estimates when $\lambda_T = 0$, but biased under MNAR. The effect of MNAR on LOGREG is less consequential when missingness depends on U^{*2}, perhaps due to the symmetrical nature of nonresponse as subjects with highly positive or highly negative values of U^* are equally likely to be missing.

4.4.2 Scenario 2: gamma distributed Z

Suppose now Z is a fully observed, gamma-distributed covariate and U is normal conditional on Z, we generate the data as follows:

$$Z \sim Gamma(1, 1)$$

$$U|Z \sim N(-2.25 + Z, 1)$$

$$Y = \begin{cases} 0 \text{ if } U < 0 \\ 1 \text{ if } U \geq 0 \end{cases}$$

We delete Y under the following models:

A. $Logit[Pr(R = 0)] = -0.5 + 0.5Z$ $(\lambda_T = 0)$

B. $Logit[Pr(R = 0)] = 0.1 + 0.25(0.98Z + U)$ $(\lambda_T = 1)$

C. $Logit[Pr(R = 0)] = 0.6 + 0.5U$ $(\lambda_T = \infty)$

D. $Logit[Pr(R = 0)] = -0.6 + 0.5Z^2$ $(\lambda_T = 0)$

E. $Logit[Pr(R = 0)] = -1.5 + 0.5(0.6Z + U)^2$ $(\lambda_T = 1)$

F. $Logit[Pr(R = 0)] = -2.5 + U^2$ $(\lambda_T = \infty)$

Results for which $\lambda_A = \lambda_T$ under nonresponse mechanisms A-F are summarized in Figures 4.2a-b. At $\lambda_A = \lambda_T = 0$, all methods are unbiased, yield similar RMSE and CIW, and achieve nominal 5% non-coverage rate with the exception of CC and bin-PPMA (BA).

Since the marginal distribution of *X*, and hence *U*, is non-normal, estimates from bin-PPMA (BA) based on the normal density function for *U* are biased. Consequently it results in higher RMSE, CIW, and non-coverage rates compared to other methods at all λ_T. When $\lambda_T = 0$, the data is MAR and the regression of *U* on *X* over the complete cases is unbiased for the nonrespondents. Thus PPMA-MI provides valid imputations for *Y* when the correct assumption of $\lambda_A = 0$ is made. When $\lambda_T > 0$, the regression of *X* on *U** is no longer linear due to lack of normality, resulting in biased estimates of *U** and *Y* under bin-PPMA even when the correct assumption of λ is made. LOGREG, which assumes MAR, also becomes biased and lack nominal non-coverage when $\lambda_T > 0$. As expected bin-PPMA (MI) is more robust to deviations from normality than bin-PPMA (BA), as it yields better RMSE and non-coverage than its Bayesian counterpart. BinS-PPMA, which is able to model nonlinearity, shows noticeable improvements in bias and RMSE compared to both bin-PPMA methods at all values of λ_A when $\lambda_T > 0$, and achieves approximately nominal non-coverage when $\lambda_A = \lambda_T$.

4.4.3 Scenario 3: nonlinear on Z

Suppose *Z* is normal but *U* is quadratic on *Z*, we generate the data as follows:

$$Z \sim N(0, 1)$$

$$U|Z \sim N(-1.75 + Z^2, 1)$$

$$Y = \begin{cases} 0 \text{ if } U < 0 \\ 1 \text{ if } U \geq 0 \end{cases}$$

Let *Y* be missing under the following models:

A. $V|Z, U \sim N(-0.5 + 0.5Z^2, 1)$ $(\lambda_T = 0)$

B. $V|Z, U \sim N(0.25(0.94Z^2 + U), 1)$ $(\lambda_T = 1)$

C. $V|Z, U \sim N(0.5 + 0.5U, 1)$ $(\lambda_T = \infty)$

D. $V|Z, U \sim N(-0.5 + 0.5Z^4, 1)$ $(\lambda_T = 0)$

E. $V|Z, U \sim N(-0.7 + 0.2(0.68Z + U)^2, 1)$ $(\lambda_T = 1)$

F. $V|Z, U \sim N(-0.7 + 0.25U^2, 1)$ $(\lambda_T = \infty)$

where *Y* is observed when $V < 0$ and missing otherwise.

In addition to the pattern mixture models, in the last two scenarios we include estimates from a latent variable analysis that estimates *V* given information about ($X^{(d)}$, $U^{(d)}$) for each set *d*. Steps are summarized as follows for $\lambda > 0$, where $\widehat{U}^{*(d,R=0,\,i=1)}$ are the imputed values of $U^{*(d)}$ at the i^{th} iteration of the d^{th} set and $\widehat{U}^{*(d,\,i=1)}$ are the observed and imputed values for the whole sample:

1. At the d^{th} set of ($X^{(d)}$, $U^{(d)}$), initialize $\widehat{U}^{*(d,R=0,\,i=1)}$ by setting $\widehat{U}^{*(d,R=0,\,i=1)}$ for nonrespondents to be predictions from the regression of $U^{*(d,\ R=1)}|X^{(d,\ R=1)}$ using the complete cases, then draw

$$\widehat{V}^{(i=1)}\,|\,\widehat{U}^{*(d,\,i=1)} \sim N(Z_{\widehat{\pi}} - \overline{U}^{*(d,\,i=1)} + \widehat{U}^{*(d,\,i=1)}, 1)$$

for the sample, where $\widehat{\pi}$ is the nonresponse rate, Z_α is the α^{th} percentile of the standard normal distribution, and $\overline{U}^{*(d,\,i=1)}$ is the mean of $\widehat{U}^{*(d,\,i=1)}$ for the sample. For respondents, positive values of $\widehat{V}^{(i=1)}$ are discarded and redrawn until all values are negative. Likewise for nonrespondents, we discard and redraw negative values of $\widehat{V}^{(i=1)}$.

2. At the *i*th iteration, obtain posterior predictive draws of $\widehat{U}^{*(d,\,R=0,i)}|\widehat{V}^{(R=0,i-1)}$, $X^{(d,R=0)}$ for nonrespondents under a linear regression model with parameters estimated from $\widehat{U}^{*(d,\,i-1)}|\widehat{V}^{(i-1)}$, $X^{(d)}$ using the entire imputed sample, with values of $\widehat{U}^{*(d,\ \ R=0,\,i-1)}$ and $\widehat{V}^{(i-1)}$ drawn from the previous iteration.
3. Obtain posterior predictive draws of $\widehat{V}^{(i)}|\widehat{U}^{*(d,\,i)}$, $X^{(d)}$ for the sample based on a linear regression model of $\widehat{V}^{(i-1)}|\widehat{U}^{*(d,\,i)}$, $X^{(d)}$ using the entire imputed sample. We again discard and redraw all positive values of $\widehat{V}^{(i)}$ for respondents and negative values of $\widehat{V}^{(i)}$ for nonrespondents.
4. Repeat 2 – 3 over *i* = 1, 2, …, 200. Then, draw a value *k* from *i* = 101, …, 200 and obtain a posterior draw for μ_Y as

$$\hat{\mu}_Y^{(d)} = \sum_{j=1}^{n} I\{\,\widehat{U}_j^{*(d,i=k)} > 0\}\,/\,n$$

redrawing respondent values of $\widehat{U}^{*(d,\,i=k)}$ based on drawn regression parameters at the *k*th iteration. We repeat steps 1 – 4 for *d* = 10, 20, 30, …, 1000 to obtain posterior draws and its associated median and 95% credibility intervals for μ_Y. For λ = 0, we produce posterior draws for μ_Y based on a linear regression model of $U^{(d)}$ on $X^{(d)}$. Since *V* and *U** are bivariate normal conditional on *X*, and the distribution of *U** given *V* and *X* is the same for respondents and nonrespondents (since response is determined by *V*), the procedure produces unbiased estimates for *U**.

Here we estimate the proxy X based on the regression of U on Z^2, with U and X estimated iteratively. For LOGREG, we impute from the regression of Y on Z^2 for the respondents. Results from the various methods under the correct λ_A are shown in Figure 4.3a-b. As in the previous scenarios, for $\lambda_A = \lambda_T = 0$ all methods except CC bin-PPMA (BA) are unbiased and achieve nominal non-coverage, while yielding similar RMSE and CIW. Since the marginal distribution of U is non-normal, bin-PPMA (BA) is biased even when data is MAR. For other values of λ_A when $\lambda_T = 0$, binS-PPMA yields lower RMSE than both bin-PPMA estimates. When $\lambda_T > 0$, linearity assumptions between X and U fail, leading to biased estimates from bin-PPMA. Not surprisingly, binS-PPMA shows gains in both bias and RMSE at the correct λ_A, while achieving nominal coverage. When $\lambda_T > 0$ but $\lambda_A \neq \lambda_T$, binS-PPMA demonstrates consistent improvements in RMSE compared to bin-PPMA when $\lambda_A > 0$, while having a minor penalty under $\lambda_A = 0$. The latent variable model yields higher RMSE and CIW compared to binS-PPMA when missingness depends linearly on U^*, possibly due to a low correlation between U and V further attenuated by the iterative samplings required to estimate both latent variables. When the nonresponse mechanism is quadratic on U^*, the latent variable model produces biased estimates at $\lambda_T > 0$, as the model is misspecified in assuming V is linear on U^*. Bias is particularly apparent at $\lambda_T = \infty$, where non-coverage is high. As expected LOGREG produces biased estimates with low coverage under MNAR, particularly when missingness depends linearly on U^*.

4.4.4 Scenario 4: binary Z

In this scenario we simulate data in a 2x2 contingency table where both the predictor, Z, and the outcome, Y, are binary. We again generate Y through a latent variable U:

$$Z \sim Bernoulli(0.5)$$

$$U|Z \sim N(-0.75 + 1.5Z, 1)$$

$$Y = \begin{cases} 0 \text{ if } U < 0 \\ 1 \text{ if } U \geq 0 \end{cases}$$

Let the response of Y be determined by the following models:

A. $V|Z, U \sim N(-0.5 + Z, 1)$ $(\lambda_T = 0)$

B. $V|Z, U \sim N(-0.25 + 0.5I\{1.7Z + U > 0\}, 1)$ $(\lambda_T = 1)$

C. $V|Z, U \sim N(-0.25 + 0.5I\{U > 0\}, 1)$ $(\lambda_T = \infty)$

where Y is observed if $V < 0$ and missing otherwise.

We again iteratively estimate U and X based on a regression of U on Z for the respondents and apply the methods. Figure 4.4 displays average bias, RMSE, CIW, and non-coverage for each method where $\lambda_A = \lambda_T$. Estimates from CC are severely biased at all nonresponse mechanisms, while LOGREG is valid only under MAR. When $\lambda_A = \lambda_T = 0$, there are little differences between the methods in terms of bias, RMSE, CIW, and non-coverage, except for bin-PPMA (BA) which sees a small increase in RMSE due to deviation from normality. When $\lambda_T > 0$, performances of bin-PPMA (MI), binS-PPMA, and the latent variable model are similar in terms of RMSE and CIW. Non-coverages under the correct assumption of λ are also near the nominal 5% for all methods (except CC and LOGREG) when $\lambda_T = 0$ or 1. However both bin-PPMA and binS-PPMA undercover when λ_A

= λ_T = ∞, with binS-PPMA to a larger extent due to a more anti-conservative coverage. At other values of λ_A that do not correspond to the true λ_T, bin-PPMA (MI) and binS-PPMA produce similar RMSE and CIW.

4.5. Example: asthma symptoms study

We consider data from a child asthma study conducted at the University of Michigan. The aim of the study is to evaluate the effectiveness of an educational intervention in reducing asthma symptoms for children. Data is collected from children in Detroit elementary and middle schools. The primary outcome is the average number of nights the child experiences asthma symptoms per month, collected at baseline and one-year follow-up. For this exercise, we are interested in estimating the proportion of children in the control group that experienced a decrease in monthly symptoms from baseline to follow-up. However, since response may be influenced by the health of the child, we apply S-PPMA and PPMA to assess for nonresponse bias.

In our analysis we limit our sample to children who have experienced 1 to 15 nights of symptoms per month at baseline, since children with no symptoms at baseline will not observe any improvement in outcome. Out of 472 children at baseline in our analysis sample, 167 (35%) were lost to follow-up. Improvement of symptoms is highly associated with both age at baseline ($p = 0.01$) and baseline nights of symptoms per month ($p=0.04$). Moreover, age is highly predictive of response status ($p<0.01$). Thus, we use age and baseline monthly symptoms as predictors to obtain our proxy X, and apply binS-PPMA and bin-PPMA over $\lambda = 0, 1, 4$, and ∞ via the data augmentation approach to estimate proportion with improvement.

Figure 4.5 shows estimates from binS-PPMA (PS) and bin-PPMA of proportion of children who experienced a decrease in monthly asthma symptoms at follow-up. We apply bin-PPMA under both Bayes (BA) and multiple imputation (MI). Each line represents the estimated mean and 95% confidence interval. We can see that estimates of the proportion with improvement in monthly symptoms increase as we place more weight on the outcome with respect to nonresponse, suggesting that healthier children were less likely to remain in the study. Differences between binS-PPMA and bin-PPMA are small in general, with binS-PPMA being slightly less sensitive to assumptions about λ, as its range of estimates over λ are smaller than those of bin-PPMA. At $\lambda = 0$, binS-PPMA and bin-PPMA are similar to complete case analysis. Estimates at $\lambda > 0$ are noticeably higher than those at $\lambda = 0$, which is an indication that inferences are sensitive to assumptions of MAR. However, results show little differences from $\lambda = 1$ to $\lambda = \infty$.

Since healthier children may have less of an incentive to participate in an asthma study, it is reasonable to assume MNAR in our data. Based on the results, one may choose an intermediate value of $\lambda = 1$ as it represents a middle ground between MAR and MNAR. However, even at $\lambda = \infty$, the most extreme case of MNAR, estimates tend to be similar.

4.6. Discussion

In this chapter we extend S-PPMA to binary outcomes through data augmentation. MAR assumptions are often not reasonable, in which case potential nonresponse bias due to MNAR should be explored. Bin-PPMA and binS-PPMA allow us to assess for nonresponse bias without requiring MAR. Unfortunately, the data provides no information about the true value of λ, thus we must rely on a sensitivity analysis over a range of λ to reflect our uncertainty about λ. Moreover, a sensitivity analysis captures potential bias and uncertainty about the missing data mechanism without the need to specify a propensity model. Although negative values of λ may be included in the analysis, since *X* is a prediction of *U* it is reasonable to assume that λ is positive. As with bin-PPMA, binS-PPMA summarizes information in Z_1, …, Z_p by reducing them into a single variable *X* that is predictive of our outcome, which facilitates nonresponse assessment. However, unlike bin-PPMA, binS-PPMA does not assume bivariate normality of *X* and *U*, making it more robust to deviations from normality. Simulations have shown binS-PPMA produces gains in bias and RMSE compared to bin-PPMA when data is non-normal, and performs similarly under normality.

Our sensitivity analysis considers mechanisms where missingness depends on the value of $X + \lambda U$. Since *X* is a function of the covariates Z_1, …, Z_p, the model implicitly weighs the importance of each Z_p on response based on their estimated coefficients from the regression of *U* on Z_1, …, Z_p. In reality missingness may depend on some other combinations or a subset of *Z*, which can have an effect on the estimates. Additional sensitivity parameters can be used to address this issue, at the expense of reducing model simplicity.

Performance gains of binS-PPMA over bin-PPMA depends highly on the degree of model misspecification at the threshold (i.e. $U = 0$), since Y is imputed based on the sign of the estimated U. Thus, when nonlinearity is present near the threshold, we expect gains from binS-PPMA. However, if nonlinearity is only apparent at values of U far away from 0, we would not see much gains from the spline. The simulations in this study are set up to accentuate the differences between binS-PPMA and bin-PPMA. When both the predictor and outcome are binary, binS-PPMA still produces gains in RMSE over other methods in our simulation. In terms of confidence interval estimation and coverage, however, the proposed hotdeck procedure of imputation is less successful as it fails to achieve nominal coverages. Further adjustments to this procedure for categorical predictors are needed.

In our examples we assume that the regression models used to estimate the proxy are correctly specified, which may not always be the case. An incorrectly specified model may introduce bias. Thus robustness to model misspecification should be further explored. Furthermore, the mechanism that generates Y may not be the result of a latent variable. For example, Y may be generated under a logistic model given Z. Validity of the methods depends on whether there exists a set of α such that (8) approximates the true mechanism.

Figure 4.1a. Results for scenario 1 when missingness depends on $U^* = X + \lambda_T U$ and $\lambda_A = \lambda_T$.

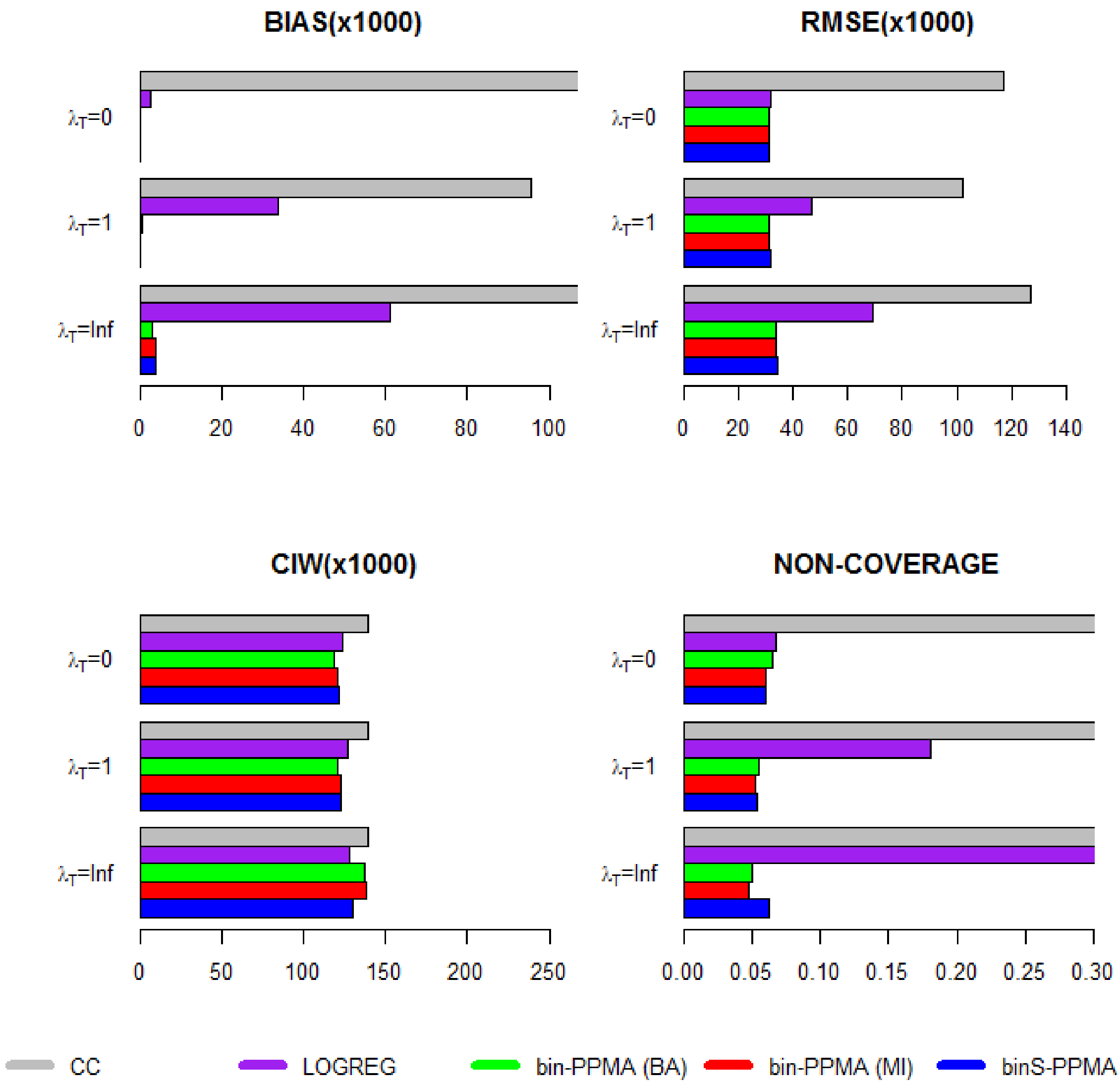


Figure 4.1b. Results for scenario 1 when missingness depends on $U^{*2} = (X + \lambda_T U)^2$ and $\lambda_A = \lambda_T$.

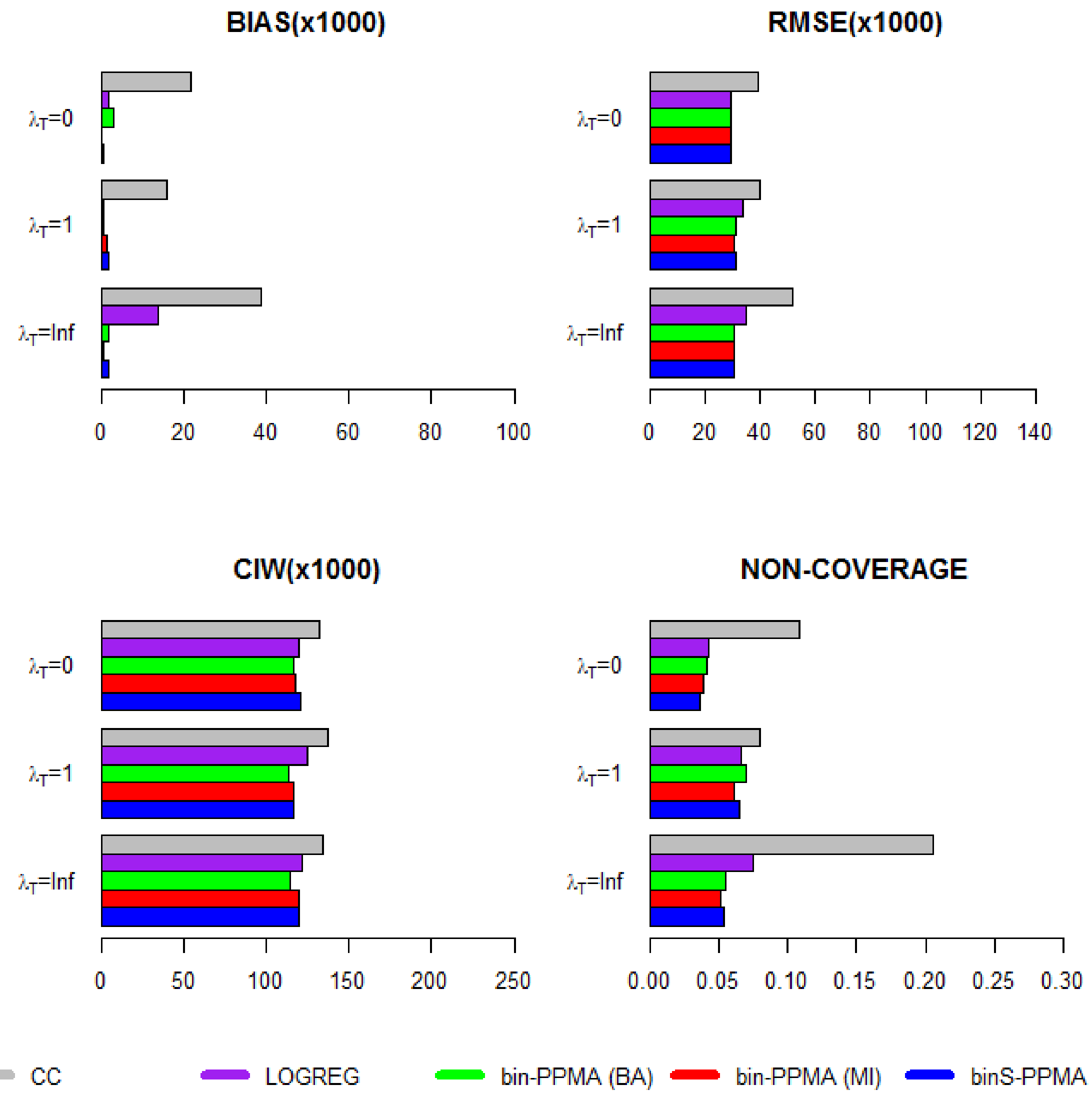


Figure 4.2a. Results for scenario 2 when missingness depends on $U^* = X + \lambda_T U$ and $\lambda_A = \lambda_T$.

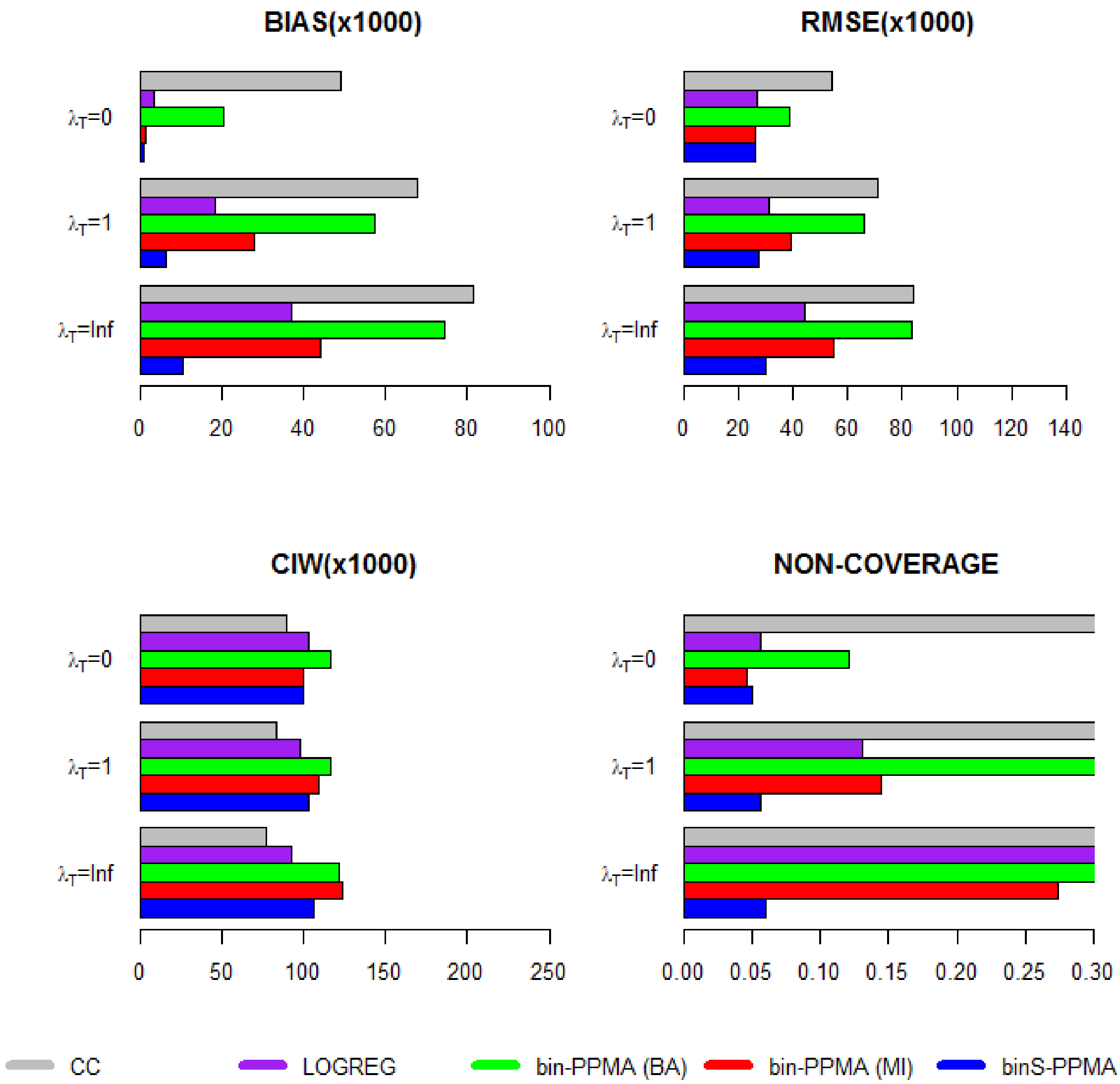


Figure 4.2b. Results for scenario 2 when missingness depends on $U^{*2} = (X + \lambda_T U)^2$ and $\lambda_A = \lambda_T$.

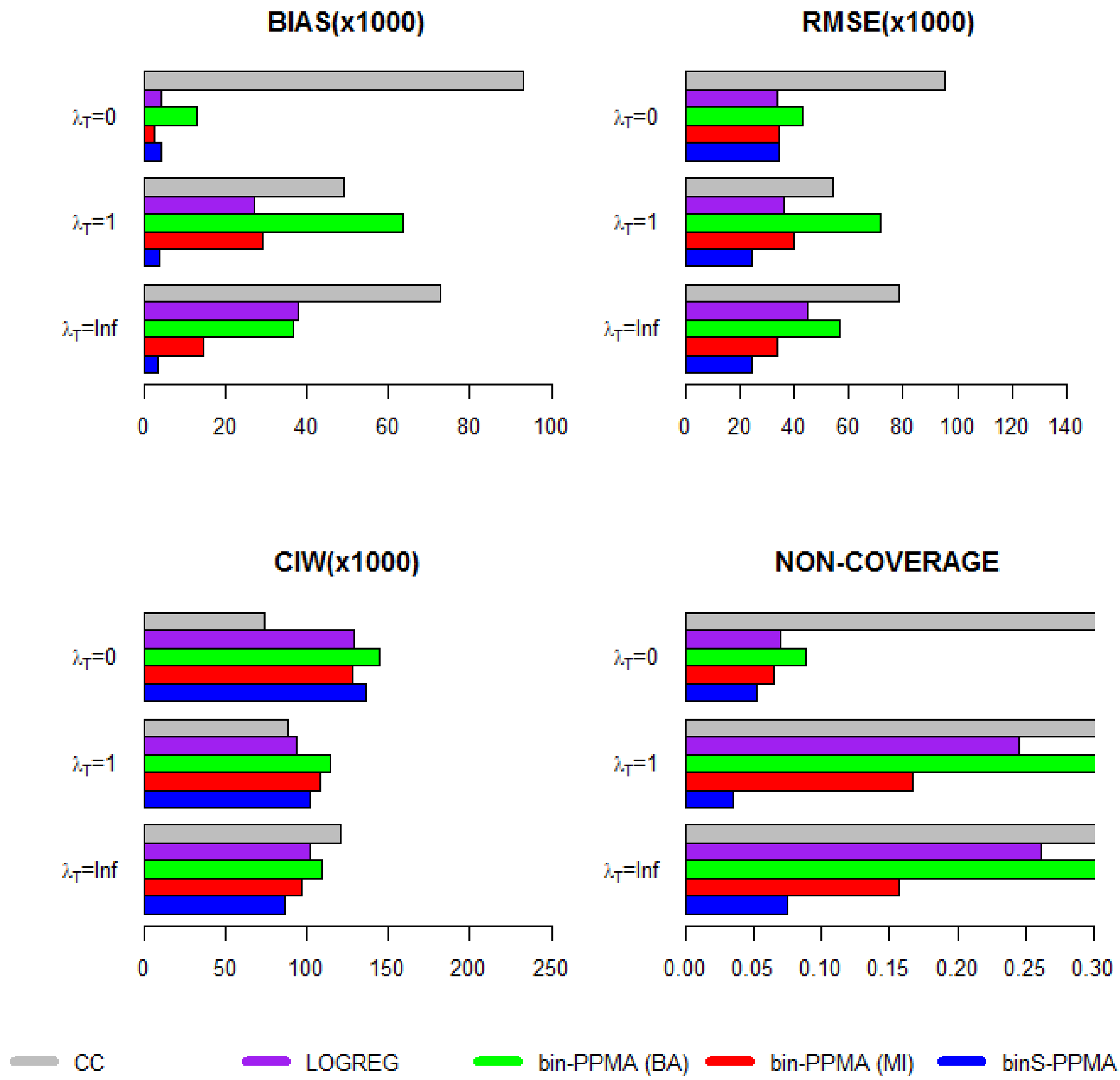


Figure 4.3a. Results for scenario 3 when missingness depends on $U^* = X + \lambda_T U$ and $\lambda_A = \lambda_T$.

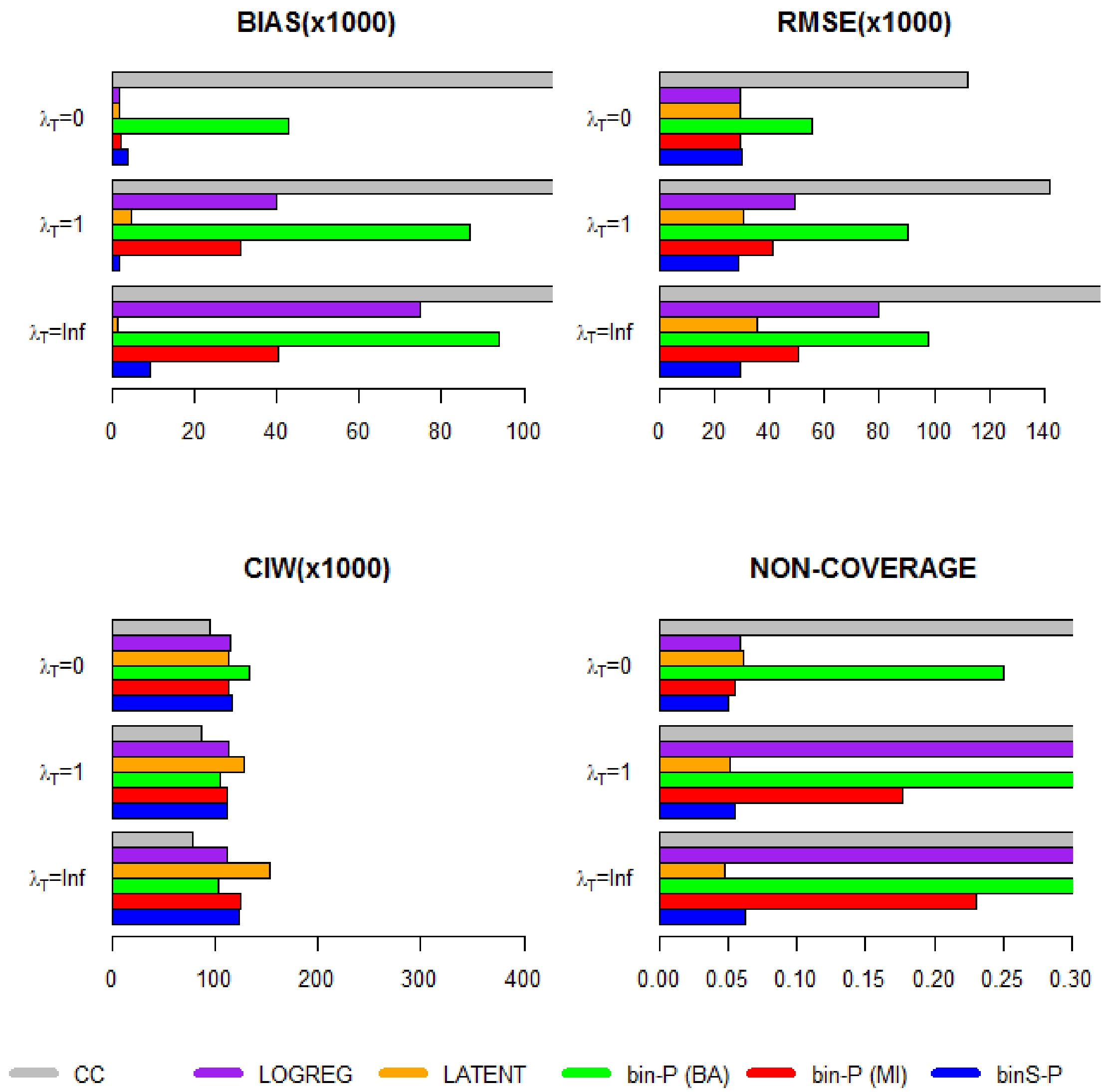


Figure 4.3b. Results for scenario 3 when missingness depends on $U^{*2} = (X + \lambda_T U)^2$ and $\lambda_A = \lambda_T$.

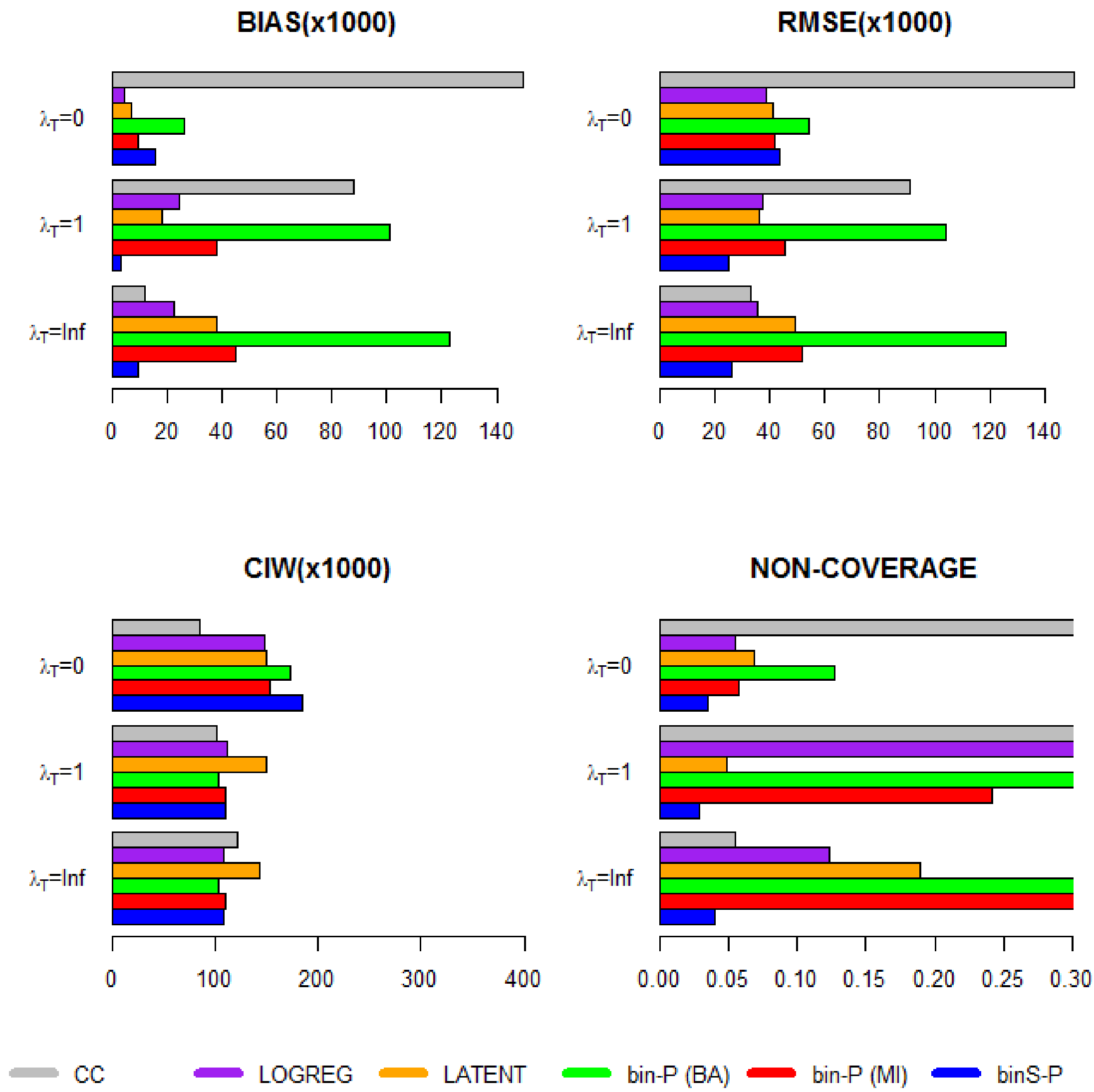


Figure 4.4. Results for scenario 4 when missingness depends on $U^* = X + \lambda_T U$ and $\lambda_A = \lambda_T$.

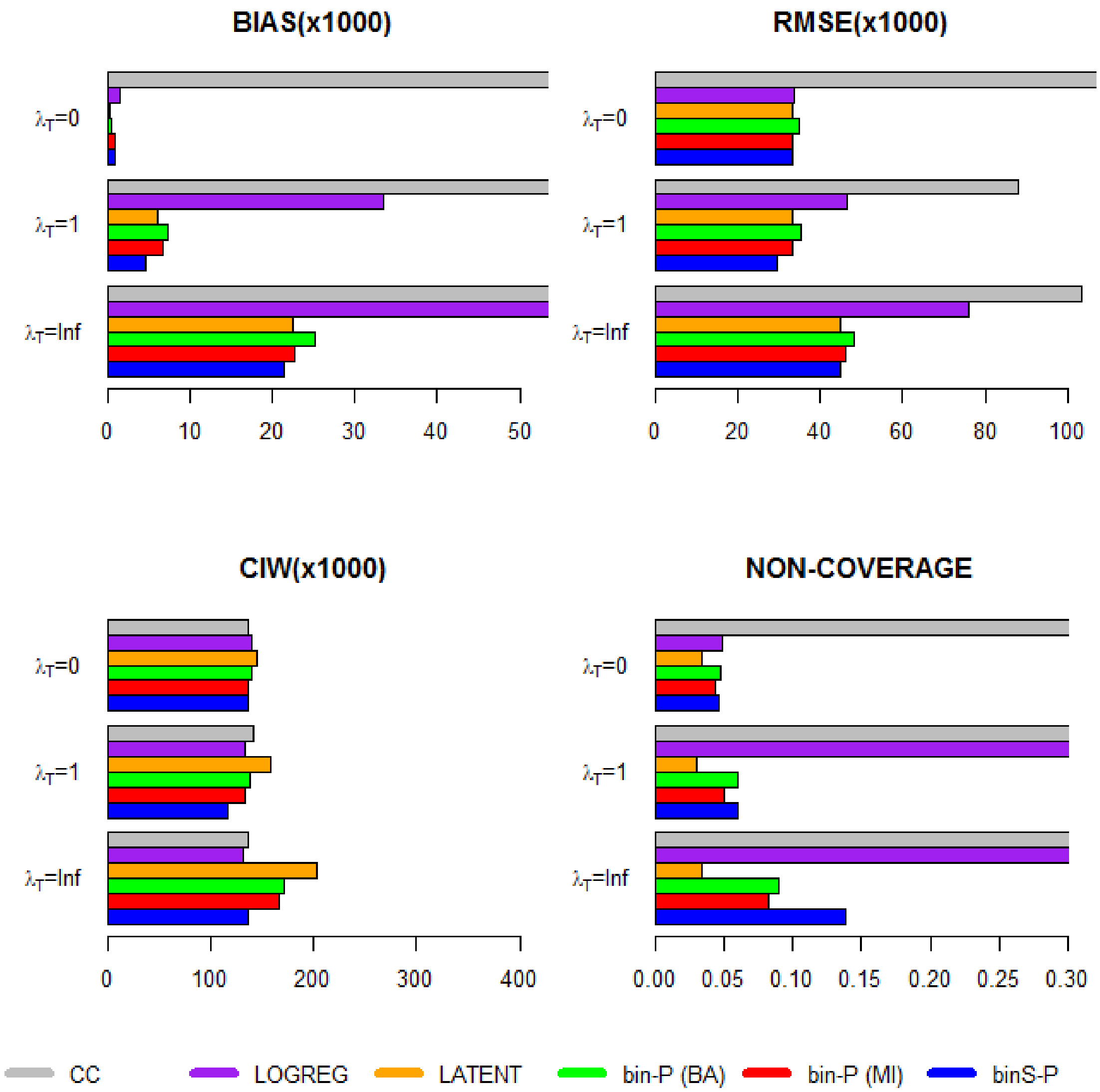
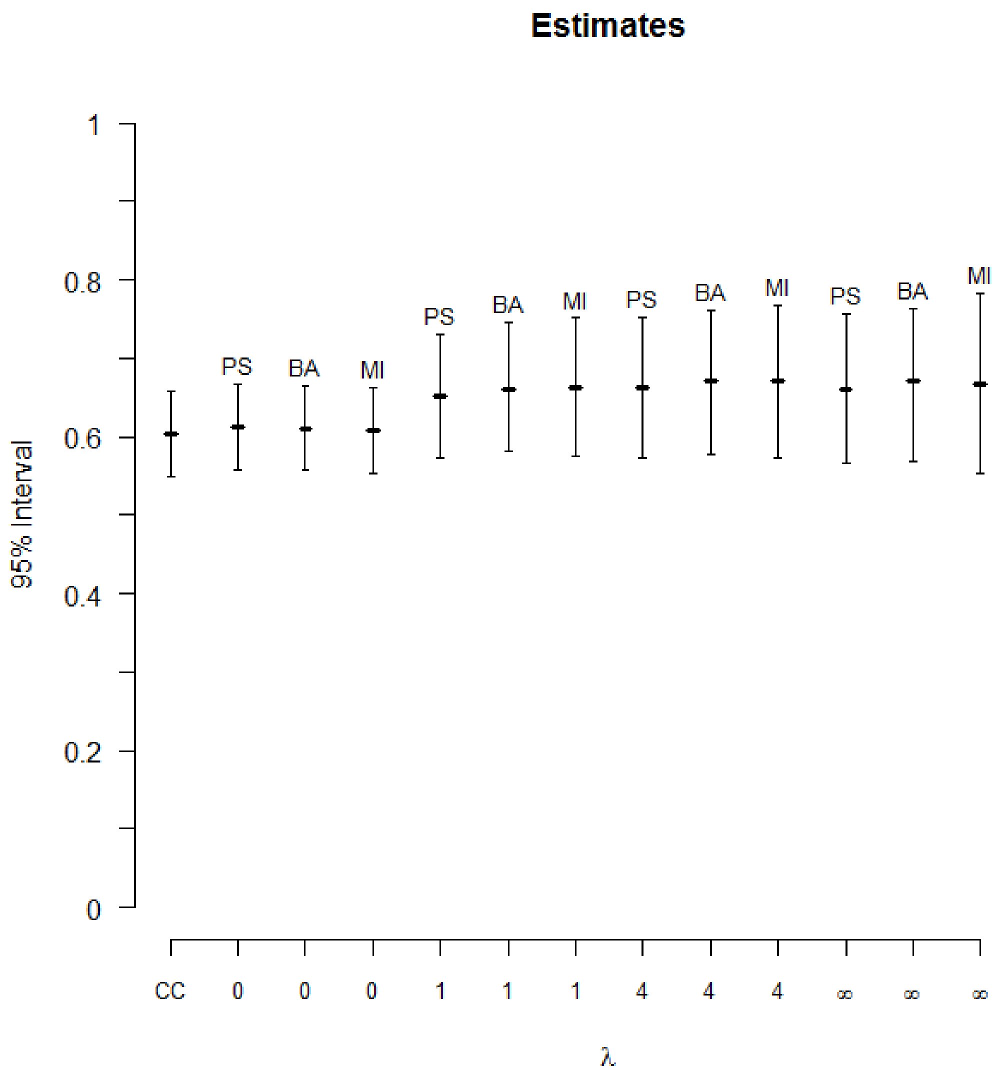


Figure 4.5. Estimates for proportion with reduced asthma symptoms at follow-up.



CHAPTER V

Summary and Future Work

This dissertation focuses on developing and comparing robust estimators of the mean of a single variable subject to missing data, in the presence of information from observed covariates. In the second chapter, we assume data that are MAR. Many methods of estimating the mean are available, such as complete case analysis, imputation, and weighting methods. However, concerns about model misspecification have led to the development of DR estimators. An estimator is DR is it is consistent when either the model for the response propensity or the model for the mean is correctly specified. Here we compare performances of five DR estimators with respect to RMSE, CIW, and coverage. The results show that when the propensity model is correctly specified but the mean model is not, DR outperforms the incorrect regression model, as promised by their DR property. Overall, PSPP and the robust calibration method of Cao, et al (2009) yield the lowest RMSE and CIW. The calibration mean regression by weighted least squares calibration produces gains over the ordinary least squares counterpart, while the division of the weighted residuals by the sum of the weights also yields minor but consistent gains over its division by *n*. When the mean function is correctly specified, we find no distinguishable differences between the DR methods.

DR estimators are biased when missingness is not a random. The third chapter considers data that are potentially MNAR. We propose a spline pattern mixture model to the relationship between a continuous response variable *Y* and a fully observed covariate *X* for a given assumption about the missing data mechanism. In the case of multiple observed covariates, *X* is taken to be predicted value of *X* for the sample obtained by regression of *Y* on *X* over the respondents. The spline modelling approach is a modification of the pattern mixture models proposed in Little (1994) and Andridge and Little (2011), which replaces the linear regression between *X* and *Y* via a spline model, allowing for a non-linear relationship between *X* and *Y* and hence relaxes the bivariate normal assumption. Simulations show that the spline pattern mixture approach provides improved robustness to normality assumptions, while trading off some precision when normality holds compared to the linear models in Little (1994) and Andridge and Little (2011). As in all pattern mixture models we recommend a sensitivity analysis to reflect our uncertainty about the nonresponse mechanism.

The fourth chapter extends the idea of the spline pattern mixture model to categorical outcomes. We assumed a continuous latent variable which determines the value of the categorical outcome. We then apply the spline pattern mixture model to the observed *X* and the iteratively estimated latent variable to obtain our estimates of the mean, where *X* is the predicted value of the latent variable for the sample. Simulation results show that, similar to results for continuous outcomes, the spline pattern mixture model provides robustness to bivariate normality assumptions between the latent variable and *X*. When *X* is continuous, the spline model yields approximately unbiased estimates of the mean with close to nominal coverage. The method, however, is less successful when *X* is categorical, where the model undercovers the true mean.

Although Chapter IV restricts attention to binary outcomes, the ideas presented can be generalized to categorical variables of more two categories. In the case of ordinal outcomes with $k > 2$ levels, we may specify k -1 threshold points for the latent *U*, and impute *Y* from the proxy pattern mixture model based on the imputed value of *U* with respect to the thresholds. For nominal outcomes, we may apply the binary pattern mixture models over $k - 1$ steps, where at each step *j* we model the probability that *Y* belongs in group *j* given *Y* does not belong in group *j* -1. For example, suppose *Y* is nominal taking values of 1, 2, or 3. We may apply the proposed methods to first model the probability that $Y = 1$, then re-apply the methods over subjects for which $Y \neq 1$ to model the probability that $Y = 2 \mid Y \neq 1$. This approach, however, assumes $k - 1$ latent variables and hence requires $k - 1$ sensitivity parameters, which increases complexity of the model. This can be troublesome when the value of *k* is large.

Methods discussed in this dissertation concern data with a single missing outcome and a set of fully observed auxiliary variables. We may incorporate S-PPMA into chained equations where more than one variable is missing. Suppose Z_1, …, Z_k are fully observed, Y_1, .., Y_j are partially missing, and R_j is the response indicator for Y_j. At each iteration *d*, we create imputations for Y_1, .., Y_j sequentially:

1. Estimate proxy $X_1^{(d)}$ from $Y_1 | Y_2^{(d-1)}, \ldots, Y_j^{(d-1)}, Z_1, \ldots, Z_k, R_1$ = 1, where $Y_j^{(d)}$ is Y_j with missing values imputed at the d^{th} iteration and $X_1^{(d)}$ is the prediction of Y_1 for the whole sample. Apply S-PPMA to $X_1^{(d)}$ and Y_1 for a given λ_1 to obtain $Y_1^{(d)}$.
2. Estimate proxy $X_2^{(d)}$ from $Y_2 | Y_1^{(d)}, Y_3^{(d-1)}, \ldots, Y_j^{(d-1)}, Z_1, \ldots, Z_k, R_2$ = 1. Apply S-PPMA to $X_2^{(d)}$ and Y_2 for a given λ_2 to obtain $Y_2^{(d)}$.

.
.
.

j. Estimate proxy $X_j^{(d)}$ from $Y_j | Y_1^{(d)}, \ldots, Y_{j-1}^{(d)}, Z_1, \ldots, Z_k, R_j$ = 1. Apply S-PPMA to $X_j^{(d)}$ and Y_j for a given λ_j to obtain $Y_j^{(d)}$.

This approach does not assume a particular missing data pattern, thus may be used to for a variety of multivariate missing data. For a categorical Y_j, we replace S-PPMA with binS-PPMA. The equations make untestable assumptions about the nonresponse mechanism for each missing Y_i, and any sensitivity analysis can become cumbersome, particularly with a large number of missing variables. Not all variables may be predictive of each other, though this can be modified to include only those that are predictive of Y_i in its regression. Lastly, convergence properties would need to be assessed.

Finally, we have only discussed methods for estimating the marginal mean of a variable. Models for estimating subgroup means need to be further explored. Moreover, there is considerable interest in estimating regression coefficients under data that is

MNAR. Methods incorporating a sensitivity analysis to assess nonresponse bias in regression coefficients, similar to the idea in pattern mixture models, may be developed.

Appendix

A.1. Gibbs sampling procedure for Bayesian penalized spline model

In this dissertation we consider a Bayesian penalized spline with homoscedastic errors. Suppose we model *Y* on a penalized spline of *P*, where

$$Y \sim N(X\beta + Z\gamma, \sigma^2)$$

$$X = \begin{pmatrix} 1 & P_1 \\ \vdots & \vdots \\ 1 & P_r \end{pmatrix}$$

$$Z = \begin{pmatrix} (P_1 - \kappa_1)_+ & \cdots & (P_r - \kappa_k)_+ \\ \vdots & \ddots & \vdots \\ (P_r - \kappa_1)_+ & \cdots & (P_r - \kappa_k)_+ \end{pmatrix}$$

$$(P_i - \kappa_k)_+ = \begin{cases} (P_i - \kappa_k)\ if\ (P_r - \kappa_k) > 0 \\ 0\ otherwise \end{cases}$$

and *r* is the number of respondents. Here, we have *K* knots in the model represented by *Z*. We assign the following non-informative priors:

$$\beta \sim 1$$

$$\gamma \sim N(0, \tau^2 I)$$

$$\sigma^2 \sim InvGamma(10^{-5}, 10^{-5})$$

$$\tau^2 \sim InvGamma(10^{-5}, 10^{-5})$$

Estimates of the joint posterior distributions of the parameters, along with the posterior predictive distribution of the missing values of *Y* are obtained via the Gibbs sampling algorithm. The procedure at the d^{th} iteration is summarized as follows:

1. Draw $[\beta^{(d)}, \gamma^{(d)}] | X, Z, \sigma^{2(d-1)}, \tau^{2(d-1)} \sim N((C'C + \frac{\sigma^{2(d-1)}}{\tau^{2(d-1)}}D)^{-1}C'Y, \sigma^{2(d-1)}(C'C + \frac{\sigma^{2(d-1)}}{\tau^{2(d-1)}}D)^{-1})$, $C = [X\ Z]$, $D = \begin{pmatrix} 0_{2x2} & 0_{2xK} \\ 0_{Kx2} & 1_{KxK} \end{pmatrix}$.
2. Draw $\tau^{2(d)} | X, Z, \sigma^{2(d-1)}, \beta^{(d)}, \gamma^{(d)} \sim InvGamma(10^{-5} + \frac{K}{2}, 10^{-5} + \frac{1}{2}\|\gamma\|^2)$
3. Draw $\sigma^{2(d)} | X, Z, \tau^{2(d)}, \beta^{(d)}, \gamma^{(d)} \sim InvGamma(10^{-5} + \frac{r}{2}, 10^{-5} + \frac{1}{2}(Y - X\beta^{(d)} - Z\gamma^{(d)})' (Y - X\beta^{(d)} - Z\gamma^{(d)}))$
4. Impute $Y_{mis} | X, Z, \sigma^{2(d)}, \tau^{2(d)}, \beta^{(d)}, \gamma^{(d)} \sim N(X\beta^{(d)} + Z\gamma^{(d)}, \sigma^{2(d)})$
5. Repeat steps 1-4 for total 10000 iterations, discarding the first 1000 as burn-in.

The following tables display complete results for simulations in Chapters II-IV.

Figure A2.1. Y vs. X_1 for respondents of n = 800 from Chapter II simulation 1

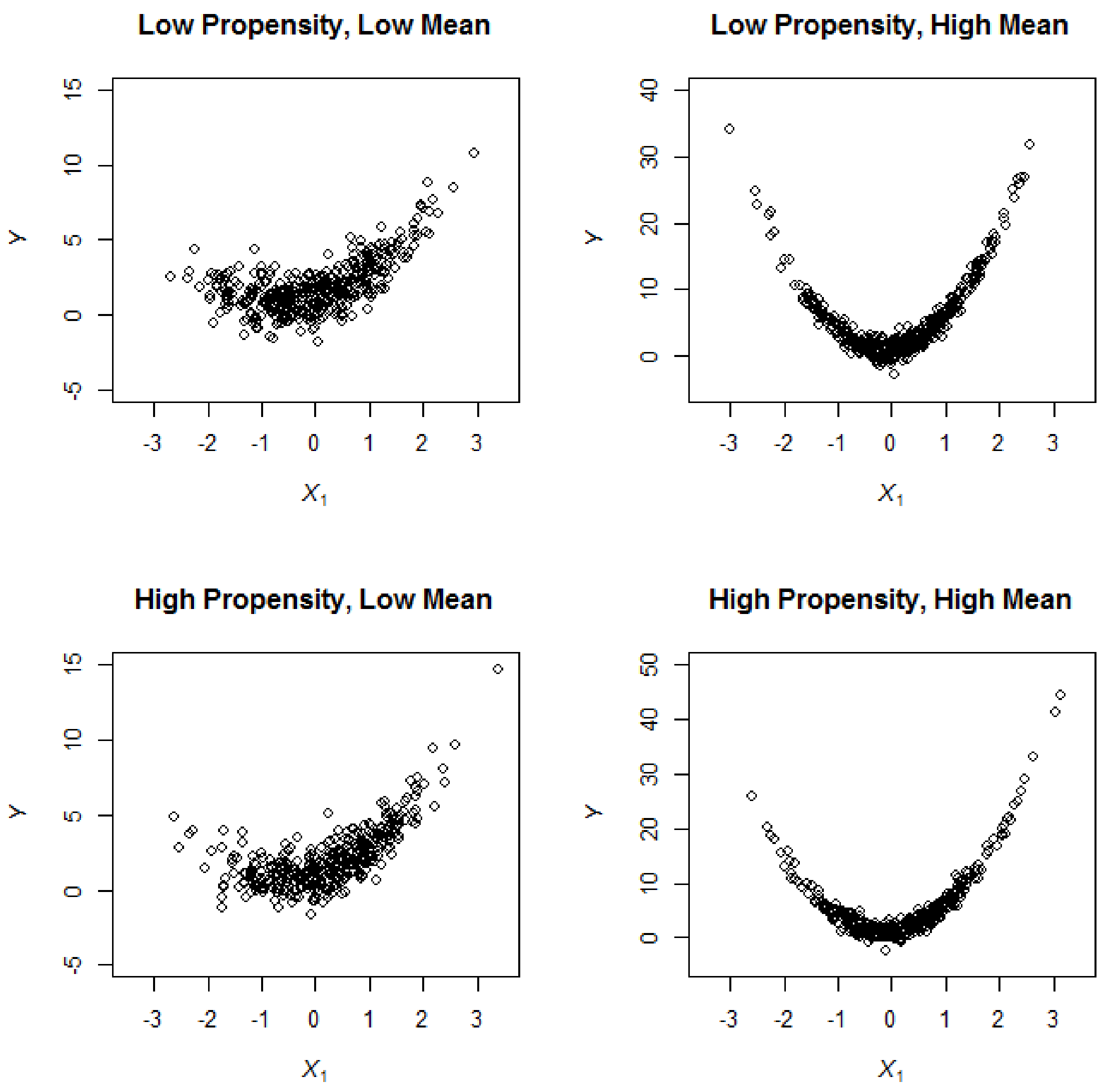


Figure A2.2. Y vs. X_1 and X_2 for respondents of n = 800 from Chapter II simulation 2

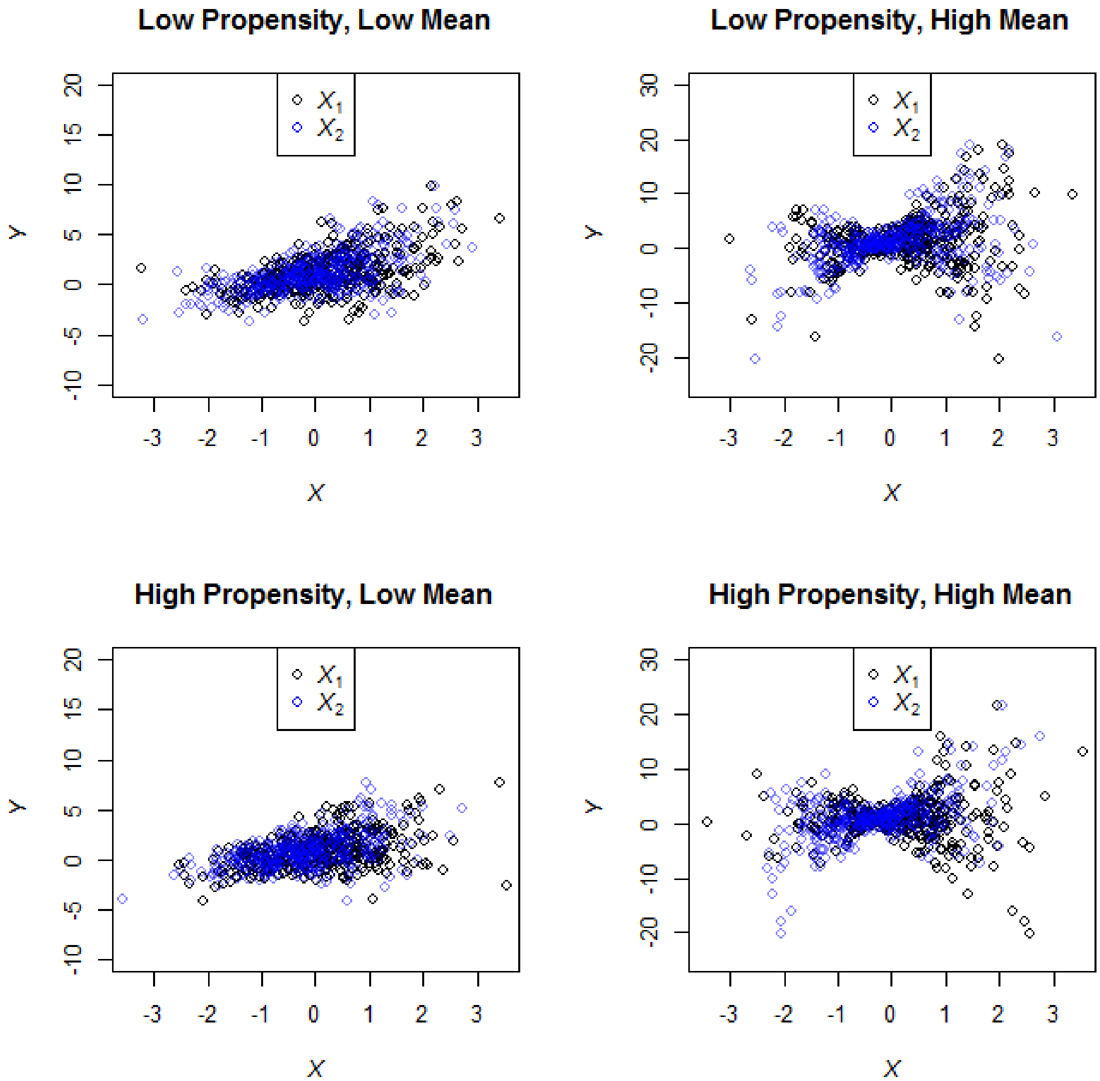


Figure A2.3. Y vs. X_1 and X_2 for respondents of n = 800 from Chapter II simulation 3

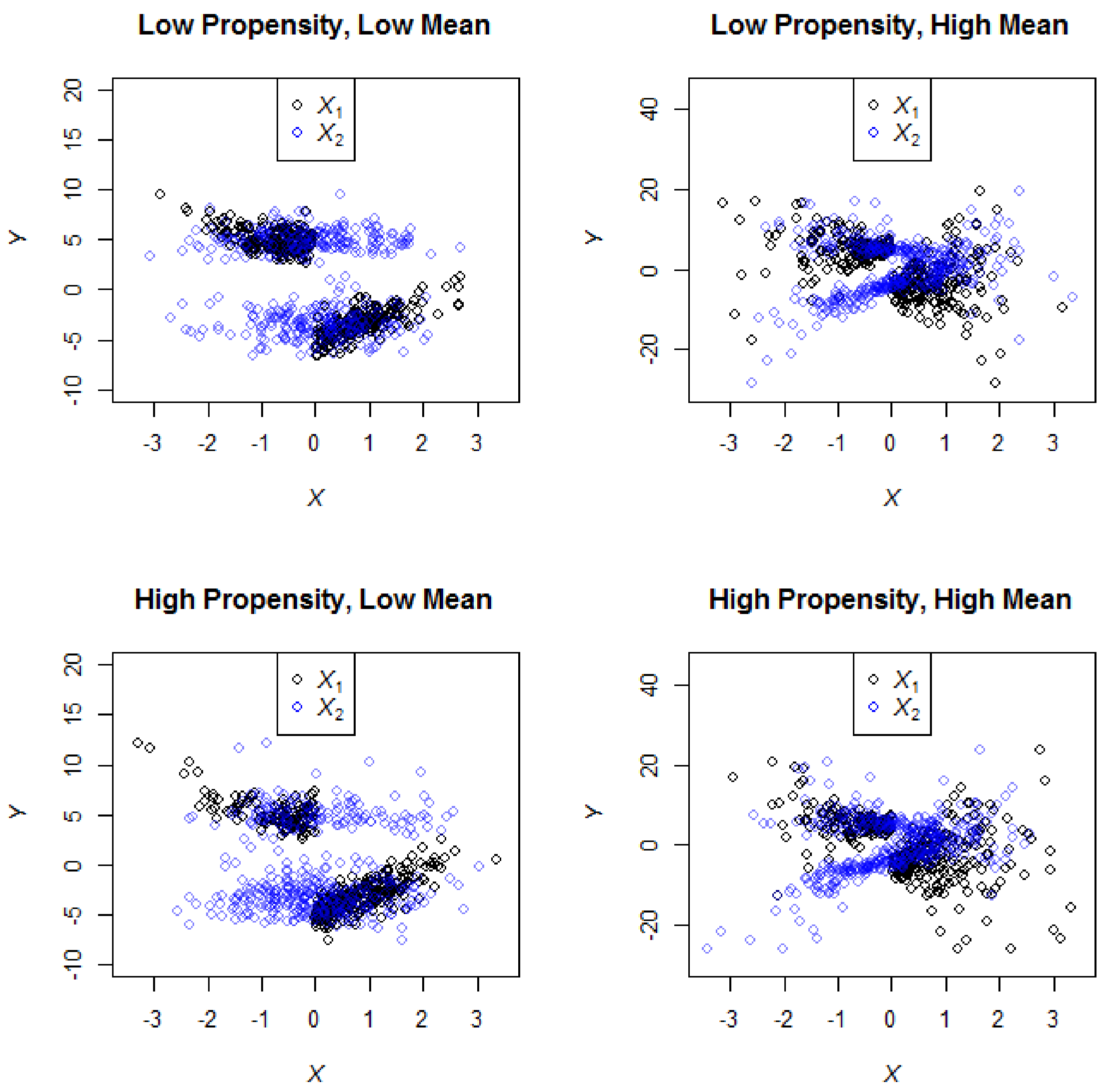


Table A2.1a. Results from Chapter II simulation 1 (LL)

N	BD	CORR	MISS	CAL	MCAL	WCAL	RCAL	$PSPP_{REML}$	$PSPP_{BAYES}$
				Bias x 1000					
50	10	6	-29	-1	-1	-3	-8	-19	-26
100	-6	-2	-19	-1	-1	-1	-8	-8	-12
200	-1	2	-12	-1	-1	-1	-4	-1	-3
400	-1	-2	-11	-3	-3	-3	-4	-4	-4
800	3	2	-2	3	3	3	3	1	1
				% increase in RMSE over BD					
50	0	17	36	40	39	37	35	22	22
100	0	15	31	32	32	31	25	16	16
200	0	15	30	31	31	30	26	16	15
400	0	15	28	28	28	28	23	16	15
800	0	12	29	29	29	29	21	13	12
				% increase in CIW over BD					
50	0	24	32	38	36	32	45	27	22
100	0	16	31	34	34	31	30	21	18
200	0	15	31	32	32	31	27	18	17
400	0	14	31	32	32	31	27	15	16
800	0	15	31	32	32	31	26	15	15
				Coverage out of 1000					
50	946	971	950	952	952	949	970	962	962
100	925	935	932	938	938	936	945	937	934
200	936	940	942	949	949	949	941	946	949
400	947	948	952	954	954	953	951	948	952
800	955	965	965	964	964	964	973	966	968

Table A2.1b. Results from Chapter II simulation 1 (LH)

N	BD	CORR	MISS	CAL	MCAL	WCAL	RCAL	$PSPP_{REML}$	$PSPP_{BAYES}$
				Bias x 1000					
50	-44	-46	-157	-18	-19	-28	-75	-136	-132
100	-13	-10	-161	-70	-71	-72	-75	-57	-55
200	-21	-20	-62	-5	-5	-7	-24	-37	-37
400	0	-4	-61	-22	-22	-22	-18	-13	-13
800	-1	-2	-23	7	7	7	-5	-6	-6
				% increase in RMSE over BD					
50	0	2	43	48	47	43	36	3	3
100	0	2	51	53	53	50	39	3	3
200	0	2	46	48	48	46	35	2	2
400	0	2	43	43	43	42	27	2	2
800	0	2	42	42	42	42	26	2	2
				% increase in CIW over BD					
50	0	3	41	54	50	42	35	4	0
100	0	2	38	44	43	39	33	2	1
200	0	2	40	43	42	40	33	1	1
400	0	1	40	42	42	41	31	1	2
800	0	2	41	42	42	41	28	1	2
				Coverage out of 1000					
50	930	930	898	910	910	905	922	918	911
100	936	938	898	904	905	903	931	932	931
200	923	918	908	922	922	921	933	920	922
400	946	945	922	929	929	927	955	944	945
800	949	944	932	937	937	937	952	938	945

Table A2.1c. Results from Chapter II simulation 1 (HL)

N	BD	CORR	MISS	CAL	MCAL	WCAL	RCAL	$PSPP_{REML}$	$PSPP_{BAYES}$
				Bias x 1000					
50	5	4	-119	-30	-31	-35	-25	-35	-56
100	-1	-3	-103	-16	-17	-20	-11	-19	-28
200	-4	-2	-97	-5	-5	-8	-4	-9	-12
400	1	1	-93	-4	-4	-5	-1	-3	-5
800	5	4	-89	2	1	1	3	1	1
				% increase in RMSE over BD					
50	0	14	41	45	43	37	37	18	21
100	0	16	48	51	50	41	24	21	18
200	0	18	59	50	49	42	20	22	20
400	0	14	68	42	41	36	14	15	15
800	0	15	87	42	42	37	15	16	15
				% increase in CIW over BD					
50	0	32	36	58	45	36	86	36	29
100	0	19	35	48	45	36	36	29	23
200	0	17	36	48	47	39	22	23	20
400	0	17	35	45	45	38	18	19	18
800	0	16	35	46	45	39	17	17	17
				Coverage out of 1000					
50	930	967	923	940	939	934	974	957	951
100	943	953	921	946	945	944	960	959	953
200	949	947	898	950	949	946	952	951	951
400	939	955	878	950	950	949	956	956	951
800	941	942	805	940	940	940	947	947	944

Table A2.1d. Results from Chapter II simulation 1 (HH)

N	BD	CORR	MISS	CAL	MCAL	WCAL	RCAL	$PSPP_{REML}$	$PSPP_{BAYES}$
				Bias x 1000					
50	-35	-37	-585	-86	-105	-154	-135	-193	-189
100	-16	-13	-494	-4	-10	-47	-55	-79	-78
200	0	2	-491	-32	-33	-39	-15	-30	-32
400	15	17	-450	8	7	2	7	-1	-1
800	4	5	-470	-25	-25	-25	-1	-3	-4
				% increase in RMSE over BD					
50	0	3	65	90	80	57	21	5	5
100	0	1	73	96	91	62	6	2	2
200	0	2	90	69	68	56	3	2	3
400	0	2	115	71	70	58	2	3	2
800	0	1	166	69	68	57	1	1	1
				% increase in CIW over BD					
50	0	4	44	91	64	45	33	5	-1
100	0	2	45	77	67	47	17	3	1
200	0	2	44	65	63	49	6	2	1
400	0	2	45	67	65	52	2	2	1
800	0	2	46	67	66	54	1	1	2
				Coverage out of 1000					
50	912	921	832	887	880	881	935	902	891
100	941	935	837	916	914	904	951	929	926
200	942	950	806	912	913	913	948	939	938
400	951	948	777	928	928	928	947	947	951
800	936	942	624	927	927	927	942	940	941

Table A2.2a. Results from Chapter II simulation 2 (LL)

N	BD	CORR	MISS	CAL	MCAL	WCAL	RCAL	$PSPP_{REML}$	$PSPP_{BAYES}$
					Bias x 1000				
50	12	16	22	15	15	15	18	13	18
100	-3	5	9	2	2	3	5	5	7
200	0	3	8	-1	-1	-1	-3	-1	2
400	-3	-2	10	1	1	1	-1	2	3
800	-3	-3	6	-3	-3	-3	-5	-3	-3
					% increase in RMSE over BD				
50	0	17	27	28	28	27	46	28	27
100	0	14	21	22	22	21	29	29	22
200	0	12	19	20	20	20	20	20	18
400	0	16	26	26	26	25	24	25	23
800	0	17	25	25	25	25	22	23	22
					% increase in CIW over BD				
50	0	21	25	31	28	25	105	53	30
100	0	16	23	25	25	23	35	41	25
200	0	14	21	23	23	22	24	31	22
400	0	14	22	23	23	22	22	26	22
800	0	13	21	22	22	22	20	23	22
					Coverage out of 1000				
50	944	943	942	949	946	940	978	964	946
100	945	948	951	950	950	948	963	965	956
200	948	948	940	942	942	942	950	956	947
400	960	943	938	946	946	945	944	954	943
800	950	930	935	940	940	940	939	942	941

Table A2.2b. Results from Chapter II simulation 2 (LH)

N	BD	CORR	MISS	CAL	MCAL	WCAL	RCAL	$PSPP_{REML}$	$PSPP_{BAYES}$
					Bias x 1000				
50	-26	-25	12	-22	-22	-23	-6	-20	-11
100	14	17	62	26	26	26	9	27	29
200	7	5	58	13	13	13	-1	17	16
400	7	7	47	1	1	1	-5	2	6
800	4	5	51	6	6	6	0	7	8
					% increase in RMSE over BD				
50	0	3	40	48	47	42	64	49	41
100	0	3	45	57	56	48	48	58	45
200	0	4	40	41	41	39	37	48	38
400	0	3	41	42	42	41	34	41	35
800	0	3	46	46	46	44	32	38	36
					% increase in CIW over BD				
50	0	5	43	63	52	42	101	93	41
100	0	3	39	51	48	40	52	87	36
200	0	3	40	44	44	41	42	73	34
400	0	3	39	43	43	41	37	55	33
800	0	3	39	42	42	40	34	45	33
					Coverage out of 1000				
50	939	948	952	963	958	953	979	984	949
100	964	966	948	959	957	952	965	986	946
200	954	952	948	949	949	949	959	981	948
400	944	947	935	946	946	944	951	970	942
800	929	933	923	935	935	935	942	947	932

Table A2.2c. Results from Chapter II simulation 2 (HL)

N	BD	CORR	MISS	CAL	MCAL	WCAL	RCAL	$PSPP_{REML}$	$PSPP_{BAYES}$
					Bias x 1000				
50	-14	-14	35	4	5	6	-10	-2	15
100	0	1	41	4	4	7	-4	6	19
200	0	3	46	5	6	7	0	9	16
400	0	0	43	2	2	2	-4	4	8
800	-3	-2	41	-2	-2	-2	-4	2	2
					% increase in RMSE over BD				
50	0	19	29	32	31	29	70	35	29
100	0	16	28	32	32	28	34	36	28
200	0	17	28	27	27	25	26	32	25
400	0	16	33	31	31	28	22	28	24
800	0	19	41	31	30	28	22	27	26
					% increase in CIW over BD				
50	0	26	28	42	32	28	168	67	32
100	0	18	24	32	29	25	55	52	24
200	0	16	23	28	27	24	30	41	21
400	0	15	24	28	28	25	23	32	20
800	0	15	23	28	27	25	20	27	18
					Coverage out of 1000				
50	941	954	940	945	936	933	996	964	954
100	947	942	946	950	949	944	970	966	936
200	944	936	936	945	945	944	953	959	935
400	952	945	924	936	935	936	940	942	938
800	947	934	911	944	944	941	942	948	937

Table A2.2d. Results from Chapter II simulation 2 (HH)

N	BD	CORR	MISS	CAL	MCAL	WCAL	RCAL	$PSPP_{REML}$	$PSPP_{BAYES}$
					Bias x 1000				
50	-2	-5	227	88	91	88	45	76	120
100	-1	1	218	38	40	46	-14	66	82
200	18	19	234	48	49	51	-2	44	64
400	10	11	225	16	16	19	-3	24	29
800	-4	-5	217	5	5	6	-6	5	6
					% increase in RMSE over BD				
50	0	4	51	66	63	51	81	58	52
100	0	4	51	59	58	49	47	72	53
200	0	3	60	59	58	50	34	53	43
400	0	3	75	60	59	50	25	50	42
800	0	4	98	56	56	48	18	45	37
					% increase in CIW over BD				
50	0	5	42	76	55	41	138	102	41
100	0	4	41	59	54	42	64	104	36
200	0	3	42	57	56	46	44	99	33
400	0	3	43	57	56	47	31	69	30
800	0	3	43	57	56	49	23	55	30
					Coverage out of 1000				
50	951	955	937	952	943	937	985	983	946
100	944	940	926	938	936	931	961	977	926
200	955	957	912	947	947	944	960	980	929
400	946	951	893	944	944	942	959	963	929
800	942	941	826	956	956	954	956	967	947

Table A2.3a. Results from Chapter II simulation 3 (LL)

N	BD	CORR	MISS	CAL	MCAL	WCAL	RCAL	$PSPP_{REML}$	$PSPP_{BAYES}$
				Bias x 1000					
50	9	-2	-50	-21	-21	-21	-35	-29	-24
100	-4	-4	12	30	30	29	18	1	6
200	-17	-22	-33	-23	-23	-23	-26	-25	-24
400	-1	-5	-10	-2	-2	-2	-3	-7	-6
800	-12	-12	-15	-10	-10	-10	-11	-12	-12
				% increase in RMSE over BD					
50	0	7	28	26	26	25	26	14	14
100	0	3	23	22	22	22	23	7	8
200	0	3	23	22	22	22	22	6	7
400	0	2	20	20	20	20	20	6	6
800	0	3	21	21	21	21	21	5	6
				% increase in CIW over BD					
50	0	1846	29	29	27	27	52	35	22
100	0	10	24	22	22	22	25	11	12
200	0	3	22	22	22	22	22	7	9
400	0	2	22	21	21	21	21	6	8
800	0	2	21	20	20	20	20	6	7
				Coverage out of 1000					
50	940	989	943	949	948	945	959	963	950
100	934	946	950	950	950	951	953	948	949
200	942	941	942	943	943	943	938	946	951
400	953	951	957	954	954	955	955	956	953
800	948	951	938	939	939	940	941	936	939

Table A2.3b. Results from Chapter II simulation 3 (LH)

N	BD	CORR	MISS	CAL	MCAL	WCAL	RCAL	$PSPP_{REML}$	$PSPP_{BAYES}$
				Bias x 1000					
50	-89	1055	-75	-47	-47	-46	-55	-51	-58
100	18	20	-9	10	10	9	1	9	6
200	-12	-14	-25	-15	-15	-14	-12	-30	-27
400	5	3	-19	-12	-12	-12	-10	-14	-17
800	4	1	-4	1	1	1	1	6	6
				% increase in RMSE over BD					
50	0	3680	42	45	45	42	42	45	42
100	0	2	35	36	36	35	35	44	39
200	0	2	32	32	32	32	31	36	31
400	0	2	30	31	31	30	30	31	28
800	0	1	30	31	31	30	31	27	26
				% increase in CIW over BD					
50	0	8539	38	43	42	37	56	67	38
100	0	31	33	36	36	33	37	60	31
200	0	2	32	34	34	32	33	55	29
400	0	1	33	33	33	33	33	42	28
800	0	1	32	33	33	32	32	35	28
				Coverage out of 1000					
50	950	989	940	942	940	935	958	966	934
100	948	960	948	951	951	947	954	968	940
200	939	935	934	935	935	930	935	950	928
400	942	938	948	949	949	948	952	960	948
800	945	939	945	944	944	944	944	956	955

Table A2.3c. Results from Chapter II simulation 3 (HL)

N	BD	CORR	MISS	CAL	MCAL	WCAL	RCAL	$PSPP_{REML}$	$PSPP_{BAYES}$
					Bias x 1000				
50	5	-2	-76	50	48	56	-86	14	22
100	-3	-2	-113	-7	-7	-4	-101	-20	-18
200	7	7	-96	5	5	6	-44	-3	-1
400	-9	-7	-99	-2	-2	-2	-32	-9	-7
800	2	3	-88	6	6	7	-9	2	3
					% increase in RMSE over BD				
50	0	34	33	28	26	26	61	28	24
100	0	6	28	20	20	21	30	8	9
200	0	2	28	19	19	19	19	4	5
400	0	4	36	23	23	23	22	8	9
800	0	3	38	22	22	22	19	7	7
					% increase in CIW over BD				
50	0	3484	35	34	29	28	106	57	27
100	0	49	28	23	22	22	39	16	16
200	0	5	26	21	21	20	26	8	12
400	0	3	24	19	19	19	20	7	9
800	0	3	24	19	19	19	18	7	8
					Coverage out of 1000				
50	948	987	957	957	955	954	980	973	967
100	948	962	942	945	943	943	957	949	953
200	943	950	939	940	940	934	948	952	951
400	949	947	924	946	946	947	937	947	951
800	947	947	919	944	944	947	949	945	950

Table A2.3d. Results from Chapter II simulation 3 (HH)

N	BD	CORR	MISS	CAL	MCAL	WCAL	RCAL	$PSPP_{REML}$	$PSPP_{BAYES}$
					Bias x 1000				
50	-12	57	-122	12	10	7	-60	23	12
100	-5	8	-74	31	31	29	-60	12	22
200	1	4	-91	15	15	16	-40	14	4
400	1	-3	-99	3	3	0	-39	-22	-18
800	-9	-7	-94	3	3	2	-16	3	0
					% increase in RMSE over BD				
50	0	798	42	55	54	44	66	65	53
100	0	9	39	50	50	41	48	69	55
200	0	2	38	49	49	39	36	67	53
400	0	2	44	51	51	44	43	59	54
800	0	0	41	45	45	39	38	53	49
					% increase in CIW over BD				
50	0	2079	44	55	50	42	107	101	46
100	0	68	38	48	46	38	53	91	40
200	0	3	36	45	45	37	40	93	37
400	0	2	37	46	46	38	38	74	34
800	0	1	36	45	45	38	34	59	33
					Coverage out of 1000				
50	954	974	955	952	948	949	979	971	947
100	939	955	938	941	939	938	953	958	928
200	951	953	944	947	946	950	949	967	926
400	958	956	945	947	947	948	951	964	920
800	951	961	952	965	965	966	953	959	937

Table A2.4a. Results from Chapter II simulation 4 (LL)

N	BD	CORR	MISS	CAL	MCAL	WCAL	RCAL	$PSPP_{REML}$	$PSPP_{BAYES}$
				Bias x 1000					
50	7	1	9	1	1	1	2	1	1
100	7	11	22	11	11	11	9	10	11
200	2	2	12	2	2	1	2	2	2
400	1	0	12	0	0	0	0	0	0
800	2	3	12	3	3	3	3	3	3
				% increase in RMSE over BD					
50	0	15	23	15	15	16	20	18	16
100	0	10	18	10	10	10	11	10	10
200	0	15	23	15	15	15	16	15	15
400	0	12	21	12	12	12	12	12	12
800	0	13	25	13	13	13	13	14	13
				% increase in CIW over BD					
50	0	22	26	22	22	22	66	58	24
100	0	15	23	15	15	15	23	24	18
200	0	14	22	14	14	14	16	17	16
400	0	14	22	14	14	14	14	15	15
800	0	13	21	14	14	13	14	14	14
				Coverage out of 1000					
50	938	947	942	949	949	948	971	964	951
100	940	948	943	949	949	949	955	959	954
200	945	931	925	932	932	934	930	937	939
400	948	958	948	958	958	957	955	959	962
800	951	950	943	950	950	951	951	950	950

Table A2.4b. Results from Chapter II simulation 4 (LH)

N	BD	CORR	MISS	CAL	MCAL	WCAL	RCAL	$PSPP_{REML}$	$PSPP_{BAYES}$
				Bias x 1000					
50	1	0	58	0	0	0	0	-2	1
100	2	6	56	6	6	6	5	7	6
200	15	15	66	15	15	15	15	15	14
400	4	3	47	3	3	3	3	4	3
800	4	5	54	5	5	5	5	5	5
				% increase in RMSE over BD					
50	0	3	45	3	3	3	5	5	3
100	0	4	49	4	4	4	5	5	5
200	0	3	43	3	3	3	3	3	3
400	0	3	44	3	3	3	3	3	3
800	0	4	44	4	4	4	4	4	4
				% increase in CIW over BD					
50	0	5	41	5	5	5	17	15	6
100	0	3	40	3	3	3	5	5	5
200	0	3	40	3	3	3	3	4	3
400	0	3	39	3	3	3	3	3	3
800	0	3	39	3	3	3	3	3	3
				Coverage out of 1000					
50	959	950	933	949	949	948	965	959	956
100	947	940	942	940	940	941	944	949	947
200	953	950	939	950	950	949	949	952	953
400	950	950	946	950	950	950	949	952	953
800	954	945	937	945	945	944	947	948	951

Table A2.4c. Results from Chapter II simulation 4 (HL)

N	BD	CORR	MISS	CAL	MCAL	WCAL	RCAL	$PSPP_{REML}$	$PSPP_{BAYES}$
				Bias x 1000					
50	-11	-8	27	-8	-8	-8	-2	-5	-7
100	-7	-5	41	-5	-5	-5	-4	-5	-5
200	2	5	48	5	5	5	5	4	5
400	-1	-1	42	-1	-1	-1	-1	-1	-1
800	-1	2	43	2	2	1	2	2	2
				% increase in RMSE over BD					
50	0	19	27	19	19	20	27	22	20
100	0	15	26	15	15	15	18	17	15
200	0	15	27	15	15	15	15	15	15
400	0	15	31	15	15	15	15	15	15
800	0	11	36	11	11	11	11	11	11
				% increase in CIW over BD					
50	0	25	27	25	25	25	68	59	26
100	0	17	24	17	17	17	25	26	19
200	0	15	23	15	15	15	17	18	17
400	0	15	24	15	15	15	16	16	16
800	0	15	23	15	15	15	15	16	15
				Coverage out of 1000					
50	934	941	939	941	941	940	962	962	949
100	948	948	948	948	948	948	968	972	960
200	945	947	944	947	947	947	949	947	950
400	942	950	932	948	948	947	951	949	954
800	951	951	929	951	951	951	951	951	953

Table A2.4d. Results from Chapter II simulation 4 (HH)

N	BD	CORR	MISS	CAL	MCAL	WCAL	RCAL	$PSPP_{REML}$	$PSPP_{BAYES}$
				Bias x 1000					
50	27	36	299	36	36	34	32	37	36
100	11	9	233	9	9	9	8	8	9
200	2	2	226	2	2	2	3	2	2
400	2	2	213	2	2	2	2	2	2
800	-4	-6	213	-6	-6	-6	-6	-6	-6
				% increase in RMSE over BD					
50	0	6	61	6	6	6	8	7	6
100	0	3	58	3	3	3	3	3	3
200	0	4	67	4	4	4	4	4	4
400	0	2	66	2	2	2	2	2	2
800	0	3	99	3	3	3	2	2	3
				% increase in CIW over BD					
50	0	6	45	6	6	6	19	16	7
100	0	4	43	4	4	4	5	6	5
200	0	3	42	3	3	3	4	4	4
400	0	3	43	3	3	3	3	3	4
800	0	3	43	3	3	3	3	3	4
				Coverage out of 1000					
50	957	944	928	944	944	944	964	959	955
100	952	956	932	955	955	955	954	956	951
200	949	948	927	948	948	948	948	948	959
400	939	940	908	939	939	939	940	938	937
800	951	951	830	952	952	952	951	950	948

Table A2.5. Results from Chapter II simulation 5 (CC)

N	BD	CORR	MISS	CAL	MCAL	WCAL	RCAL	$PSPP_{REML}$	$PSPP_{BAYES}$
				Bias x 1000					
50	-169	-168	-215	-177	-172	-170	-119	-160	-166
100	-24	-27	-141	-28	-27	-27	-32	-29	-26
200	54	59	-459	58	58	58	62	61	59
400	27	28	-762	28	28	28	28	28	28
800	-19	-16	-821	-14	-14	-15	-15	-14	-15
				% increase in RMSE over BD					
50	0	0	35	0	0	0	2	0	0
100	0	0	30	0	0	0	1	0	0
200	0	0	32	0	0	0	1	0	0
400	0	0	33	0	0	0	0	0	0
800	0	0	40	0	0	0	0	0	0
				% increase in CIW over BD					
50	0	0	37	0	0	0	7	1	1
100	0	0	31	0	0	0	1	0	1
200	0	0	29	0	0	0	0	0	1
400	0	0	28	0	0	0	0	0	1
800	0	0	29	0	0	0	0	0	0
				Coverage out of 1000					
50	942	939	937	939	938	938	947	941	946
100	937	934	937	936	936	936	936	936	943
200	955	953	946	955	955	955	954	953	954
400	944	945	926	945	945	945	945	945	948
800	940	938	920	939	939	939	941	940	941

Table A2.6. Results from Chapter II simulation 5 (MC)

N	BD	CORR	MISS	CAL	MCAL	WCAL	RCAL	$PSPP_{REML}$	$PSPP_{BAYES}$
				Bias x 1000					
50	-301	-305	-222	-308	-308	-307	-315	-318	-315
100	54	53	-313	46	51	52	50	56	51
200	-100	-101	-671	-98	-103	-103	-105	-104	-102
400	-35	-37	-733	-38	-39	-36	-37	-37	-36
800	10	10	-786	22	12	9	9	9	9
				% increase in RMSE over BD					
50	0	0	32	0	0	0	1	1	0
100	0	0	30	0	0	0	0	0	0
200	0	0	31	0	0	0	0	0	0
400	0	0	40	4	0	0	0	0	0
800	0	0	45	11	0	0	0	0	0
				% increase in CIW over BD					
50	0	0	35	1	0	0	1	2	2
100	0	0	31	3	0	0	0	1	1
200	0	0	29	9	0	0	0	0	1
400	0	0	29	5	0	0	0	0	1
800	0	0	29	10	0	0	0	0	0
				Coverage out of 1000					
50	953	954	954	956	954	953	952	955	958
100	939	939	946	939	939	939	938	941	944
200	950	949	952	950	949	949	949	951	960
400	959	960	937	959	959	959	961	960	965
800	952	950	921	952	947	949	950	949	951

Table A2.7. Results from Chapter II simulation 7 (CM)

N	BD	CORR	MISS	CAL	MCAL	WCAL	RCAL	$PSPP_{REML}$	$PSPP_{BAYES}$
				Bias x 1000					
50	-32	-28	-61	712	802	849	-658	278	703
100	-128	-126	-310	487	535	617	-568	120	266
200	-49	-52	-531	296	317	466	-279	48	124
400	46	46	-572	273	289	415	-5	82	177
800	-6	-6	-804	90	97	203	17	-3	24
				% increase in RMSE over BD					
50	0	0	32	41	33	28	28	37	26
100	0	0	38	45	41	28	18	40	21
200	0	0	31	42	39	20	6	24	9
400	0	0	38	50	47	20	2	18	10
800	0	0	48	53	51	20	1	16	8
				% increase in CIW over BD					
50	0	0	36	73	37	33	71	99	18
100	0	0	31	56	35	25	35	84	12
200	0	0	30	40	34	20	10	64	6
400	0	0	29	37	34	17	4	36	4
800	0	0	29	36	35	17	2	22	2
				Coverage out of 1000					
50	952	952	954	977	968	964	989	991	955
100	957	959	936	963	957	956	977	989	941
200	944	944	945	948	942	944	954	983	939
400	949	949	923	949	945	943	951	967	932
800	958	955	926	950	949	950	960	956	941

Table A2.8. Results from Chapter II simulation 8 (MM)

N	BD	CORR	MISS	CAL	MCAL	WCAL	RCAL	$PSPP_{REML}$	$PSPP_{BAYES}$
				Bias					
50	-215	-230	-23	-1624	-1236	-996	-1716	-2198	-1073
100	-14	-14	-356	-3613	-2740	-1696	-1391	-2828	-1859
200	-2	-4	-679	-5334	-4166	-2293	-1312	-2763	-2223
400	51	52	-659	-43887	-6085	-2528	-1263	-2268	-2160
800	-71	-73	-835	-27360	-7661	-2962	-1503	-2427	-2409
				% increase in RMSE over BD					
50	0	0	33	104	45	35	62	72	40
100	0	0	33	242	97	39	47	149	62
200	0	0	36	432	202	59	31	141	59
400	0	0	36	55744	554	90	32	103	73
800	0	0	44	34324	930	158	58	132	119
				% increase in CIW over BD					
50	0	0	36	93	44	38	112	160	37
100	0	0	31	322	57	30	66	190	32
200	0	0	29	470	84	26	40	183	23
400	0	0	28	10638	147	25	21	118	17
800	0	0	29	4107	241	26	14	91	14
				Coverage out of 1000					
50	950	955	948	962	947	941	978	978	935
100	950	949	950	973	951	933	960	974	928
200	948	946	931	949	907	863	949	978	857
400	946	945	934	911	844	785	915	936	797
800	944	945	906	756	657	541	824	830	605

Table A3.1a. Results from Chapter III scenario 1 under $\lambda_T = 0$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-250	290	397	645
	0	S-BNPM	1	161	625	61
		BNPM-ML	2	159	595	70
		BNPM-BAYES	0	159	607	64
	1	S-BNPM	269	334	678	323
		BNPM-ML	262	324	712	267
		BNPM-BAYES	260	322	722	335
	∞	S-BNPM	808	905	1706	691
		BNPM-ML	852	995	42061	265
		BNPM-BAYES	835	938	3390	863
400		CC	-249	259	196	984
	0	S-BNPM	-1	73	299	45
		BNPM-ML	0	72	289	52
		BNPM-BAYES	0	72	290	53
	1	S-BNPM	251	266	326	862
		BNPM-ML	248	262	337	853
		BNPM-BAYES	248	261	336	871
	∞	S-BNPM	750	769	620	1000
		BNPM-ML	755	773	707	1000
		BNPM-BAYES	755	773	701	1000

Table A3.1b. Results from Chapter III scenario 1 under $\lambda_T = 1$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-500	519	395	982
	0	S-BNPM	-250	291	628	356
		BNPM-ML	-250	290	596	400
		BNPM-BAYES	-252	292	607	370
	1	S-BNPM	20	183	682	54
		BNPM-ML	15	178	721	43
		BNPM-BAYES	13	178	728	52
	∞	S-BNPM	568	692	1671	238
		BNPM-ML	643	1401	46654	30
		BNPM-BAYES	592	723	3391	493
400		CC	-501	506	196	1000
	0	S-BNPM	-250	262	301	885
		BNPM-ML	-250	261	291	897
		BNPM-BAYES	-250	262	293	892
	1	S-BNPM	5	93	329	66
		BNPM-ML	3	91	339	51
		BNPM-BAYES	3	91	340	55
	∞	S-BNPM	513	545	634	946
		BNPM-ML	521	552	728	930
		BNPM-BAYES	521	552	720	977

Table A3.1c. Results from Chapter III scenario 1 under $\lambda_T = \infty$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-1001	1011	394	1000
	0	S-BNPM	-746	762	619	991
		BNPM-ML	-747	763	592	996
		BNPM-BAYES	-749	765	600	995
	1	S-BNPM	-486	519	671	771
		BNPM-ML	-494	524	705	753
		BNPM-BAYES	-496	526	712	691
	∞	S-BNPM	55	405	1623	70
		BNPM-ML	81	444	25807	35
		BNPM-BAYES	74	415	3014	34
400		CC	-996	998	197	1000
	0	S-BNPM	-745	748	301	1000
		BNPM-ML	-745	748	292	1000
		BNPM-BAYES	-745	749	292	1000
	1	S-BNPM	-492	500	328	998
		BNPM-ML	-494	502	339	999
		BNPM-BAYES	-495	502	339	998
	∞	S-BNPM	6	174	621	76
		BNPM-ML	12	174	710	45
		BNPM-BAYES	12	174	702	59

Table A3.2a. Results from Chapter III scenario 2 under $\lambda_T = 0$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-2401	2412	932	1000
		True	-37	503	2028	56
	0	S-BNPM	-35	553	2372	46
		BNPM-ML	-37	503	2006	65
		BNPM-BAYES	-63	504	2015	51
	1	S-BNPM	368	761	2572	39
		BNPM-ML	716	963	2546	121
		BNPM-BAYES	693	943	2497	197
	∞	S-BNPM	986	1370	3262	89
		BNPM-ML	1490	1744	3805	243
		BNPM-BAYES	1476	1730	3360	527
400		CC	-2399	2402	463	1000
		True	-1	245	989	47
	0	S-BNPM	1	276	1384	12
		BNPM-ML	-1	245	985	48
		BNPM-BAYES	-8	245	984	49
	1	S-BNPM	503	809	2446	8
		BNPM-ML	719	782	1206	640
		BNPM-BAYES	714	777	1202	695
	∞	S-BNPM	864	1066	2704	52
		BNPM-ML	1443	1499	1586	986
		BNPM-BAYES	1440	1496	1525	993

Table A3.2b. Results from Chapter III scenario 2 under $\lambda_T = 1$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-2082	2101	1107	1000
		True	27	472	2216	23
	0	S-BNPM	-123	499	1927	81
		BNPM-ML	-122	475	1788	91
		BNPM-BAYES	-141	478	1798	73
	1	S-BNPM	72	537	1994	61
		BNPM-ML	241	569	2031	47
		BNPM-BAYES	225	559	2035	62
	∞	S-BNPM	304	677	2196	53
		BNPM-ML	618	869	2498	78
		BNPM-BAYES	607	859	2389	151
400		CC	-2107	2112	545	1000
		True	3	236	1061	28
	0	S-BNPM	-147	279	985	101
		BNPM-ML	-142	269	876	123
		BNPM-BAYES	-147	272	878	120
	1	S-BNPM	52	292	1140	40
		BNPM-ML	206	328	977	115
		BNPM-BAYES	202	326	988	130
	∞	S-BNPM	183	367	1184	44
		BNPM-ML	564	637	1146	458
		BNPM-BAYES	561	634	1133	520

Table A3.2c. Results from Chapter III scenario 2 under $\lambda_T = \infty$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-1917	1940	1155	999
		True	40	463	2051	31
	0	S-BNPM	-181	488	1792	99
		BNPM-ML	-179	478	1702	107
		BNPM-BAYES	-194	481	1725	95
	1	S-BNPM	-47	477	1826	80
		BNPM-ML	71	487	1859	56
		BNPM-BAYES	59	482	1898	54
	∞	S-BNPM	94	527	1945	61
		BNPM-ML	332	641	2149	48
		BNPM-BAYES	321	632	2129	74
400		CC	-1911	1917	573	1000
		True	10	220	984	25
	0	S-BNPM	-201	296	899	153
		BNPM-ML	-195	288	837	173
		BNPM-BAYES	-198	291	837	165
	1	S-BNPM	-73	265	1013	80
		BNPM-ML	38	229	901	50
		BNPM-BAYES	34	229	916	45
	∞	S-BNPM	55	304	1124	44
		BNPM-ML	280	375	1002	150
		BNPM-BAYES	278	374	1012	171

Table A3.3a. Results from Chapter III scenario 3 under $\lambda_T = 0$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-1146	1191	939	981
	0	S-BNPM	5	291	1185	50
		BNPM-ML	6	288	1165	53
		BNPM-BAYES	2	288	1171	53
	1	S-BNPM	131	326	1200	63
		BNPM-ML	126	320	1203	60
		BNPM-BAYES	123	319	1201	63
	∞	S-BNPM	253	401	1245	104
		BNPM-ML	246	394	1250	97
		BNPM-BAYES	246	393	1255	117
400		CC	-1157	1168	467	1000
	0	S-BNPM	3	148	584	54
		BNPM-ML	3	147	576	52
		BNPM-BAYES	2	146	574	51
	1	S-BNPM	131	202	599	132
		BNPM-ML	126	196	593	124
		BNPM-BAYES	126	196	588	131
	∞	S-BNPM	256	300	617	345
		BNPM-ML	250	295	614	326
		BNPM-BAYES	250	295	612	345

Table A3.3b. Results from Chapter III scenario 3 under $\lambda_T = 1$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-1349	1393	914	995
	0	S-BNPM	-117	319	1197	65
		BNPM-ML	-116	317	1180	70
		BNPM-BAYES	-122	319	1178	70
	1	S-BNPM	26	309	1216	42
		BNPM-ML	18	303	1225	46
		BNPM-BAYES	14	302	1210	47
	∞	S-BNPM	163	366	1267	64
		BNPM-ML	151	353	1280	57
		BNPM-BAYES	151	353	1276	72
400		CC	-1351	1361	458	1000
	0	S-BNPM	-139	198	587	140
		BNPM-ML	-138	197	576	152
		BNPM-BAYES	-140	198	576	145
	1	S-BNPM	-2	147	601	41
		BNPM-ML	-5	145	594	42
		BNPM-BAYES	-5	145	590	46
	∞	S-BNPM	133	204	619	113
		BNPM-ML	129	201	617	116
		BNPM-BAYES	129	201	617	120

Table A3.3c. Results from Chapter III scenario 3 under $\lambda_T = \infty$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-1314	1357	914	995
	0	S-BNPM	-233	372	1139	133
		BNPM-ML	-231	370	1125	136
		BNPM-BAYES	-235	373	1129	132
	1	S-BNPM	-115	319	1148	87
		BNPM-ML	-118	316	1160	78
		BNPM-BAYES	-119	316	1156	81
	∞	S-BNPM	0	307	1193	51
		BNPM-ML	-5	304	1204	50
		BNPM-BAYES	-5	304	1207	48
400		CC	-1312	1322	452	1000
	0	S-BNPM	-240	277	560	387
		BNPM-ML	-238	275	552	392
		BNPM-BAYES	-239	276	553	395
	1	S-BNPM	-118	185	572	133
		BNPM-ML	-119	185	568	135
		BNPM-BAYES	-120	185	567	129
	∞	S-BNPM	1	149	590	47
		BNPM-ML	-1	148	589	44
		BNPM-BAYES	-1	147	588	42

Table A3.3d. Results from Chapter III scenario 3 when nonresponse depends on Z_2.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-1221	1265	930	988
	0	S-BNPM	7	309	1201	58
		BNPM-ML	6	307	1219	55
		BNPM-BAYES	3	306	1178	65
	1	S-BNPM	145	354	1220	68
		BNPM-ML	137	344	1269	53
		BNPM-BAYES	134	343	1214	68
	∞	S-BNPM	277	436	1267	123
		BNPM-ML	267	425	1327	89
		BNPM-BAYES	267	425	1273	134
400		CC	-1204	1216	461	1000
	0	S-BNPM	9	150	589	48
		BNPM-ML	8	149	597	44
		BNPM-BAYES	8	149	577	49
	1	S-BNPM	148	214	607	159
		BNPM-ML	141	208	619	137
		BNPM-BAYES	141	208	593	167
	∞	S-BNPM	280	325	626	408
		BNPM-ML	274	318	646	364
		BNPM-BAYES	274	318	617	421

Table A3.3e. Results from Chapter III scenario 3 when nonresponse depends on $2Z_2 + Y$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-1581	1610	840	1000
	0	S-BNPM	-239	375	1196	127
		BNPM-ML	-230	365	1160	127
		BNPM-BAYES	-235	368	1153	126
	1	S-BNPM	-41	318	1238	54
		BNPM-ML	-48	304	1231	55
		BNPM-BAYES	-51	303	1206	60
	∞	S-BNPM	149	375	1315	60
		BNPM-ML	134	355	1315	55
		BNPM-BAYES	134	355	1300	68
400		CC	-1576	1583	417	1000
	0	S-BNPM	-235	277	592	342
		BNPM-ML	-226	267	566	355
		BNPM-BAYES	-227	268	565	353
	1	S-BNPM	-40	161	625	56
		BNPM-ML	-42	155	598	57
		BNPM-BAYES	-42	155	589	54
	∞	S-BNPM	152	226	658	137
		BNPM-ML	142	214	636	133
		BNPM-BAYES	141	214	631	148

Table A3.4a. Results from Chapter III scenario 4 under $\lambda_T = 0$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-2146	2169	891	1000
	0	S-BNPM	5	479	1968	45
		BNPM-ML	2	467	1872	55
		BNPM-BAYES	-14	465	1845	53
	1	S-BNPM	220	556	2089	43
		BNPM-ML	271	568	2052	51
		BNPM-BAYES	256	559	2009	58
	∞	S-BNPM	445	720	2289	53
		BNPM-ML	540	775	2240	92
		BNPM-BAYES	528	765	2223	134
400		CC	-2155	2161	442	1000
	0	S-BNPM	3	231	999	36
		BNPM-ML	1	227	916	51
		BNPM-BAYES	-3	227	907	54
	1	S-BNPM	186	318	1123	51
		BNPM-ML	265	358	1000	146
		BNPM-BAYES	261	356	988	167
	∞	S-BNPM	358	463	1233	137
		BNPM-ML	528	592	1087	458
		BNPM-BAYES	526	589	1079	503

Table A3.4b. Results from Chapter III scenario 4 under $\lambda_T = 1$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-2208	2229	883	1000
	0	S-BNPM	-98	503	1938	62
		BNPM-ML	-93	494	1847	66
		BNPM-BAYES	-110	495	1823	67
	1	S-BNPM	116	546	2051	37
		BNPM-ML	166	548	2022	47
		BNPM-BAYES	153	542	1974	66
	∞	S-BNPM	334	679	2236	50
		BNPM-ML	426	717	2204	70
		BNPM-BAYES	415	709	2181	119
400		CC	-2201	2206	437	1000
	0	S-BNPM	-136	270	967	96
		BNPM-ML	-137	263	888	116
		BNPM-BAYES	-141	265	882	111
	1	S-BNPM	38	264	1079	37
		BNPM-ML	121	271	971	55
		BNPM-BAYES	117	269	959	69
	∞	S-BNPM	198	352	1185	61
		BNPM-ML	379	466	1057	267
		BNPM-BAYES	376	463	1050	309

Table A3.4c. Results from Chapter III scenario 4 under $\lambda_T = \infty$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-2279	2301	868	1000
	0	S-BNPM	-281	537	1869	115
		BNPM-ML	-275	524	1780	125
		BNPM-BAYES	-291	532	1760	106
	1	S-BNPM	-79	503	1983	60
		BNPM-ML	-24	481	1950	55
		BNPM-BAYES	-38	478	1913	55
	∞	S-BNPM	136	576	2181	51
		BNPM-ML	226	584	2127	47
		BNPM-BAYES	215	579	2115	66
400		CC	-2280	2285	431	1000
	0	S-BNPM	-255	336	950	186
		BNPM-ML	-257	333	875	221
		BNPM-BAYES	-261	336	868	204
	1	S-BNPM	-85	253	1058	59
		BNPM-ML	-5	226	955	45
		BNPM-BAYES	-9	226	942	42
	∞	S-BNPM	76	280	1167	32
		BNPM-ML	247	351	1039	109
		BNPM-BAYES	244	349	1031	132

Table A3.4d. Results from Chapter III scenario 4 when nonresponse depends on Z_2.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-1392	1453	1124	950
	0	S-BNPM	11	462	1788	59
		BNPM-ML	11	460	1771	60
		BNPM-BAYES	2	458	1777	57
	1	S-BNPM	98	485	1823	53
		BNPM-ML	115	486	1846	52
		BNPM-BAYES	107	483	1848	54
	∞	S-BNPM	174	523	1893	50
		BNPM-ML	219	537	1926	52
		BNPM-BAYES	211	532	1934	59
400		CC	-1406	1420	559	1000
	0	S-BNPM	-3	229	887	52
		BNPM-ML	-2	228	876	58
		BNPM-BAYES	-4	228	873	55
	1	S-BNPM	65	240	907	51
		BNPM-ML	98	253	911	61
		BNPM-BAYES	96	251	907	68
	∞	S-BNPM	120	266	934	67
		BNPM-ML	198	311	949	105
		BNPM-BAYES	197	310	945	117

Table A3.4e. Results from Chapter III scenario 4 when nonresponse depends on $2Z_2 + Y$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-1712	1748	1041	997
	0	S-BNPM	-123	455	1770	72
		BNPM-ML	-122	451	1731	73
		BNPM-BAYES	-131	452	1746	66
	1	S-BNPM	-18	451	1810	57
		BNPM-ML	13	450	1824	51
		BNPM-BAYES	4	448	1838	43
	∞	S-BNPM	83	478	1893	41
		BNPM-ML	148	496	1921	39
		BNPM-BAYES	142	493	1939	48
400		CC	-1714	1723	519	1000
	0	S-BNPM	-113	244	884	87
		BNPM-ML	-114	243	865	89
		BNPM-BAYES	-116	244	862	89
	1	S-BNPM	-23	224	918	46
		BNPM-ML	19	222	910	46
		BNPM-BAYES	17	222	905	48
	∞	S-BNPM	51	240	957	43
		BNPM-ML	151	276	957	69
		BNPM-BAYES	149	275	952	79

Table A3.5a. Results from Chapter III scenario 5 under $\lambda_T = 0$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-1304	1327	730	998
	0	S-BNPM	-13	374	1519	57
		BNPM-ML	-15	368	1502	50
		BNPM-BAYES	-28	367	1441	54
	1	S-BNPM	184	440	1645	39
		BNPM-ML	255	480	1742	22
		BNPM-BAYES	245	473	1646	61
	∞	S-BNPM	388	614	1886	42
		BNPM-ML	525	709	1996	86
		BNPM-BAYES	516	701	1913	186
400		CC	-1287	1294	368	1000
	0	S-BNPM	-2	188	763	46
		BNPM-ML	-1	185	718	47
		BNPM-BAYES	-4	185	712	52
	1	S-BNPM	153	259	864	58
		BNPM-ML	255	328	821	191
		BNPM-BAYES	252	326	805	231
	∞	S-BNPM	274	364	953	125
		BNPM-ML	511	565	930	597
		BNPM-BAYES	509	564	920	630

Table A3.5b. Results from Chapter III scenario 5 under $\lambda_T = 1$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-1521	1543	701	1000
	0	S-BNPM	-143	427	1616	101
		BNPM-ML	-149	416	1639	89
		BNPM-BAYES	-164	419	1492	105
	1	S-BNPM	106	460	1800	45
		BNPM-ML	166	469	1950	27
		BNPM-BAYES	154	461	1716	57
	∞	S-BNPM	361	652	2143	31
		BNPM-ML	481	721	2279	35
		BNPM-BAYES	473	714	2065	116
400		CC	-1510	1516	350	1000
	0	S-BNPM	-145	245	805	130
		BNPM-ML	-145	241	745	155
		BNPM-BAYES	-149	244	725	154
	1	S-BNPM	53	232	953	32
		BNPM-ML	163	274	869	85
		BNPM-BAYES	160	271	832	120
	∞	S-BNPM	216	342	1072	64
		BNPM-ML	472	543	1004	427
		BNPM-BAYES	470	541	976	492

Table A3.5c. Results from Chapter III scenario 5 under $\lambda_T = \infty$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-1444	1465	713	1000
	0	S-BNPM	-226	426	1470	129
		BNPM-ML	-226	417	1451	118
		BNPM-BAYES	-238	422	1398	121
	1	S-BNPM	-34	398	1603	60
		BNPM-ML	29	390	1683	38
		BNPM-BAYES	19	387	1594	42
	∞	S-BNPM	154	493	1833	29
		BNPM-ML	285	541	1930	24
		BNPM-BAYES	277	536	1858	71
400		CC	-1443	1449	356	1000
	0	S-BNPM	-233	295	731	261
		BNPM-ML	-233	292	688	277
		BNPM-BAYES	-236	295	679	277
	1	S-BNPM	-82	212	833	75
		BNPM-ML	17	194	789	50
		BNPM-BAYES	15	194	771	49
	∞	S-BNPM	43	229	929	32
		BNPM-ML	268	350	897	163
		BNPM-BAYES	265	348	885	215

Table A3.5d. Results from Chapter III scenario 5 when nonresponse depends on Z_2.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-728	822	1033	691
	0	S-BNPM	8	529	1413	196
		BNPM-ML	8	528	2141	54
		BNPM-BAYES	3	526	1409	194
	1	S-BNPM	74	555	1426	195
		BNPM-ML	83	561	2307	42
		BNPM-BAYES	78	558	1469	197
	∞	S-BNPM	118	582	1464	196
		BNPM-ML	158	610	2479	41
		BNPM-BAYES	154	606	1539	202
400		CC	-731	756	520	980
	0	S-BNPM	1	237	690	147
		BNPM-ML	1	237	938	53
		BNPM-BAYES	0	236	688	154
	1	S-BNPM	63	252	698	161
		BNPM-ML	65	256	994	42
		BNPM-BAYES	65	256	709	165
	∞	S-BNPM	93	269	701	181
		BNPM-ML	129	291	1052	49
		BNPM-BAYES	129	291	736	199

Table A3.5e. Results from Chapter III scenario 5 when nonresponse depends on $4Z_2 + Y$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-1217	1261	887	971
	0	S-BNPM	-279	518	1357	233
		BNPM-ML	-279	517	1728	138
		BNPM-BAYES	-287	519	1347	231
	1	S-BNPM	-174	482	1389	188
		BNPM-ML	-156	485	1901	82
		BNPM-BAYES	-161	484	1436	167
	∞	S-BNPM	-99	484	1451	158
		BNPM-ML	-32	500	2081	59
		BNPM-BAYES	-36	499	1551	126
400		CC	-1207	1217	442	1000
	0	S-BNPM	-283	349	661	439
		BNPM-ML	-283	350	789	322
		BNPM-BAYES	-284	351	654	434
	1	S-BNPM	-193	285	676	282
		BNPM-ML	-172	277	852	142
		BNPM-BAYES	-174	278	692	235
	∞	S-BNPM	-143	261	690	214
		BNPM-ML	-62	243	920	69
		BNPM-BAYES	-63	243	737	146

Table A3.6a. Results from Chapter III scenario 6 under $\lambda_T = 0$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-1220	1250	790	999
	0	S-BNPM	2	311	1251	54
		BNPM-ML	2	306	1209	58
		BNPM-BAYES	-4	304	1185	62
	1	S-BNPM	358	547	1517	72
		BNPM-ML	210	405	1328	65
		BNPM-BAYES	206	401	1281	97
	∞	S-BNPM	688	871	1923	164
		BNPM-ML	418	584	1460	160
		BNPM-BAYES	416	581	1437	241
400		CC	-1217	1225	395	1000
	0	S-BNPM	5	155	644	34
		BNPM-ML	4	150	593	45
		BNPM-BAYES	2	150	584	54
	1	S-BNPM	330	394	855	279
		BNPM-ML	207	268	644	238
		BNPM-BAYES	206	268	627	278
	∞	S-BNPM	671	742	1226	610
		BNPM-ML	409	456	701	602
		BNPM-BAYES	409	455	690	634

Table A3.6b. Results from Chapter III scenario 6 under $\lambda_T = 1$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-1056	1096	844	989
	0	S-BNPM	-137	310	1150	86
		BNPM-ML	-134	307	1133	85
		BNPM-BAYES	-139	308	1136	77
	1	S-BNPM	20	303	1217	41
		BNPM-ML	-6	298	1200	47
		BNPM-BAYES	-9	297	1192	43
	∞	S-BNPM	179	397	1375	42
		BNPM-ML	121	350	1278	49
		BNPM-BAYES	120	349	1279	70
400		CC	-1048	1059	423	1000
	0	S-BNPM	-128	191	574	154
		BNPM-ML	-127	189	560	151
		BNPM-BAYES	-128	189	559	154
	1	S-BNPM	7	156	613	46
		BNPM-ML	-2	153	588	53
		BNPM-BAYES	-3	153	584	61
	∞	S-BNPM	131	223	683	82
		BNPM-ML	124	210	622	123
		BNPM-BAYES	123	209	621	135

Table A3.6c. Results from Chapter III scenario 6 under $\lambda_T = \infty$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-1118	1156	835	992
	0	S-BNPM	-219	356	1140	145
		BNPM-ML	-215	353	1121	150
		BNPM-BAYES	-220	354	1124	139
	1	S-BNPM	-65	315	1202	69
		BNPM-ML	-90	314	1188	77
		BNPM-BAYES	-93	314	1182	78
	∞	S-BNPM	95	378	1369	44
		BNPM-ML	35	335	1268	54
		BNPM-BAYES	33	334	1270	56
400		CC	-1134	1145	418	1000
	0	S-BNPM	-238	279	565	404
		BNPM-ML	-235	276	551	423
		BNPM-BAYES	-237	277	550	407
	1	S-BNPM	-109	188	600	125
		BNPM-ML	-115	192	579	147
		BNPM-BAYES	-116	192	575	146
	∞	S-BNPM	8	171	658	49
		BNPM-ML	6	167	612	55
		BNPM-BAYES	5	167	610	56

Table A3.6d. Results from Chapter III scenario 6 when nonresponse depends on Z_2.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-745	789	726	937
	0	S-BNPM	-11	332	1219	61
		BNPM-ML	-12	328	1352	40
		BNPM-BAYES	-19	327	1173	81
	1	S-BNPM	221	484	1444	81
		BNPM-ML	149	408	1566	45
		BNPM-BAYES	144	405	1320	112
	∞	S-BNPM	452	713	1812	135
		BNPM-ML	310	545	1799	86
		BNPM-BAYES	306	542	1516	181
400		CC	-746	758	362	1000
	0	S-BNPM	3	164	604	77
		BNPM-ML	2	161	644	59
		BNPM-BAYES	0	161	571	91
	1	S-BNPM	153	254	708	131
		BNPM-ML	162	247	733	120
		BNPM-BAYES	160	246	635	195
	∞	S-BNPM	310	410	841	265
		BNPM-ML	321	389	832	305
		BNPM-BAYES	320	388	715	448

Table A3.6e. Results from Chapter III scenario 6 when nonresponse depends on $5Z_2 + Y$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-1043	1075	750	996
	0	S-BNPM	-220	379	1178	157
		BNPM-ML	-211	371	1240	119
		BNPM-BAYES	-217	372	1138	148
	1	S-BNPM	41	424	1396	70
		BNPM-ML	-43	361	1406	64
		BNPM-BAYES	-48	359	1253	86
	∞	S-BNPM	313	632	1773	57
		BNPM-ML	125	446	1591	51
		BNPM-BAYES	121	443	1422	104
400		CC	-1040	1048	377	1000
	0	S-BNPM	-220	266	583	345
		BNPM-ML	-210	257	595	303
		BNPM-BAYES	-212	258	555	334
	1	S-BNPM	-49	184	672	78
		BNPM-ML	-59	177	658	73
		BNPM-BAYES	-60	178	600	101
	∞	S-BNPM	127	280	806	68
		BNPM-ML	92	213	729	79
		BNPM-BAYES	91	213	663	120

Table A4.1a. Results from Chapter IV scenario 1 when missingness depends on *X*.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-108	129	281	323
		LOGREG	-7	62	253	55
	0	Bin-PPMA (BA)	1	61	233	65
		Bin-PPMA (MI)	0	62	243	63
		BinS-PPMA	0	62	244	62
	1	Bin-PPMA (BA)	53	77	214	158
		Bin-PPMA (MI)	54	78	222	177
		BinS-PPMA	54	79	220	180
	∞	Bin-PPMA (BA)	80	96	208	281
		Bin-PPMA (MI)	81	96	217	325
		BinS-PPMA	82	97	206	371
400		CC	-111	117	139	879
		LOGREG	-2	31	124	67
	0	Bin-PPMA (BA)	0	31	119	64
		Bin-PPMA (MI)	0	31	121	60
		BinS-PPMA	0	31	121	60
	1	Bin-PPMA (BA)	51	58	108	430
		Bin-PPMA (MI)	52	59	110	456
		BinS-PPMA	52	59	110	452
	∞	Bin-PPMA (BA)	85	89	107	854
		Bin-PPMA (MI)	86	90	109	871
		BinS-PPMA	86	90	103	896

Table A4.1b. Results from Chapter IV scenario 1 when missingness depends on (*X* + *Y*).

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-90	115	282	228
		LOGREG	-35	71	256	68
	0	Bin-PPMA (BA)	-30	69	237	84
		Bin-PPMA (MI)	-30	69	248	75
		BinS-PPMA	-30	70	249	77
	1	Bin-PPMA (BA)	3	61	239	55
		Bin-PPMA (MI)	5	62	249	59
		BinS-PPMA	5	63	244	61
	∞	Bin-PPMA (BA)	29	69	251	67
		Bin-PPMA (MI)	30	69	262	94
		BinS-PPMA	31	70	240	128
400		CC	-95	102	139	759
		LOGREG	-34	47	126	181
	0	Bin-PPMA (BA)	-32	45	121	196
		Bin-PPMA (MI)	-32	45	124	176
		BinS-PPMA	-32	46	124	174
	1	Bin-PPMA (BA)	0	31	120	55
		Bin-PPMA (MI)	0	31	122	52
		BinS-PPMA	0	32	122	53
	∞	Bin-PPMA (BA)	29	44	134	140
		Bin-PPMA (MI)	31	45	135	146
		BinS-PPMA	31	46	122	197

Table A4.1c. Results from Chapter IV scenario 1 when missingness depends on *Y*.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-121	140	282	391
		LOGREG	-65	92	257	169
	0	Bin-PPMA (BA)	-62	88	239	179
		Bin-PPMA (MI)	-61	89	251	146
		BinS-PPMA	-61	89	251	148
	1	Bin-PPMA (BA)	-29	70	243	74
		Bin-PPMA (MI)	-28	70	253	73
		BinS-PPMA	-28	71	254	74
	∞	Bin-PPMA (BA)	-3	66	260	53
		Bin-PPMA (MI)	-1	67	271	55
		BinS-PPMA	-1	67	256	65
400		CC	-122	127	139	939
		LOGREG	-61	69	127	465
	0	Bin-PPMA (BA)	-60	67	123	487
		Bin-PPMA (MI)	-60	67	125	468
		BinS-PPMA	-60	68	125	460
	1	Bin-PPMA (BA)	-28	41	122	133
		Bin-PPMA (MI)	-27	41	124	118
		BinS-PPMA	-27	41	127	113
	∞	Bin-PPMA (BA)	3	34	137	49
		Bin-PPMA (MI)	4	34	138	47
		BinS-PPMA	4	34	130	62

Table A4.1d. Results from Chapter IV scenario 1 when missingness depends on X^2.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	22	70	267	71
		LOGREG	-3	58	243	36
	0	Bin-PPMA (BA)	2	58	228	42
		Bin-PPMA (MI)	0	58	237	47
		BinS-PPMA	0	58	241	42
	1	Bin-PPMA (BA)	-19	57	212	61
		Bin-PPMA (MI)	-23	58	225	50
		BinS-PPMA	-25	60	230	48
	∞	Bin-PPMA (BA)	-29	60	205	89
		Bin-PPMA (MI)	-32	61	226	58
		BinS-PPMA	-33	62	239	47
400		CC	22	39	132	108
		LOGREG	-1	29	119	42
	0	Bin-PPMA (BA)	3	29	116	41
		Bin-PPMA (MI)	0	29	118	38
		BinS-PPMA	0	29	120	36
	1	Bin-PPMA (BA)	-19	32	107	91
		Bin-PPMA (MI)	-23	35	111	99
		BinS-PPMA	-27	38	112	124
	∞	Bin-PPMA (BA)	-30	39	102	200
		Bin-PPMA (MI)	-34	42	110	187
		BinS-PPMA	-36	44	118	172

Table A4.1e. Results from Chapter IV scenario 1 when missingness depends on $(X + Y)^2$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-16	71	276	52
		LOGREG	-5	62	252	52
	0	Bin-PPMA (BA)	-3	62	238	52
		Bin-PPMA (MI)	-1	62	246	55
		BinS-PPMA	-1	63	249	52
	1	Bin-PPMA (BA)	-2	58	230	49
		Bin-PPMA (MI)	-1	58	242	41
		BinS-PPMA	0	58	243	51
	∞	Bin-PPMA (BA)	-6	57	233	46
		Bin-PPMA (MI)	-4	57	252	32
		BinS-PPMA	-2	58	253	38
400		CC	-16	40	137	80
		LOGREG	-1	33	124	66
	0	Bin-PPMA (BA)	1	33	122	70
		Bin-PPMA (MI)	1	33	122	75
		BinS-PPMA	1	33	124	73
	1	Bin-PPMA (BA)	0	31	113	70
		Bin-PPMA (MI)	1	31	116	61
		BinS-PPMA	2	31	117	65
	∞	Bin-PPMA (BA)	-5	30	110	80
		Bin-PPMA (MI)	-3	30	117	58
		BinS-PPMA	-2	30	123	48

Table A4.1f. Results from Chapter IV scenario 1 when missingness depends on Y^2.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-42	81	271	86
		LOGREG	-21	65	247	64
	0	Bin-PPMA (BA)	-19	64	232	67
		Bin-PPMA (MI)	-17	63	241	54
		BinS-PPMA	-17	64	242	57
	1	Bin-PPMA (BA)	-7	59	228	56
		Bin-PPMA (MI)	-4	59	239	50
		BinS-PPMA	-3	59	242	49
	∞	Bin-PPMA (BA)	-3	59	236	49
		Bin-PPMA (MI)	0	59	251	40
		BinS-PPMA	1	59	248	43
400		CC	-39	52	134	205
		LOGREG	-14	35	121	74
	0	Bin-PPMA (BA)	-14	34	118	81
		Bin-PPMA (MI)	-12	34	119	63
		BinS-PPMA	-13	34	120	67
	1	Bin-PPMA (BA)	-4	31	114	58
		Bin-PPMA (MI)	-2	30	116	52
		BinS-PPMA	-1	30	119	48
	∞	Bin-PPMA (BA)	-2	30	114	55
		Bin-PPMA (MI)	0	30	120	51
		BinS-PPMA	2	30	120	53

Table A4.2a. Results from Chapter IV scenario 2 when missingness depends on *X*.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-50	66	178	259
		LOGREG	9	51	210	34
	0	Bin-PPMA (BA)	24	70	217	108
		Bin-PPMA (MI)	1	51	196	64
		BinS-PPMA	0	51	197	63
	1	Bin-PPMA (BA)	81	105	233	344
		Bin-PPMA (MI)	49	74	230	113
		BinS-PPMA	33	63	218	78
	∞	Bin-PPMA (BA)	110	131	242	499
		Bin-PPMA (MI)	77	99	253	206
		BinS-PPMA	56	80	231	125
400		CC	-49	54	89	574
		LOGREG	3	26	103	56
	0	Bin-PPMA (BA)	20	39	117	121
		Bin-PPMA (MI)	1	26	100	46
		BinS-PPMA	1	26	100	49
	1	Bin-PPMA (BA)	81	87	115	787
		Bin-PPMA (MI)	48	56	110	396
		BinS-PPMA	28	40	105	181
	∞	Bin-PPMA (BA)	120	124	116	968
		Bin-PPMA (MI)	84	90	124	770
		BinS-PPMA	51	58	111	413

Table A4.2b. Results from Chapter IV scenario 2 when missingness depends on (*X* + *Y*).

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-69	81	165	423
		LOGREG	-13	51	203	79
	0	Bin-PPMA (BA)	-4	66	204	116
		Bin-PPMA (MI)	-21	55	186	138
		BinS-PPMA	-22	55	186	137
	1	Bin-PPMA (BA)	53	89	229	236
		Bin-PPMA (MI)	26	64	227	91
		BinS-PPMA	10	56	212	83
	∞	Bin-PPMA (BA)	83	114	240	393
		Bin-PPMA (MI)	54	86	252	150
		BinS-PPMA	33	69	226	112
400		CC	-68	71	83	831
		LOGREG	-18	31	98	131
	0	Bin-PPMA (BA)	-7	33	112	84
		Bin-PPMA (MI)	-21	32	96	161
		BinS-PPMA	-21	33	96	158
	1	Bin-PPMA (BA)	57	66	116	527
		Bin-PPMA (MI)	28	39	109	145
		BinS-PPMA	6	27	102	56
	∞	Bin-PPMA (BA)	99	105	121	867
		Bin-PPMA (MI)	66	73	126	536
		BinS-PPMA	31	41	111	155

Table A4.2c. Results from Chapter IV scenario 2 when missingness depends on *Y*.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-81	91	153	554
		LOGREG	-29	54	194	122
	0	Bin-PPMA (BA)	-27	67	189	140
		Bin-PPMA (MI)	-38	61	175	210
		BinS-PPMA	-39	61	175	219
	1	Bin-PPMA (BA)	27	75	224	159
		Bin-PPMA (MI)	7	56	222	71
		BinS-PPMA	-9	52	205	93
	∞	Bin-PPMA (BA)	57	97	238	263
		Bin-PPMA (MI)	34	73	248	96
		BinS-PPMA	13	59	219	77
400		CC	-81	84	77	952
		LOGREG	-37	44	93	381
	0	Bin-PPMA (BA)	-29	43	105	216
		Bin-PPMA (MI)	-39	46	90	418
		BinS-PPMA	-40	46	90	421
	1	Bin-PPMA (BA)	32	47	114	267
		Bin-PPMA (MI)	7	28	105	64
		BinS-PPMA	-15	30	97	127
	∞	Bin-PPMA (BA)	74	83	121	689
		Bin-PPMA (MI)	44	55	124	274
		BinS-PPMA	10	30	106	60

Table A4.2d. Results from Chapter IV scenario 2 when missingness depends on X^2.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-91	98	147	649
		LOGREG	10	58	248	63
	0	Bin-PPMA (BA)	18	75	243	102
		Bin-PPMA (MI)	0	62	232	92
		BinS-PPMA	-2	62	235	92
	1	Bin-PPMA (BA)	132	151	266	499
		Bin-PPMA (MI)	100	120	287	315
		BinS-PPMA	77	99	278	188
	∞	Bin-PPMA (BA)	162	179	272	567
		Bin-PPMA (MI)	131	148	301	472
		BinS-PPMA	109	128	293	331
400		CC	-93	95	73	989
		LOGREG	4	34	129	70
	0	Bin-PPMA (BA)	13	43	145	88
		Bin-PPMA (MI)	-2	34	128	64
		BinS-PPMA	-4	34	136	52
	1	Bin-PPMA (BA)	156	159	122	988
		Bin-PPMA (MI)	123	127	131	985
		BinS-PPMA	84	90	135	694
	∞	Bin-PPMA (BA)	195	197	114	982
		Bin-PPMA (MI)	165	168	134	996
		BinS-PPMA	131	135	156	977

Table A4.2e. Results from Chapter IV scenario 2 when missingness depends on $(X + Y)^2$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-49	66	177	263
		LOGREG	-20	48	192	81
	0	Bin-PPMA (BA)	-20	55	188	75
		Bin-PPMA (MI)	-28	53	179	130
		BinS-PPMA	-29	53	182	135
	1	Bin-PPMA (BA)	57	83	224	223
		Bin-PPMA (MI)	31	59	230	46
		BinS-PPMA	13	48	225	29
	∞	Bin-PPMA (BA)	82	102	222	389
		Bin-PPMA (MI)	52	73	241	95
		BinS-PPMA	30	58	244	22
400		CC	-49	54	88	572
		LOGREG	-27	36	93	245
	0	Bin-PPMA (BA)	-26	37	98	185
		Bin-PPMA (MI)	-31	39	90	294
		BinS-PPMA	-32	40	94	297
	1	Bin-PPMA (BA)	64	72	114	613
		Bin-PPMA (MI)	29	40	108	167
		BinS-PPMA	4	24	102	35
	∞	Bin-PPMA (BA)	97	102	108	922
		Bin-PPMA (MI)	55	62	115	455
		BinS-PPMA	26	37	125	44

Table A4.2f. Results from Chapter IV scenario 2 when missingness depends on Y^2.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	74	97	243	183
		LOGREG	48	70	213	89
	0	Bin-PPMA (BA)	61	85	208	247
		Bin-PPMA (MI)	40	64	201	70
		BinS-PPMA	39	64	202	70
	1	Bin-PPMA (BA)	50	83	207	225
		Bin-PPMA (MI)	25	57	199	69
		BinS-PPMA	14	48	185	50
	∞	Bin-PPMA (BA)	49	86	214	244
		Bin-PPMA (MI)	23	59	212	68
		BinS-PPMA	12	49	188	50
400		CC	73	79	120	673
		LOGREG	38	45	102	262
	0	Bin-PPMA (BA)	53	60	106	546
		Bin-PPMA (MI)	34	42	99	230
		BinS-PPMA	34	42	100	225
	1	Bin-PPMA (BA)	41	54	103	402
		Bin-PPMA (MI)	19	32	93	132
		BinS-PPMA	7	23	85	49
	∞	Bin-PPMA (BA)	37	57	109	378
		Bin-PPMA (MI)	15	34	97	157
		BinS-PPMA	3	24	86	74

Table A4.3a. Results from Chapter IV scenario 3 when missingness depends on *X*.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-108	119	191	585
		LOGREG	0	57	237	52
		LATENT	1	63	222	88
	0	Bin-PPMA (BA)	44	87	258	130
		Bin-PPMA (MI)	-2	62	232	68
		BinS-PPMA	-5	62	236	65
	1	Bin-PPMA (BA)	117	129	218	574
		Bin-PPMA (MI)	64	85	234	171
		BinS-PPMA	45	74	237	110
	∞	Bin-PPMA (BA)	136	146	215	711
		Bin-PPMA (MI)	89	105	243	284
		BinS-PPMA	63	85	249	136
400		CC	-109	112	95	983
		LOGREG	-1	29	114	58
		LATENT	-2	29	112	61
	0	Bin-PPMA (BA)	43	55	133	251
		Bin-PPMA (MI)	-2	29	114	54
		BinS-PPMA	-4	30	117	50
	1	Bin-PPMA (BA)	113	116	102	995
		Bin-PPMA (MI)	61	67	110	610
		BinS-PPMA	39	48	112	260
	∞	Bin-PPMA (BA)	135	137	97	999
		Bin-PPMA (MI)	91	95	114	900
		BinS-PPMA	53	59	121	370

Table A4.3b. Results from Chapter IV scenario 3 when missingness depends on (*X* + *Y*).

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-140	146	174	823
		LOGREG	-41	68	232	116
		LATENT	4	70	261	77
	0	Bin-PPMA (BA)	-3	75	253	88
		Bin-PPMA (MI)	-45	75	222	159
		BinS-PPMA	-47	76	225	160
	1	Bin-PPMA (BA)	80	101	225	322
		Bin-PPMA (MI)	27	64	241	57
		BinS-PPMA	3	58	240	54
	∞	Bin-PPMA (BA)	103	120	224	485
		Bin-PPMA (MI)	54	81	253	114
		BinS-PPMA	25	63	252	50
400		CC	-140	142	87	1000
		LOGREG	-40	49	113	290
		LATENT	4	30	128	51
	0	Bin-PPMA (BA)	1	38	136	72
		Bin-PPMA (MI)	-41	51	112	320
		BinS-PPMA	-43	52	115	314
	1	Bin-PPMA (BA)	87	90	104	926
		Bin-PPMA (MI)	31	41	112	177
		BinS-PPMA	1	29	112	55
	∞	Bin-PPMA (BA)	114	116	99	994
		Bin-PPMA (MI)	66	72	118	608
		BinS-PPMA	22	35	123	58

Table A4.3c. Results from Chapter IV scenario 3 when missingness depends on *Y*.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-169	174	154	939
		LOGREG	-70	88	228	241
		LATENT	-11	80	282	120
	0	Bin-PPMA (BA)	-38	85	245	146
		Bin-PPMA (MI)	-78	97	212	334
		BinS-PPMA	-80	99	216	335
	1	Bin-PPMA (BA)	48	85	233	184
		Bin-PPMA (MI)	-7	65	244	95
		BinS-PPMA	-33	70	237	134
	∞	Bin-PPMA (BA)	74	103	232	305
		Bin-PPMA (MI)	21	72	258	90
		BinS-PPMA	-8	65	250	89
400		CC	-171	172	77	1000
		LOGREG	-75	80	110	731
		LATENT	-1	36	153	47
	0	Bin-PPMA (BA)	-40	55	138	219
		Bin-PPMA (MI)	-78	83	109	775
		BinS-PPMA	-79	84	113	771
	1	Bin-PPMA (BA)	60	66	109	614
		Bin-PPMA (MI)	0	29	114	54
		BinS-PPMA	-37	47	112	267
	∞	Bin-PPMA (BA)	94	98	104	934
		Bin-PPMA (MI)	40	50	124	230
		BinS-PPMA	-9	29	123	62

Table A4.3d. Results from Chapter IV scenario 3 when missingness depends on X^2.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-150	156	167	880
		LOGREG	-12	65	288	38
		LATENT	-17	85	264	101
	0	Bin-PPMA (BA)	14	90	295	94
		Bin-PPMA (MI)	-21	81	283	107
		BinS-PPMA	-26	81	300	93
	1	Bin-PPMA (BA)	121	144	272	463
		Bin-PPMA (MI)	84	112	309	287
		BinS-PPMA	61	94	310	156
	∞	Bin-PPMA (BA)	134	155	282	443
		Bin-PPMA (MI)	102	126	324	360
		BinS-PPMA	82	110	334	207
400		CC	-150	151	84	1000
		LOGREG	-4	38	148	54
		LATENT	-7	41	150	68
	0	Bin-PPMA (BA)	26	54	173	127
		Bin-PPMA (MI)	-10	42	152	57
		BinS-PPMA	-16	44	184	35
	1	Bin-PPMA (BA)	147	149	107	963
		Bin-PPMA (MI)	115	119	124	971
		BinS-PPMA	80	85	147	613
	∞	Bin-PPMA (BA)	162	164	108	941
		Bin-PPMA (MI)	139	142	128	973
		BinS-PPMA	109	113	183	759

Table A4.3e. Results from Chapter IV scenario 3 when missingness depends on $(X + Y)^2$.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-89	103	203	417
		LOGREG	-25	61	225	89
		LATENT	18	70	307	40
	0	Bin-PPMA (BA)	2	71	244	79
		Bin-PPMA (MI)	-30	65	220	105
		BinS-PPMA	-34	66	228	104
	1	Bin-PPMA (BA)	95	110	223	391
		Bin-PPMA (MI)	36	62	229	52
		BinS-PPMA	9	50	236	27
	∞	Bin-PPMA (BA)	113	124	222	484
		Bin-PPMA (MI)	55	74	236	89
		BinS-PPMA	26	56	263	20
400		CC	-88	91	101	903
		LOGREG	-24	37	111	152
		LATENT	18	36	150	48
	0	Bin-PPMA (BA)	3	36	130	64
		Bin-PPMA (MI)	-26	39	110	167
		BinS-PPMA	-30	42	122	158
	1	Bin-PPMA (BA)	101	104	102	970
		Bin-PPMA (MI)	38	46	109	242
		BinS-PPMA	3	25	110	29
	∞	Bin-PPMA (BA)	122	124	97	993
		Bin-PPMA (MI)	62	67	113	592
		BinS-PPMA	24	35	139	23

Table A4.3f. Results from Chapter IV scenario 3 when missingness depends on Y^2.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	10	62	244	53
		LOGREG	24	60	221	57
		LATENT	45	81	294	92
	0	Bin-PPMA (BA)	81	106	231	293
		Bin-PPMA (MI)	21	59	213	59
		BinS-PPMA	20	59	215	59
	1	Bin-PPMA (BA)	107	124	215	536
		Bin-PPMA (MI)	32	63	214	70
		BinS-PPMA	12	53	210	48
	∞	Bin-PPMA (BA)	119	134	209	652
		Bin-PPMA (MI)	43	70	223	86
		BinS-PPMA	15	54	220	44
400		CC	12	33	121	55
		LOGREG	23	36	108	123
		LATENT	38	49	143	189
	0	Bin-PPMA (BA)	80	87	120	730
		Bin-PPMA (MI)	22	35	106	117
		BinS-PPMA	21	35	107	107
	1	Bin-PPMA (BA)	108	112	108	954
		Bin-PPMA (MI)	31	41	105	196
		BinS-PPMA	9	27	104	53
	∞	Bin-PPMA (BA)	123	126	103	984
		Bin-PPMA (MI)	45	51	110	352
		BinS-PPMA	9	26	107	40

Table A4.4a. Results from Chapter IV scenario 4 when missingness depends on *X*.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-105	126	276	328
		LOGREG	-6	69	281	50
		LATENT	0	71	280	68
	0	Bin-PPMA (BA)	6	85	273	79
		Bin-PPMA (MI)	-3	71	271	68
		BinS-PPMA	-3	71	271	68
	1	Bin-PPMA (BA)	94	119	264	256
		Bin-PPMA (MI)	80	102	267	227
		BinS-PPMA	81	102	233	305
	∞	Bin-PPMA (BA)	138	153	260	509
		Bin-PPMA (MI)	121	136	274	437
		BinS-PPMA	114	130	236	491
400		CC	-104	110	136	848
		LOGREG	-1	34	139	48
		LATENT	0	33	145	34
	0	Bin-PPMA (BA)	0	35	140	47
		Bin-PPMA (MI)	-1	33	136	43
		BinS-PPMA	-1	33	137	46
	1	Bin-PPMA (BA)	88	94	131	728
		Bin-PPMA (MI)	81	87	128	704
		BinS-PPMA	87	92	113	849
	∞	Bin-PPMA (BA)	147	151	135	982
		Bin-PPMA (MI)	137	142	140	968
		BinS-PPMA	129	134	121	971

Table A4.4b. Results from Chapter IV scenario 4 when missingness depends on (*X* + *Y*).

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-84	108	284	202
		LOGREG	-37	73	263	78
		LATENT	-6	66	313	30
	0	Bin-PPMA (BA)	-38	82	261	98
		Bin-PPMA (MI)	-37	74	259	84
		BinS-PPMA	-37	74	259	84
	1	Bin-PPMA (BA)	-8	76	277	56
		Bin-PPMA (MI)	-9	66	274	46
		BinS-PPMA	-2	61	235	52
	∞	Bin-PPMA (BA)	18	84	318	68
		Bin-PPMA (MI)	16	75	313	46
		BinS-PPMA	15	72	256	80
400		CC	-81	88	140	623
		LOGREG	-33	46	132	153
		LATENT	-6	33	158	30
	0	Bin-PPMA (BA)	-36	49	133	168
		Bin-PPMA (MI)	-33	46	129	151
		BinS-PPMA	-33	46	130	149
	1	Bin-PPMA (BA)	-7	35	138	59
		Bin-PPMA (MI)	-7	33	133	50
		BinS-PPMA	5	30	116	60
	∞	Bin-PPMA (BA)	21	44	159	91
		Bin-PPMA (MI)	20	42	153	65
		BinS-PPMA	21	41	127	123

Table A4.4c. Results from Chapter IV scenario 4 when missingness depends on *Y*.

n	λ_A	Model	Bias (x1000)	RMSE (x1000)	CIW (x1000)	Non-Coverage
100		CC	-98	120	276	274
		LOGREG	-66	93	261	173
		LATENT	-18	84	389	27
	0	Bin-PPMA (BA)	-73	102	255	195
		Bin-PPMA (MI)	-68	95	258	188
		BinS-PPMA	-68	95	258	186
	1	Bin-PPMA (BA)	-49	91	282	104
		Bin-PPMA (MI)	-44	84	283	92
		BinS-PPMA	-36	77	245	112
	∞	Bin-PPMA (BA)	-27	90	333	74
		Bin-PPMA (MI)	-24	84	332	62
		BinS-PPMA	-23	82	274	98
400		CC	-97	103	136	792
		LOGREG	-68	76	131	541
		LATENT	-22	45	203	34
	0	Bin-PPMA (BA)	-73	80	130	594
		Bin-PPMA (MI)	-69	76	129	561
		BinS-PPMA	-69	76	129	560
	1	Bin-PPMA (BA)	-49	60	139	276
		Bin-PPMA (MI)	-46	57	137	254
		BinS-PPMA	-34	46	120	209
	∞	Bin-PPMA (BA)	-25	48	171	89
		Bin-PPMA (MI)	-23	46	167	82
		BinS-PPMA	-21	45	137	138

BIBLIOGRAPHY

Albert, J.H. and Chib, S. (1993). "Bayesian Analysis of Binary and Polychotomous Response Data." *J. Am. Statist. Assoc.*, 88, 669-679.

Andridge, R. and Little, R.J. (2009). "Extensions of Proxy Pattern-Mixture Analysis for Survey Nonresponse." *Proceedings of the Survey Research Methods Section, American Statistical Association*, 2009, 2468-2482.

Andridge, R. and Little, R.J. (2010). "A Review of Hot Deck Imputation for Survey Nonresponse." *International Statistical Review,* 78, 1, 40-64.

Andridge, R. and Little, R.J. (2011). "Proxy Pattern-Mixture Analysis for Survey Nonresponse." *Journal of Official Statistics*, 27, 153-180.

Baker, S. and Laird, N. (1988). "Regression Analysis for Categorical Variables with Outcome Subject to Nonignorable Nonresponse." *J. Am. Statist. Assoc.*, 83, 62-69.

Bang, H. and Robins, J.M. (2005). "Doubly robust estimation in missing data and causal inference models." *Biometrics*, 61, 962–972.

Binder, D.A. (1982). "Non-Parametric Bayesian Models for Samples from Finite Populations." *J. R. Statist. Soc.*, 44, 388-393.

Brewer, K.R. and Mellor, R.W. (1973). "The Effect of Sample Structure on Analytical Surveys." *Austal. J. Statist.*, 15, 145-152.

Cao, W., Tsiatis, A.A., and Davidian, M. (2009). "Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data." *Biometrika*, 96, 723-734.

Daniels, M.J. and Hogan, J.W. (2000). "Reparameterizing the Pattern Mixture Model for Sensitivity Analyses under Informative Dropout." *Biometrics*, 56, 1241-1248.

DuMouchel, W.H. and Duncan, G.J. (1983). "Using Sample Survey Weights in Multiple Regression Analysis of Stratified Samples." *J. Am. Statist. Assoc.*, 78, 535-543.

Fay, R. (1986). "Causal Models For Patterns of Nonresponse." *J. Am. Statist. Assoc.*, 81, 354-336.

Gelman, A. (2007). "Struggles with Survey Weighting and Regression Modelling." *Statist. Sci.*, 22, 153-164.

Heckman, J. (1976). "The Common Structure of Statistical Models of Truncation, Sample Selection, Limited Dependent Variables and a Simple Estimator for Such Models." *Ann. Econ. Social Meas.*, 5, 475-492.

Kang, D.Y.J. and Schafer, J.L. (2007). "Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data." *Statist. Sci.*, 22, 523–539.

Little, R.J. (1986). "Survey nonresponse adjustments for estimates of means." *Int. Statist. Rev.*, 54, 139-157.

Little, R.J. (1993). "Pattern-Mixture Models for Multivariate Incomplete Data." *J. Am. Statist. Assoc.*, 88, 125-134.

Little, R.J. (1994). "A Class of Pattern-Mixture Models for Normal Incomplete Data." *Biometrika*, 81, 471-483.

Little, R.J. and Rubin, D.B. (2002). Statistical Analysis with Missing Data (Second Edition). New York: Wiley.

Little, R.J. and An, H. (2004). "Robust Likelihood-Based Analysis of Multivariate Data with Missing Values." *Statistica Sinica*, 14, 949–968.

Little, R.J. and Vartivarian, S. (2005). "Does weighting for nonresponse increase the variance of survey means?" *Statistics Canada*, 31, 161-168.

Little, R.J. (2012). "Calibrated Bayes, an Alternative Inferential Paradigm for Official Statistics." *Journal of Official Statistics*, 28, 1-27.

Nandram, B., and Choi, J.W. (2002). "A Bayesian Analysis of a Proportion Under Non-Ignorable Nonresponse." *Statistics in Medicine*, 21, 1189-1212.

Nordheim, E. (1984). "Inference From Nonrandomly Missing Categorical Data: An Example From a Genetic Study on Turner's Syndrome." *J. Am. Statist. Assoc.*, 79, 772-780.

Pfeffermann, D. and Sikov, A. (2011). "Imputation and Estimation under Nonignorable Nonresponse in Household Surveys with Missing Covariate Information." *Journal of Official Statistics*, 27, 181-209.

Robins, J.M., Rotnitzky, A., and Zhao, L.P. (1994). "Estimation of regression coefficients when some regressors are not always observed." *J. Am. Statist. Assoc.*, 89, 846–866.

Rosenbaum, P.R. and Rubin, D.B. (1983). "The central role of the propensity score in observational studies for causal effects." *Biometrika*, 70,41-55.

Rosenbaum, P.R. and Rubin, D.B. (1984). “Reducing bias in observational studies using subclassification on the propensity score.” *J. Am. Statist. Assoc.*, 79, 516-524.

Rotnitzky, A., Robins, J.M., and Scharfstein, D.O. (1998). “Semiparametric regression for repeated measures outcomes with non-ignorable non-response.” *J. Am. Statist. Assoc.*, 93, 1321–1339.

Rubin, D.B. (1974). “Characterizing the Estimation of Parameters in Incomplete Data Problems.” *J. Am. Statist. Assoc.*, 69, 467-474.

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

Sarndal, C.E. and Lundstrom, S. (2008). “Assessing Auxiliary Vectors for Control of Nonresponse Bias in the Calibration Estimator.” *Journal of Official Statistics*, 24, 167-191.

Sarndal, C.E. (2011). “The 2010 Morris Hansen Lecture Dealing with Survey Nonresponse in Data Collection, in Estimation.” *Journal of Official Statistics*, 27, 1-21.

Schouten, B. (2007). “A Selection Strategy for Weighting Variables Under a Not-Missing-at-Random Assumption.” *Journal of Official Statistics*, 23, 51-68.

Sullivan, D. and Andridge, R. (2015). “A Hotdeck Imputation Procedure for Multiply Imputing Nonignorable Missing Data: The Proxy Pattern-Mixture Hotdeck.” *Computational Statistics & Data Analysis*, 82, 173-185.

Tsiatis, A.A. and Davidian, M. (2007). “Comment on ‘Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data.’” *Statist. Sci.*, 22, 569–573.

Tsiatis, A.A., Davidian, M., and Cao, W. (2011). “Improved Doubly Robust Estimation When Data Are Monotonely Coarsened, with Application to Longitudinal Studies with Dropout.” *Biometrics*, 67, 536-545.

West, B. and Little, R.J. (2013). “Non-Response Adjustment of Survey Estimates Based on Auxiliary Variables Subject to Error.” *Appl. Statist.*, 62, 213-231.

Zhang, G. and Little, R.J. (2008). “Extensions of the penalized spline propensity prediction method of imputation.” *Biometrics*, 65, 911–918.

Zhang, G. and Little, R.J. (2011). “A comparative study of doubly robust estimators of the mean with missing data.” *J. Stat. Comput. Simul.*, 81, 2039-2058.

Zheng, H. and Little, R.J. (2005). “Inference for the population total from probability-proportional-to-size samples based on predictions from a penalized spline nonparametric model.” *Journal of Official Statistics*, 21, 1-20.