Investigating the Bias of Alternative Statistical Inference Methods in Sequential Mixed-Mode Surveys

Z. Tuba Suzer-Gurtekin[1], Steven G. Heeringa [2], Richard Valliant[3]
[1,2]Michigan Program in Survey Methodology, Institute for Social Research, Ann Arbor, MI
[3] Michigan Program in Survey Methodology, Institute for Social Research, Ann Arbor, Michigan and Joint Program in Survey Methodology, University of Maryland, College Park, MD

Abstract
Sequential mixed-mode surveys combine different data collection modes sequentially to reduce nonresponse bias under certain cost constraints. However, as a result of nonignorable mode effects, nonrandom mixes of modes may yield unknown bias properties for population estimates such as means and totals. The assumption of ignorable mode effects governs the existing inference methods for sequential mixed-mode surveys. The objective of this paper is to describe and empirically evaluate the proposed multiple imputation estimation methods that account for both nonresponse and nonrandom mixtures of modes in a mixed-mode survey. This paper presents some empirical and simulation results for the bias of mean wage and salary income based on the public-use Current Population, 1973, Social Security Records Exact Match data.

Key Words: Mixed-mode Surveys, Multiple Imputation, Mode Effects, Selection Model

1. Introduction

In a sequential mixed-mode survey design, a combination of data collection modes is used sequentially to reduce nonresponse bias under certain cost constraints (Brambilla & McKinlay, 1987; Cobben, Schouten, & Bethlehem, 2006; De Leeuw, 2005; Hochstim, 1967). However, as a result of nonignorable mode effects nonrandom mixes of modes may yield unknown bias properties for the population estimates such as means and totals.
In sequential mixed-mode surveys, two potentially confounded mechanisms should be accounted for in the mode effect evaluations and the data adjustments. First, there is a potential for a mode choice bias, a conditioned response selection mechanism, in which different types of people may respond across phases and different mode alternatives. The second of the two confounded sources of survey error may be real measurement differences among the response modes. These two mechanisms, nonrandom response selection and measurement differences across modes, are inherently confounded in sequential mixed-mode survey data. Without any adjustment the survey inference will be inconsistent due to possible variations in the composition of respondents per mode. Even in the absence of nonresponse error, the estimates from sequential mixed-mode survey data would have unknown bias properties.
A simple example will illustrate the more general problem. Suppose that subgroup g in the population responds only via a particular mode. The response y_i of any unit in subgroup g can be modeled as $y_i = \mu_i + B_g + \varepsilon_i$ $(i \in U_g)$ where U_g is the set of units in

g in the population, μ_i is the true value for unit i, B_g is a bias for any unit in U_g, and ε_i is a random error with mean 0. Suppose a simple random sample of size n is selected and the sample mean, $\bar{y} = \sum_{i=1}^{n} y_i \big/ n$, is used to estimate the population mean. The bias of $\bar{y}$ can be shown to be $\sum_g P_g B_g$ where P_g is the proportion of the population in subgroup g. Thus, without some correction the sample mean is biased. Of course, the real situation is far more complicated because each mode may have different biases for different groups and there is usually no control in assigning subgroups of respondents to modes in sequential designs (Aquilino, 1994; Moore, Stinson, & Welniak, Jr., 2000). To address this issue, a proper bias adjustment should incorporate varying response effects by mode so that survey estimates that combine the observations from the multiple modes are on the same underlying scale.

While some statistical inference methods have been developed to address this nature of mixed-mode surveys, the assumption of ignorable mode effects governs the existing inference methods (Buelens & Van den Brakel, 2011; Cobben, 2009; Elliott, Little, & Lewitzky, 2000; Glynn, Laird, & Rubin, 1993; Hansen & Hurwitz, 1946; Little & Rubin, 2002; Rubin, 1987; Vannieuwenhuyze, Loosveldt, & Molenberghs, 2010, 2012). Thus, full population statistical inference methods that account for both final nonresponse and nonrandom mode effects have yet to be developed (Buelens & Van den Brakel, 2011; Cobben, 2009; Vannieuwenhuyze, Loosveldt, & Molenberghs, 2010, 2012). This paper aims to present and to evaluate multiple imputation methods which incorporate adjustments for both nonresponse and nonrandom mode effects under a sequential mixed-mode survey design. The proposed methods conceptualize the sequential mixed-mode survey response patterns as a special case of a missing data problem and use a series of multiple imputation models to create completed mode-specific data vectors conditioned on the observed data for response mode and sample unit covariates. These mode-specific completed data vectors are used to address two research questions in particular: (1) Are the measurement error differences between modes ignorable? and, (2) What are the properties of statistical inference methods that incorporate nonignorable measurement error differences under a sequential mixed-mode survey design?

To explore the first research question, multiple imputation inference techniques are applied to the completed mode-specific data vectors to compute sample means and standard errors separately (Rubin, 1987). These means and standard errors are used to compare the differences in the mean estimates of the population distribution of the variable of interest by mode. To explore the second research question, the empirical properties of alternative methods in combining separate mode-specific mean estimates are investigated.

For the empirical evaluations, the public-use Current Population Survey, 1973, and Social Security Records Exact Match is used in this paper. While the analytical methods discussed in this paper are also applicable to the other survey items, wages and salary income is chosen for testing the proposed methods. CPS is a mixed-mode survey and the person level Internal Revenue Service (IRS) income match data provide benchmarks to evaluate the proposed methods.

2. Methods

2.1 Existing Statistical Inference Methods

2.1.1 Frequentist survey inference with nonresponse follow-up

Hansen and Hurwitz (1946) proposed a double-sampling technique for nonresponse follow-up and derived a design-based unbiased mean estimator and its variance estimator under certain assumptions. This inferential method is a special case of sequential mixed-mode survey inference. In particular, it assumes deterministic nonresponse, full follow-up response and no measurement differences between modes. The literature that extends Hansen and Hurwitz's mixed-mode model with double-sampling falls under the frequentist (Rao, 1983) and the Bayesian approaches (Singh, 1983). Importantly, these frameworks extended the double-sampling method to the cases when full follow-up response is not achieved.

2.1.2 Bayesian survey inference in the context of nonignorable nonresponse

The double-sampling technique's deterministic nonresponse assumption implies nonignorable nonresponse. The literature on Bayesian inference methods has extended Hansen and Hurwitz's double-sampling technique to the nonignorable nonresponse context and relaxed the full follow-up response assumption often assuming a stochastic view of nonresponse (Lessler & Kalsbeek, 1992; Rubin, 1987). Among the three nonignorable nonresponse models that the Bayesian framework distinguishes, the selection models (Heckman, 1979; Glynn, Laird, & Rubin, 1986; 1993; Little, 1993; Little & Rubin, 2002) are of particular interest since they allow adjustment for the nonrandom selection mechanism in the imputation models. In the method that is applied in this paper, selection models are used to model the mode choice to simulate the nonrandom response selection. Suppose the full likelihood function in which data, Y, and missing data patterns, M, depend on general parameters θ and ψ is:

$$L_{full}(\theta,\psi \mid Y,M) \propto f(Y,M \mid \theta,\psi)$$

In the general Bayesian framework, the complete data Y are partitioned into observed and missing values such as (Y_{obs}, Y_{mis}) and $M_{(nx1)}$ is the missing-data indicator vector. For the purposes of this paper, Y is considered to be univariate. The $M_{(nx1)}$ vector captures the stochastic nonresponse mechanism that is not known. A nonignorable missing mechanism by definition implies that nonresponse is related to Y_{mis} and the association between the nonresponse and Y_{mis} cannot be fully modeled when conditioned on any available covariates. The full likelihood function of θ and ψ is proportional to the joint distribution of Y and M conditioned on θ and ψ, which are assumed to be distinct parameters. The likelihood function of Y and M conditioned on θ and ψ can be written as a joint distribution of y_i and M_i, where the i subscript represents an observation. The nonignorable nonresponse models differ on the parameterization of the joint distribution:

$$f(M,Y \mid \theta,\psi) = \prod_U f(M_i, y_i \mid \theta,\psi),$$

assuming that observations are independent and identically distributed.

According to the Bayesian framework, the selection models factor the joint distribution of Y and M into two probability distributions: (1) a complete data model which is the probability distribution for Y with density $f(Y \mid \theta)$ indexed by θ, an unknown parameter vector and (2) a model for the missing data mechanism which is the probability distribution of $f(M \mid Y,\psi)$ for M given Y indexed by ψ, another unknown parameter vector.

$$f(M_i, y_i \mid \theta,\psi) = f(y_i \mid \theta) f(M_i \mid y_i,\psi)$$

While the assumption of ignorable mode effects governs the existing inference methods for sequential mixed-mode surveys, the recent literature has studied some methods that evaluate and assess the measurement effects in the mixed-mode surveys.

2.1.3 Existing methods to assess measurement effects in mixed-mode surveys

In the literature, currently there are two methods that aim to assess measurement effects in mixed-mode surveys. The first method initially assigns survey modes to respondents randomly, and then reinterviews the respondents in the other survey mode (Biemer, 2001; Jäckle, Roberts, & Lynn, 2010). The second method conducts a single mode survey in parallel to a mixed-mode survey and assesses the mode effects by mixture distribution assumptions (Vannieuwenhuyze et al., 2010, 2012).

2.1.4 Existing methods to adjust measurement effects in mixed-mode surveys

Buelens and Van den Brakel (2011) calibrate the mode proportions to fixed proportions by including the mode as a variable in the calibration estimator. This method does not eliminate the bias, but instead aims to calibrate the bias in the survey mean estimator to yield unbiased change estimates. In an alternative method, Cobben (2009) uses selection models to adjust for nonresponse for the sequential nature of mode choice and identifies the adjustment for measurement effects as an open research area (Cobben et al., 2006).

2.2 Proposed Methods

Unlike the previous methods, this paper proposes a multiple imputation method through which mode choice and measurement effects are isolated analytically. In this method, sequential mixed mode survey response patterns are treated as a special case of missing data problem in which the preceding or the following response data are considered as nonresponse. The data for the nonreporting units in each phase are imputed to create a completed data vector as if all the units had reported in that particular mode. The general approach includes five steps:

(1) Expand the respondents for a given wave of data collection to the full sample via multiple imputation models.

(2) Estimate means from each mode's completed (observed + imputed) data set. Use the multiple imputation method to reflect the uncertainty associated with imputation (Rubin, 1987). While M=5 is often used, more recent evidence shows that a greater number of imputations is required when the missing fraction is high (Graham, Olchowski, & Gilreath, 2007). In this paper, M=5 is used, but the value of M (number of imputations) will be determined empirically in the future research extensions.

(3) Compare estimates from the different modes to assess mode differences.

(4) Compare results from different modes and the competing method to available benchmark values to determine whether mode estimates are biased.

(5) If significant mode effects are noted, compare results from alternative ways of combining mode estimates to available benchmark values.

2.2.1 Model for selection models for continuous variables

Most of the literature imposes the normality assumption on *Y* in selection bias modeling (Greene, 2011; Rubin, 1987; Little & Rubin, 2002), although many important variables collected in surveys are non-normal. In the maximum likelihood estimation approach, the full likelihood is built up as a multiplication of respondents' and nonrespondents' likelihood functions conditioned on the selection mechanism and the distributional assumption for the response variable (Greenlees, Reece, & Zieschang, 1982). Greenlees, Reece, and Zieschang (1982) (from now on denoted by GRZ) used a logistic function instead of a probit function in modeling the response mechanism. Define $X_i^{(M)}$ and $X_i^{(Y)}$ to be the covariates on which the missing mechanism and response variable are conditioned. $\beta^{(M)}$ and $\beta^{(Y)}$ are the model parameters for the selection and outcome equations, respectively. The GRZ approach allows imputation of the expected values of y_i conditional on X and missingness. Assuming a normal distribution for the response variable and a logistic function for the missing data mechanism, the selection model and full likelihood functions are as follows:

$$(Y_i \mid X_i^{(Y)};\theta) \sim N(X_i^{(Y)};\beta^{(Y)},\sigma^2) \text{ where } \theta = (\beta^{(Y)},\sigma^2)$$

$$\Pr(M_i = 0 \mid X_i^{(M)}, Y_i;\psi) = \left[1+\exp(-X_i^{(M)}\beta^{(M)} - \gamma Y_i)\right]^{-1} \text{ where } \psi = (\beta^{(M)},\gamma)$$

$$L_{full}(\theta,\psi \mid Y,M) = \prod_{i\in r} \frac{1}{\left[1+\exp(-X_i^{(M)}\beta^{(M)} - \gamma Y_i)\right]} \frac{1}{\sigma}\Phi\left(\frac{Y_i - X_i^{(Y)}\beta^{(Y)}}{\sigma}\right)\times$$

$$\prod_{i\in nr} \int_{-\infty}^{\infty} \left(1 - \frac{1}{\left[1+\exp(-X_i^{(M)}\beta^{(M)} - \gamma Y_i)\right]}\right) \frac{1}{\sigma}\Phi\left(\frac{Y_i - X_i^{(Y)}\beta^{(Y)}}{\sigma}\right) dy_i$$

Since the integral cannot be exactly evaluated, it is approximated by ten-point Gauss-Hermite quadrature. Given the full likelihood function, maximum likelihood estimation is performed using the Broyden, Fletcher, Goldfarb and Shanno (BFGH) method as implemented in the R *optim* function in the stats package[1].

2.2.2 Imputation for selection models for continuous variables

The maximum likelihood estimates $(\hat{\beta}^{(Y)},\hat{\sigma}^2,\hat{\beta}^{(M)},\hat{\gamma})$ given are plugged into an imputation model (Greenlees et al., 1982) as follows:

(1) Draw ε_i randomly from $N(0,1)$.

(2) Compute $\hat{y}_i = X_i^{(Y)}\hat{\beta}^{(Y)} + \hat{\sigma}\varepsilon_i$.

(3) Compute $\Pr(M_i = 1 \mid \hat{y}_i, X_i^{(M)}, \hat{\beta}^{(M)}, \hat{\gamma}) = 1 - \frac{1}{[1+\exp(-X_i^{(M)}\hat{\beta}^{(M)} - \hat{\gamma}\hat{y}_i]}$.

(4) Draw a random number η from a uniform distribution $[0,1]$.

(5) Save the $\hat{y}_i$ as the imputed value for observation i if $\Pr(M_i = 1 \mid \hat{y}_i, X_i^{(M)}, \hat{\beta}^{(M)}, \hat{\gamma}) \geq \eta$; otherwise repeat the imputation steps 1-5.

2.2.3 Alternative combination methods

In these investigations, three alternative methods are used to combine the mode-specific estimates:

CM_1: Mode-specific estimates are combined as a simple average.

CM_2 : This method weights the mode-specific means inversely to the variance of the mode-specific estimates

CM_3 : The third method weights mode-specific means inversely to the mean square error of the mode-specific estimates

These estimation methods yield population estimates which have theoretically known bias properties and aim for the minimum variance or the minimum mean square error for the combined single estimator.

---

[1] R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

2.3 Current Population Survey, 1973, and Social Security Records: Exact Match Data[2]

Current Population Survey (CPS) is a rotating panel survey that produces data on the U.S. labor force. The panel rotation scheme follows a 4-8-4 pattern for a selected household. A sample household is interviewed for two four consecutive waves which are eight months apart. CPS is a mixed-mode survey which includes in-person and telephone modes. Except for the first and fifth wave interviews, interviews are mostly conducted by telephone, but for the first and fifth waves the dominant mode is in-person.

In a joint project, the U.S. Census Bureau and Social Security Administration matched the 1973 CPS March data with Social Security benefit and earnings records and released the data to the public. Additionally, a limited set of tax items provided by the Internal Revenue Service (IRS) from the 1972 Federal Income Tax are also available for a subset of respondents in the same dataset.

In addition to the survey mode, there are some other measurement error sources of the CPS data collection, such as proxy reporting, and dependent interviewing, that may contribute into the varying biases. For this investigation, not all the measurement error sources are taken into account, but a subset of data is selected to eliminate other possible measurement errors to a degree. Also, since the telephone interviewing was not centralized in 1973, this may be speculated to produce larger interviewer related survey error on the survey estimates. However, the data for the interviewers are not available in this dataset to perform this evaluation.

The empirical evaluations are constrained to a subset of data that eliminates the possible measurement error sources. The analysis dataset includes household heads, who are married and whose spouses are present, who reported a non-farm residence, whose source of income is wage and salary only, who worked in a non-agricultural industry full-time full-year in 1972, who were married taxpayers filing jointly, whose wives did not work in 1972, and had IRS matched records that were identified as a good-match. In addition, among this subset who reported salary and wages less than $600 were also excluded. Since there is no variation in the CPS and the IRS top-coded records, and this proposed method is expected to be implemented on the raw data, these top-coded records were excluded from further analysis as well. The final sample size for this subset is 5,425. In the simulations, this subset was considered as the population and random replicates were sampled to perform the empirical evaluations.

Table 1 reports the response mode distribution by wave for this sample. The distribution of response mode follows a similar pattern in this subset of data to the one at the aggregate. While in-person mode is the dominant mode in the first and the fifth waves, telephone mode is preferred by about two-thirds of the sample in the other months.

In this investigation, the variable of interest is the wage and salary income as reported in the CPS and the mean wage and salary income is the estimate of interest. Without controlling for the covariates, the mean comparison suggests that reported wage and salary income by mode is different in this subset of data. On the average, a person who has responded by telephone mode earns $1,691 per year less. Table 2 reports the quintiles of the wage and salary income by mode. After controlling for the covariates, the mean difference shrinks by about 40%, but it is still significant. Controlling covariates include the personal characteristics, education, work experience, race (white vs. other), occupation type (professional, sales, craft, laborer), and industry (construction,

[2] [ICPSR 7616]. ICPSR version. Washington, DC: U.S. Dept. of Commerce, Bureau of the Census and Social Security Administration, Long-Range Research Branch [producer], 197?. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2001. doi:10.3886/ICPSR07616

manufacturing, transportation, trade, service) and residential (household) characteristics, central city, suburb, region.

Since the person level IRS data are available, they can be compared against reported wage and salary income. The relative bias, $\text{RelBias}_{CPS} = \left(\bar{y}_{CPS} - \bar{y}_{IRS} / \bar{y}_{IRS}\right)$, is not significantly different between the in-person and the telephone modes. In addition, a larger number of overreporting outliers compared to the number of underreporting outliers is observed for both the in-person and the telephone modes. This pattern is the opposite of what is reported in the literature, even though the modest bias magnitude agrees with the previous findings (Moore et al., 2000).

The selection model covariates, $X^{(M)}$, and the outcome model covariates, $X^{(Y)}$ are the same as in the GRZ selection and outcome models with two exceptions. In this paper, the response mode is the dependent variable in the selection model, and wave in sample is included as one of the selection model covariates.

GRZ and the extensions of their work have studied the properties of the imputation models for the item missing in reported wage and salary income (Greenlees et al., 1982; Glynn et al., 1986, 1993). This paper includes inclusion of item missing as one of the simulation parameters. While the overall item missing percent for this subset is 10%, 59% of the item missing is in the telephone survey mode data. The item missing includes both the refusals and the other types of missing data in this investigation.

Table 1: Response Mode Distribution by Wave in Sample

Response	**Wave in Sample**							
Mode	**W1**	**W2**	**W3**	**W4**	**W5**	**W6**	**W7**	**W8**
Telephone	3%	65%	64%	71%	7%	59%	72%	69%
In-person	97%	35%	36%	29%	93%	41%	28%	31%

Table 2: Sample Quintiles of Reported Wage and Salary Income

Response	**Quintiles**						
Mode	**0%**	**10%**	**25%**	**50%**	**75%**	**90%**	**100%**
Telephone	1000	7388.8	9467.5	12000	16000	22800	50000
In-person	80	6000	8300	11000	14700	20000	50000

Table 3: Sample Percentage of Item Missing in Reported Wage and Salary Income

Response Mode	**% of Item Missing**
Telephone	12%
In-person	8%
Overall	10%

2.4 Evaluation of Proposed Methods

A total of eight simulations were performed varying three parameters: (1) Replicate sample size (400 and 800), (2) Whether to include item missing in the imputations or not, and (3) Imputation model specification: deterministic mode choice regression model versus stochastic mode choice regression model (selection model). An equal number of respondents was drawn from each wave in each of the replicates under fixed sample sizes of 400 and 800 from the subset of the CPS data as defined. The complete datasets were then created for telephone and in-person modes using the both imputation models. The completed datasets were used to compute the mode-specific mean wage and salary income. Multiple imputations were combined using the usual multiple imputation combination rules to produce mode-specific means (Rubin, 1987). The mode-specific means were compared in terms of three evaluation criteria: (1) Number of significant differences, (2) Mean relative bias, and (3) Mean absolute relative bias. The mode-specific estimates were combined under four methods: (1) Simple average (Combination Method 1 – CM1), (2) Weighted inversely to variance (Combination Method 2 – CM2), (3) Weighted inversely to mean square error (Combination Method 3 – CM3), and (4) Ignoring the measurement differences (the competing combination method – CM4). Each simulation included 50 replicates and 5 imputations per replicate.

3. Results

The differences between the mode-specific means are not significant given the number of significant differences at 95% confidence level for any of the simulations (see Table 4). On the other hand, both the relative and the absolute relative bias is consistently larger for the telephone mode-specific means on the average than the ones for the in-person mode-specific means in all of the simulation variations (Figures 1-4). The difference between the telephone and in-person mode-specific wage and income means stays same for both variations of including or excluding the item missing variations.
In terms of the relative bias, the Combination Methods 1 and 2 yield same bias levels on the average across all the variations of the simulations. These two methods do not outperform the competing method in which telephone and in-person responses are combined without any adjustments. On the other hand, while it is not statistically different across all the variations of the simulations, the Combination Method 3 yields consistently lower levels of bias compared to the competing combination method in the deterministic mode choice regression model. In the simulation variations in which item missing is excluded, the Combination Method 3 significantly outperforms the competing combination method on the average in terms of the relative bias. The absolute relative bias results follow the same pattern.
However, under the stochastic mode choice regression model imputation simulations, none of the alternative combination methods outperforms the competing method in terms of the mean relative bias (Figures 3-4). The competing method consistently outperforms the alternative combination methods across all the variations of the simulation.

Table 4: Number of Significant Differences at 95% confidence level between Telephone and In-person Mode-specific Estimates

	400x50x5		**800x50x5**	
	Deterministic Mode Choice Regression Model	**Stochastic Mode Choice Model**	**Deterministic Mode Choice Regression Model**	**Stochastic Mode Choice Model**
Item Missing Excluded	1	1	3	8
Item Missing Included	1	1	3	9

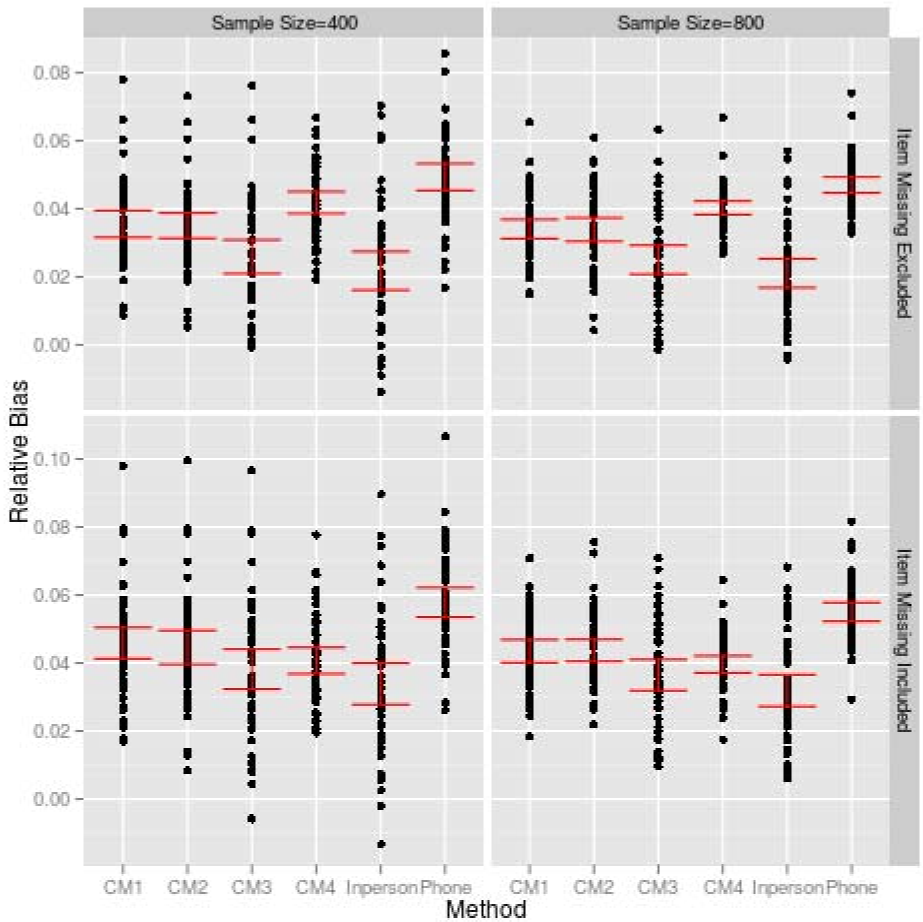

Figure 1: Relative biases ($\text{RelBias}_{CM} = \left(\bar{y}_{CM_i} - \bar{y}_{IRS} / \bar{y}_{IRS}\right)$, where CM_i for $i = 1,2,3,4$ is the combination method)[3] in 50 samples[4] of estimates of wage and salary income mean with the four alternative methods of estimation (CM1, CM2, CM3, and CM4=Competing Combination Method) and mode-specific imputed data (In-person and Telephone) by item missing treatment procedure under the deterministic mode choice regression model. Sample sizes are 400 and 800 each for the samples; five imputations were performed for each sample, the red error bars represent the 95% confidence intervals.

[3] Same formula is used for the mode specific mean estimates in which CM_i is replaced by the in-person and telephone estimates.
[4] The model parameters are not estimated in one replicate in sample size=400 simulations due to zero sample size cells.

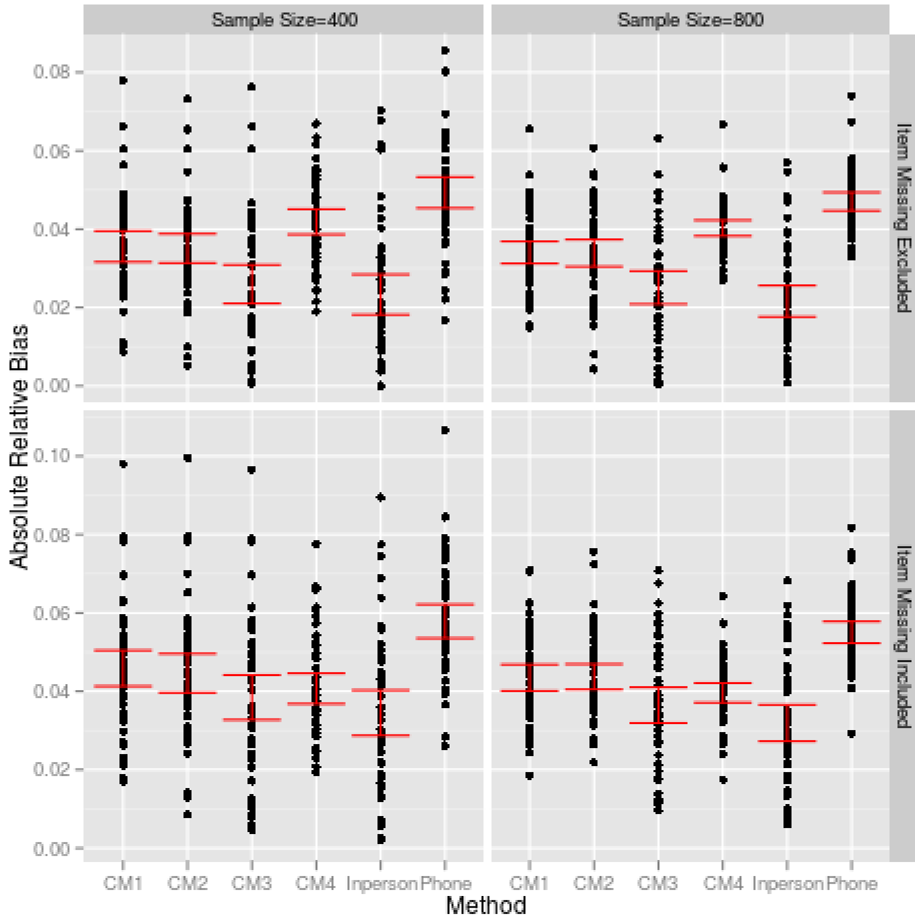

Figure 2: Absolute relative biases ($\text{Abs RelBias}_{CM} = \left(\left|\bar{y}_{CM_i} - \bar{y}_{IRS}\right| / \bar{y}_{IRS}\right)$, where CM_i for $i = 1,2,3,4$ is the combination method)[5] in 50 samples[6] of estimates of wage and salary income mean with the four alternative methods of estimation (CM1, CM2, CM3, and CM4=Competing Combination Method) and mode-specific imputed data (In-person and Telephone) by item missing treatment procedure under the deterministic mode choice regression model. Sample sizes are 400 and 800 each for the samples; five imputations were performed for each sample, the red error bars represent the 95% confidence intervals.

[5] Same formula is used for the mode specific mean estimates in which CM_i is replaced by the in-person and telephone estimates.
[6] The model parameters are not estimated in one replicate in sample size=400 simulations due to zero sample size cells.

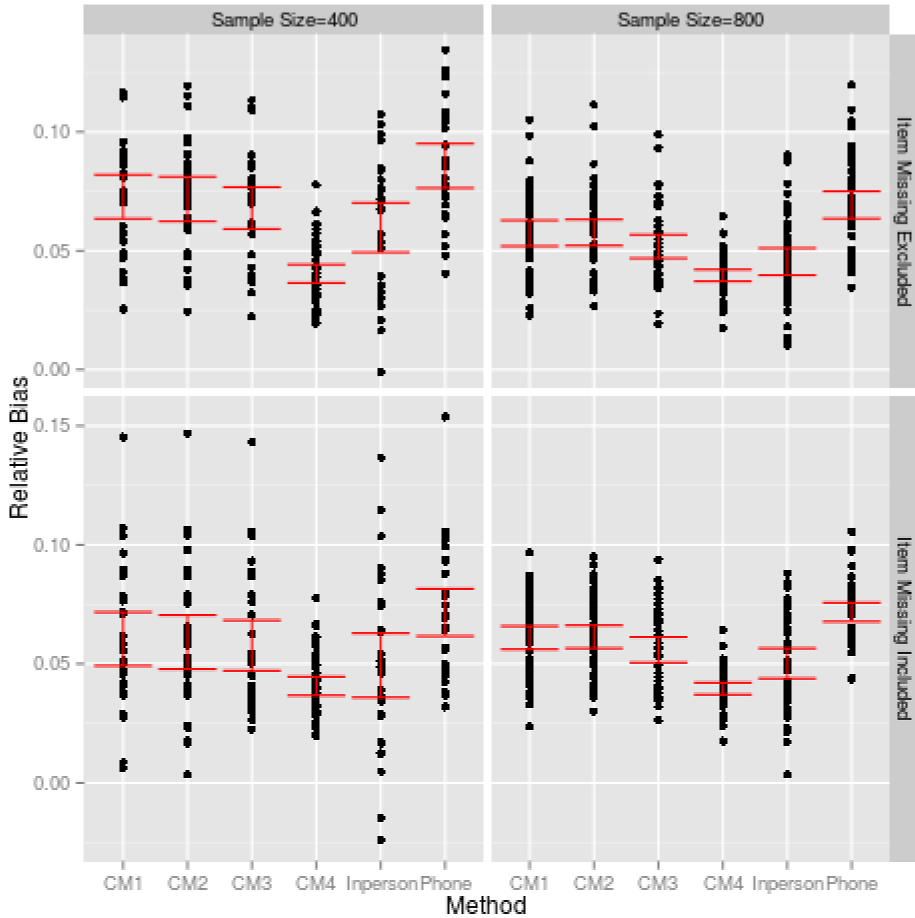

Figure 3: Relative biases ($\text{RelBias}_{CM} = \left(\bar{y}_{CM_i} - \bar{y}_{IRS} / \bar{y}_{IRS}\right)$, where CM_i for $i = 1,2,3,4$ is the combination method)[7] in 50 samples[8] of estimates of wage and salary income mean with the four alternative methods of estimation (CM1, CM2, CM3, and CM4=Competing Combination Method) and mode-specific imputed data (In-person and Telephone) by item missing treatment procedure under the stochastic mode choice regression model. Sample sizes are 400 and 800 each for the samples; five imputations were performed for each sample, the red error bars represent the 95% confidence intervals.

[7] Same formula is used for the mode specific mean estimates in which CM_i is replaced by the in-person and telephone estimates.
[8] The model parameters are not estimated in 3 and 18-20 replicates, respectively in sample size=800 and sample size=400 simulations, due to zero sample size cells.

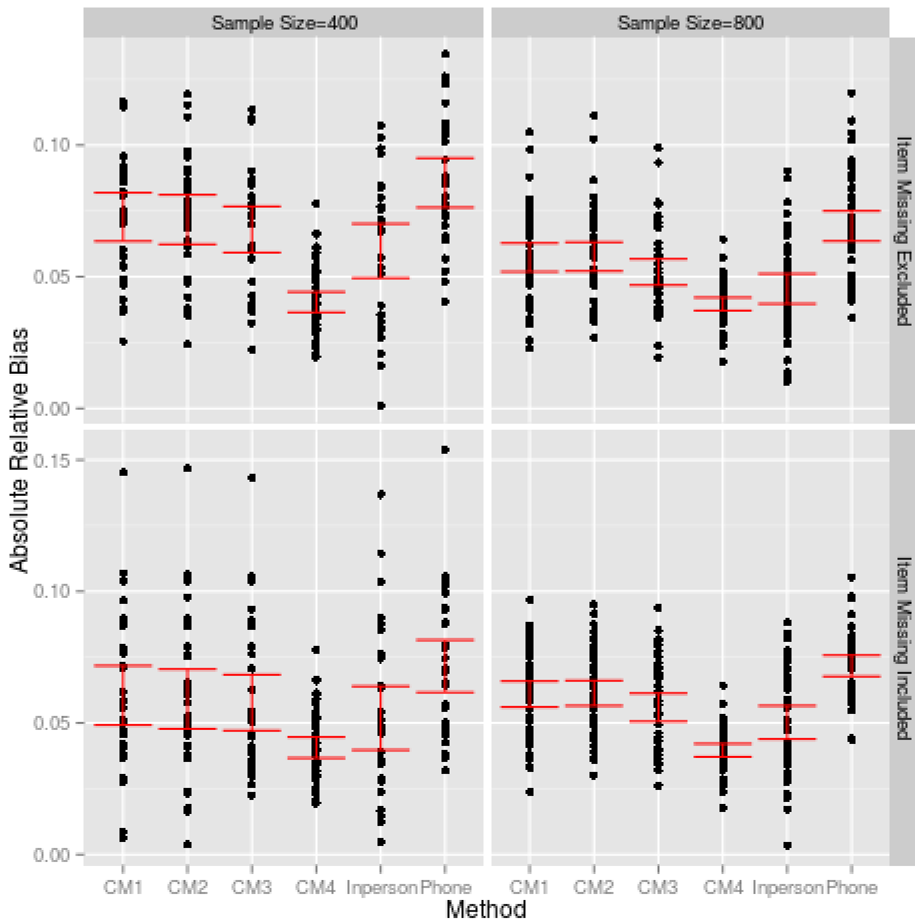

Figure 4: Absolute relative biases ($\text{Abs RelBias}_{CM} = \left(\left| \bar{y}_{CM_i} - \bar{y}_{IRS} \right| / \bar{y}_{IRS} \right)$, where CM_i for $i = 1, 2, 3, 4$ is the combination method)[9] in 50 samples[10] of estimates of wage and salary income mean with the four alternative methods of estimation (CM1, CM2, CM3, and CM4=Competing Combination Method) and mode-specific imputed data (In-person and Telephone) by item missing treatment procedure under the stochastic mode choice regression model. Sample sizes are 400 and 800 each for the samples; five imputations were performed for each sample, the red error bars represent the 95% confidence intervals.

4. Discussion and Extensions

In the previous literature, a negative bias was reported for the CPS wage and salary income (Moore et al., 2000). But our findings do not agree with the previous literature reports. The overall mean relative bias is positive in this dataset. This may be a result of differences in the dataset selection, unit of analysis and the relative bias measures

[9] Same formula is used for the mode specific mean estimates in which CM_i is replaced by the in-person and telephone estimates.
[10] The model parameters are not estimated in 3 and 18-20 replicates, respectively in sample size=800 and sample size=400 simulations, due to zero sample size cells.

(Herriot & Spiers, 1975). In order to understand the nature of the mode effects in these evaluations, this disagreement requires further research.

Overall the relative bias magnitude was moderate for the mode-specific and the combined means. Related to the first research question, we found that the relative and the absolute relative biases of the telephone mode-specific estimates were higher than those of the in-person estimates. On the other hand, the error sources in the mode-specific estimates are not known without studying the error sources in a randomized experiment, the differences in the relative bias between the mode-specific means are only evaluated by controlling the available covariates analytically.

In addressing the second research question, among the proposed combination methods, only Combination Method 3 improved the relative bias of wage and salary mean under the deterministic mode choice regression imputation model. This improvement was not observed when the item missing was included in the imputations.

In this investigation, a separate selection mechanism was not considered in the imputation of the item missing. In one variation of the simulations, the item missing was treated the same as the other mode responses and were imputed by the same multiple selection models. The Combination Method 3 did not improve the relative bias of the wage and salary income mean in this variation of the simulation under the deterministic mode choice regression model. This suggests that there may be a different mechanism that needs to be included for the treatment of item missing in the mode-specific imputations.

In the stochastic mode choice regression model simulations, none of the alternative methods outperformed the competing combination method in terms of the relative bias. The algorithm also failed to produce parameter estimates in some of the replicate samples (3 in sample size=800 simulations, 18-20 in sample size=400) as a result of no control over the cell sample sizes by survey wave and response mode. Therefore, further simulations will be performed to investigate the covariate cell size and the covariance structure requirements for the stochastic mode choice regression models in order to achieve lower relative bias compared to the competing method which combines the in-person and the telephone data without any adjustments in terms of the relative bias.

References

Aquilino, W. S. (1994). Interview mode effects in surveys of drug and alcohol use: A field experiment. Public Opinion Quarterly, 58(2), 210-240.

Biemer, P. (2001). Nonresponse bias and measurement bias in a comparison of face to face and telephone interviewing. Journal of Official Statistics, 17(2), 295-320.

Boudreaux, M., Ziegenfuss, J. Y., Graven, P., Davern, M., & Blewett, L. A. (2011). Counting Uninsurance and Means-Tested Coverage in the American Community Survey: A Comparison to the Current Population Survey. Health Services Research, 46(1), 210-231.

Brambilla, D. J., & McKinlay, S. M. (1987). A comparison of responses to mailed questionnaires and telephone interviews in a mixed mode health survey. American Journal of Epidemiology, 126(5), 962-71.

Buelens, B., & Van den Brakel, J. (2011). Inference in Surveys with Sequential Mixed-Mode Data Collection. Discussion Paper. Statistics Netherlands. Retrieved from http://www.cbs.nl/NR/rdonlyres/C70ED9D5-6199-4E27-B37D-792A161D614D/0/2011x1021.pdf

Cobben, F. (2009). Nonresponse in Sample Surveys Methods for Analysis and Adjustment. cbs.nl. Universiteit van Amsterdam. Retrieved from

http://www.cbs.nl/NR/rdonlyres/2C300D9D-C65D-4B44-B7F3-377BB6CEA066/0/2009x11cobben.pdf
Cobben, F., Schouten, B., & Bethlehem, J. (2006). A Model for Statistical Inference based on Mixed Mode Interviewing. European Conference on Quality in Survey Statistics. Retrieved from http://www.statistics.gov.uk/events/q2006/downloads/T10_Cobben.doc
De Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. Journal of Official Statistics, 21(2), 233-255.
Elliott, M. R., Little, R. J. A., & Lewitzky, S. (2000). Subsampling Callbacks to Improve Survey Efficiency. Journal of the American Statistical Association, 95(451), 730-738.
Glynn, R. J., Laird, N. M., & Rubin, D. B. (1986). Selection modeling versus mixture modeling with nonignorable nonresponse. In H. Wainer (Ed.), Drawing inferences from self-selected samples (pp. 115-142). New York: Springer- Verlag.
Glynn, R. J., Laird, N. M., & Rubin, D. B. (1993). Multiple imputation in mixture models for nonignorable nonresponse with follow-ups. Journal of the American Statistical Association, 88(423), 984-993.
Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prevention Science, 8(3), 206–213. Springer.
Greene, W. H. (2011). Econometric Analysis (7th ed.). New York: Macmillan.
Greenlees, J. S., Reece, W. S., & Zieschang, K. D. (1982). Imputation of missing values when the probability of response depends on the variable being imputed. Journal of the American Statistical Association, 77(378), 251-261.
Hansen, M., & Hurwitz, W. (1946). The problem of nonresponse in sample surveys. The American Statistician, 41(236), 517-529.
Heckman, J. J. (1979). Sample selection bias as a specification error. Econometrica: Journal of the Econometric Society, 47(1), 153-161.
Herriot, R. A., & Spiers, E. F. (1975). Measuring the Impact on Income Statistics of Reporting Differences Between the Current Population Survey and Administrative Sources. American Statistical Association Proceedings of the Social Statistics Section, 147-158.
Hochstim, J. R. (1967). A critical comparison of three strategies of collecting data from households. Journal of the American Statistical Association, 62(319), 976-989.
Jäckle, A., Roberts, C., & Lynn, P. (2010). Assessing the Effect of Data Collection Mode on Measurement. International Statistical Review, 78(1), 3-20.
Lessler, J. T., & Kalsbeek, W. D. (1992). Nonsampling error in surveys. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics Section.
Little, R.J.A., & Rubin, D. B. (2002). Statistical analysis with missing data (2nd ed.). Hoboken, New Jersey: Wiley Series in Probability and Statistics.
Little, R. J.A. (1993). Pattern-mixture models for multivariate incomplete data. Journal of the American Statistical Association, 88(421), 125-134.
Moore, J. C., Stinson, L. L., & Welniak, Jr., E. J. (2000). Income measurement error in surveys: A review. Journal of Official Statistics, 16(4), 331-361.
Rao, P. S. R. S. (1983). Nonresponse and Double sampling Randomization Approach. In W. G. Madow, I. Olkin, & D. B. Rubin (Eds.), Incomplete Data in Surveys Volume 2 Theory and Bibliographies (pp. 97-106). Academic Press, Inc.
Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. Hoboken, New Jersey: Wiley Classics Library.
Singh, B. (1983). Nonresponse and Double Sampling Bayesian Approach. In W. G. Madow, I. Olkin, & D. B. Rubin (Eds.), Incomplete Data in Surveys Volume 2 Theory and Bibliographies (pp. 107-124). Academic Press, Inc.

Vannieuwenhuyze, J., Loosveldt, G., & Molenberghs, G. (2010). A Method for Evaluating Mode Effects in Mixed-mode Surveys. Public Opinion Quarterly, 74(5), 1027-1045.

Vannieuwenhuyze, J., Loosveldt, G., & Molenberghs, G. (2012). A Method to Evaluate Mode Effects on the Mean and Variance of a Continuous Variable in Mixed-Mode Surveys. International Statistical Review.