The Empirical Properties of Survey Quality Indicators Under Non-ignorable Nonresponse

Julia Shin-Jung Lee*

Abstract

Recent developments in survey quality indicators may provide promising alternatives to the response rate. Survey quality indicators evaluate the representativeness of respondents, which reflects the quality of the survey inferences in terms of bias. We investigate the performance of these indicators and their sampling properties under different response rates and non-response mechanisms through simulation experiments. NHANES data from 1999-2008 is used as the finite population for the simulation experiments. The experiments are conducted under two assumptions. First, all covariates associated with the true response probability are available and are included in the model. Second, auxiliary variables such as variables from a frame are known at the subject-level for both respondents and nonrespondents. A total of 72 scenarios from different combinations of experimental parameters are explored. The results are derived from 10,000 replications for each scenario. We describe the findings for each indicator in terms of their ability to anticipate the level of bias that may arise in a specific survey analysis.

Key Words: Survey quality, Indicator, Representativeness, Bias, Nonresponse

1. Introduction

Survey quality takes on different meanings depending on the goals of data users. A widely accepted universal indicator for survey quality is response rate. Response rate quantifies the scope of nonresponse which implies the quality of the survey inferences in terms of bias. However, the association between nonresponse bias and response rate has been shown to be inconsistent (Keeter et al., 2000; Curtin et al., 2000; Merkle and Edelman, 2002; Groves, 2006; Groves and Peytcheva, 2008). Alternative indicators of survey quality that better reflect nonresponse bias would be of great interest to survey researchers (Groves et al. 2008). In recent years, two types of indicators appeared in the survey literature. The first type measures the representativeness of the respondent. The second type aims to aid the bias reduction at weighting adjustment. Both types of indicators provide promising alternatives to the response rate as survey quality indicators.

Specifically, several *representativeness* indicators, or R-indicators, were proposed by the RISQ (Representativity Indicators for Survey Quality) project. R-indicators measure the representativeness of the response; that is, the similarity between the respondent and the sample or target population in terms of available auxiliary information. R-indicators are derived as functions of estimated response propensity. In related work, Särndal and Lundström (2011) derived balance indicators (B-indicators). B-indicators also measure the similarity between the respondents and the sample. B-indicators are motivated by a dissimilarity measure, Mahalanobis distance, computed between the respondent and the sample on the mean auxiliary vectors.

To aid bias reduction, several indicators were proposed as analytic tools under the general title of 'auxiliary variable selection for weighting purposes', such as W-indicator (Schouten, 2007), Q-indicators (Särndal and Lundström, 2008) and H- indicators (Särndal

*Julia Shin-Jung Lee, PhD candidate, Program in Survey Methodology, 426 Thompson St, Ann Arbor, MI 48104

and Lundström, 2010). Both Q- and H-indicators assist in choosing auxiliary variables for nonresponse adjustment, whereas the W-indicator provides a confidence band for the unknown bias.

For simplicity of discussion, both representativeness indicators and bias reduction indicators are termed 'survey quality indicators'. Furthermore, R-indicators, mentioned in many manuscripts published by the RISQ project, are a general term used to describe any survey-level indicators that intend to assess survey quality. Here we use R-indicators to refer specifically to the measures discussed by Cobben, et al (2005) and Schouten, et al (2009). Schouten, et al (2007) considered q^2 measure, proposed by Särndal and Lundström (2008), as a candidate of R indicators. In this article q^2 measure is labeled as Q-indicator to distinguish it from those derived by Shlomo, et al (2009). Properties of these indicators, including motivation, formulation, and use, have previously been described (Cobben, et al (2005), Schouten, et al (2007, 2009), Schouten (2007), Särndal and Lundström (2008, 2010), and Särndal (2011)).

The survey quality indicators R, B, W, Q, and H, may hold high promise as alternatives to the response rate, but their properties under non-ignorable nonresponse has yet to be investigated. The goal of this project is to assess the empirical properties of indicators under non-ignorable nonresponse. A large national survey, the National Health and Nutrition Examination Survey (NHANES), is used as the finite population for the simulation experiment. We will draw samples as simple random samples from the NHANES data set. Simulation parameters include the response rate and bias magnitude. We report indicators of their distribution profile under each simulation scenario, and their ability to anticipate the level of bias that may arise in a specific survey analysis.

The contents of this article are arranged as follows. Section 2 details the methods used in the simulation experiment. In section 2.1, we first describe the finite population and summarize the covariates used in this study. Section 2.2 describes the study design. Section 2.3 justifies the simulation sample selection. Section 2.4 reviews the estimates of survey quality indicators and defines notation. Section 3 discusses the results of the simulation experiments by simulation parameters. Finally, section 4 gives a summary of the findings and gaps in the literature, and suggests future research.

2. Methods

2.1 Data

The simulation study uses data from National Health and Nutrition Examination Survey (NHANES). Five cycles of NHANES data from 1999 to 2008 are combined. Subjects 18 years and older are included. This gives a total sample size of $28,852$ which serves as the true finite population.

A general response model assumes a dependency between response probability and some covariates. The covariates could be survey variables and/or auxiliary variables. Auxiliary variables usually are available from outside sources such as a registry or a sampling frame. Five demographic variables from NHANES are used as auxiliary variables so that the covariance among auxiliary variables better resembles real data encountered in practice.

Specifically, subject-level demographic information, including *G*ender (X_1), *A*ge (X_2), *c*itizenship Status (X_3), *H*ousehold Size (X_4), and *E*thnicity (X_5) are used as auxiliary variables. One continuous survey variable (Y) is created based on these auxiliary variables so that the associations between the survey variable and the auxiliaries can be manipulated.

The true population R^2 is set to be 0.7 for the linear model that created Y using five auxiliary variables. Since *E*thnicity is a factor, four dummy variables are constructed, namely

$X_{5.1}$ through $X_{5.4}$. The response variable is created based on the following linear model:

$$Y_i = 0.5 + 5X_1 + 2X_2 + X_3 - .5X_4 + X_{5.1} + X_{5.2} + X_{5.3} + X_{5.4} + \epsilon_i$$

where $\epsilon_i \sim N(0, V)$ and V is such that $R^2 = 0.7$.

Population summary statistics are: average Y is 98.98 (SD = 49.33); 52% of subjects are female; the mean age is 47.3 years (SD = 20.35); 14.3% of subjects have U.S. citizenship; and the average household size is 3.2 persons (SD = 1.69, range 1 to 7). The distribution of ethnicity is: Mexican American 21.8%, other Hispanic 5.7%, non-Hispanic white 47.5%, non-Hispanic black 21%, and other 4%.

2.2 Study Design

The goal of the study is to assess the empirical properties of indicators under non-ignorable nonresponse. Specific objectives are 1) to compare bias among indicators and 2) to compare variances among indicators. We study the empirical properties of these indicators assuming that all covariates affecting true response probability are included in the response propensity model. We study the condition of a single survey variable of interest. All indicators are computed using the same covariates to ensure the comparability.

Based on the auxiliary variables and a generated survey variable as described above, three response models are studied to create response probabilities that have weak, medium, and strong dependency on the covariates. The response model has a general form of

$$logit(\theta) = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 + \gamma_4 X_4 + \gamma_5 X_{5.1} + \gamma_6 X_{5.2} + \gamma_7 X_{5.3} + \gamma_8 \times X_{5.4} + \gamma_9 Y \quad (1)$$

where θ is the response probability and γ_0, the intercept, is a function of desired response rate.

We study three missing data mechanisms, namely, missing completely at random (MCAR), missing at random (MAR) and non-ignorable missingness (MNAR). Under MCAR, the response probability is a constant, independent of any covariates. Under MAR, the response probability depends on auxiliary variables only. Under MNAR, the response probability depends on both auxiliary variables and survey variables.

Specifically, using the response model in (1), the γ coefficients for auxiliary variables and the survey variable Y can be written as $\boldsymbol{\gamma} = \{\gamma_1, \gamma_2, \ldots, \gamma_9\}$. Under MCAR, $\boldsymbol{\gamma} = \mathbf{0}$. For MAR, $\boldsymbol{\gamma} = \{1, 1, 1, 1, 1, 1, 1, 1, 0\}$, $\boldsymbol{\gamma} = \{2, 2, 2, 2, 1, 1, 1, 1, 0\}$, and $\boldsymbol{\gamma} = \{5, 5, 5, 5, 1, 1, 1, 1, 0\}$, for weak, medium, and strong dependency model, respectively. Formulas below listed the response model under MNAR. The weak dependence model:

$$logit(\theta) = \gamma_{0w} + 1X_1 + 1X_2 + 1X_3 + 1X_4 + 1X_{5.1} + 1X_{5.2} + 1X_{5.3} + 1X_{5.4} + 1Y$$

The medium dependence model:

$$logit(\theta) = \gamma_{0m} + 2X_1 + 2X_2 + 2X_3 + 2X_4 + 1X_{5.1} + 1X_{5.2} + 1X_{5.3} + 1X_{5.4} - 10Y$$

The strong dependence model:

$$logit(\theta) = \gamma_{0s} + 5X_1 + 5X_2 + 5X_3 + 5X_4 + 1X_{5.1} + 1X_{5.2} + 1X_{5.3} + 1X_{5.4} - 50Y$$

The intercept γ_0 is used to control the response rate. Since the response rate is an increasing function of intercept, one can numerically identify the intercept to get the desired response rate. Based on the response model in (1), one can write $\theta_i = 1/exp(-\boldsymbol{\gamma}\boldsymbol{X}_i)$. Here we loosely use the notation using $\boldsymbol{X}$ to denote the ith row of design matrix in (1) that includes intercept, covariates and a survey variable for the ith person. The specific procedures are:

1. To put all covariates in the response model on the same scale we standardize $\boldsymbol{X}$ to $\boldsymbol{Z}$
2. To control the stochastic aspect of response behavior we scale coefficients γ so that linear predictor has variance of 1.
3. To compute γ_0 with respect to response rate using bisection root finding algorithm.

Based on each of these models and procedures, true response probabilities are generated. Eight response rates are studied that range from 0.2 to 0.9 in 0.1 increment. In summary, true response probabilities are created under 3 models and 3 nonresponse mechanism for each of the 8 response rates, for a total of 72 simulation scenarios.

2.3 Sample Selection

Ten-thousand simple random samples are drawn from each scenario with a sampling rate of 1:10 and the sample-based indicator values are computed and compared to the population true indicator values. Bias and variances are computed. The number of simulation samples is determined by two goals of the simulation. The first goal is to compare the difference in the biases of R- and B-indicators. The second goal is to compare the difference in the variance of the same two indicators. We want to know whether one indicator is better than another with respect to bias and variance.

The difference in the bias of two estimators, $\hat{R}_1$ and $\hat{B}_3$, from their parameters R_1 and B_3 is estimated by the formulas below. In each sample that is selected, both $\hat{R}_1$ and $\hat{B}_3$ are computed. Let θ_1 denote R_1 and θ_2 denote B_3, we have

$$\begin{aligned}\hat{D} &= b(\hat{R}_1) - b(\hat{B}_3)\\ &= \frac{1}{S}\sum_{s=1}^{S}(\hat{\theta}_{1s}-\theta_1) - \frac{1}{S}\sum_{s=1}^{S}(\hat{\theta}_{2s}-\theta_2)\\ &= \frac{1}{S}\sum_{s=1}^{S}(\hat{\theta}_{1s}-\hat{\theta}_{2s}-\theta_1-\theta_2)\end{aligned}$$

By setting the desired CV of the estimated difference in bias to $k = 0.075$, we have

$$\begin{aligned}CV[\hat{D}] &= \frac{v(d)/S}{D} = k = 0.075\\ \Leftrightarrow S &= \frac{v(d)}{(k)^2} = \frac{0.03^2}{(0.075 \times 0.004)^2} = 10,000\end{aligned} \quad (2)$$

where $v(d)$ is the unit variance of the differences, $\hat{\theta}_{1s} - \hat{\theta}_{2s}$, and $D = E(\hat{D})$. Since R- and B-indicators are closely related, their D and $v(d)$ are very small which amounts to ten-thousand of samples to detect a cv of 0.075.

2.4 Estimates

For the purpose of our simulation study, the bias is defined as the differences between the sample response mean and the corresponding true population value. That is, $Bias = E(y_r) - \bar{Y}_U$. Hence, the bias varies for each sample. We show the relbias, $Bias/\bar{Y}_U$, which gives a sense of bias in relation to the population mean.

The indicators studied in this article can be categorized into two groups. The first group, representativeness indicators, including R_1, R_2, R_3, B_1, B_2, and B_3 are survey-level indicators that are proposed to facilitate the assessment of survey quality and the comparison of the fieldwork strategies. The second group, bias indicators, including Q_1, Q_2, H_1, H_2, and W, are proposed as computational tools to aid in the selection of auxiliary variables for weighting purposes. Among them, Q_1, Q_2 are survey-level whereas H_1, H_2, and W are estimate-level indicators.

For brief interpretations, R-indicators measure the departure of the response from MCAR. One yields a lower R-value when the nonresponse is non-ignorable. B-indicators measure the similarity between respondents and nonrespondents. Similar to the R-indicators, a lower B-value implies non-ignorable nonresponse. Q- and H-indicators help to select weighting variables in bias reduction. Higher values of Q and H indicate superior bias reduction in auxiliary variables. Finally, W-indicator measures confidence band of the unknown bias, hence a smaller value is preferred. The mathematical formulations of these indicators and their possible range are summarized in the table 1 below.

R, ranged from 0 to 1, was designed to measure the departure of the response from missing completely at random. We expect a lower R value when the nonresponse is non-ignorable. B, similar to R sitting on a unity scale, measures the similarity between respondents and nonrespondents. A higher value of B indicates a balanced response set as comparing to a lower B value. Hence, we expect a lower B value under non-ignorable nonresponse. W, Q, and H, developed as bias indicators on auxiliary variable selection for weighting purpose, were shown to have ability of selecting effective auxiliary sets in reducing bias. Auxiliary variables that obtain higher values of Q and H are superior in bias-reduction than those of lower indicator values. On the other hand, lower W is preferable than higher W since it measures the confidence band of the unknown bias.

2.4.1 Notation

Unless otherwise noted, we distinguish stochastic variables from their realizations by using upper-case and lower-case letters. We focus on three groups of variables: first, the survey question of interest, denoted by Y; second, the auxiliary variables and paradata that can be linked to individual subject, denoted by X; and third, the response indicator, I, where $I = 1$ if responded, otherwise $I = 0$.

The following table shows the notation used in the remainder of the discussion.

r	responding set
s	full initial sample
Y	survey variable
β	ordinary least square estimate of population coefficient
$\boldsymbol{x}$	auxiliary vector
π	inclusion probability
d	design weight $= 1/\pi$
I	response indicator, where $I = 1$ if responded, otherwise $I = 0$
θ	unknown response propensity
ϕ	unknown response influence; $= 1/\theta$
P	unweighted survey response rate
P_d	design-weighted response rate, $= \sum_r d_k / \sum_s d_k$

In addition, we adopt Särndal and Lundström's notation. In symbols with two indices separated by a semicolon, such as $\boldsymbol{x}_{r;d}$, the first index, r, shows the set of units over which the quantity is defined, and the second, d, shows the weighting. In symbols with three indices and a bar (|), such as $\boldsymbol{B}_{\boldsymbol{x}|r;d}$, the first index, $\boldsymbol{x}$, shows the estimate B was computed upon, and the second, r, and the third, d are the set of units and the weighting as mentioned above.

Table 1: Candidate indicators, formula, range, and y-dependency

Indicator Name		Formula	Range	Y-dependent
R_1	=	$1 - 2S_{\hat{\theta}\|s}$	0 to 1	no
R_2	=	$1 - 4S^2_{\hat{\theta}\|s}$	0 to 1	no
R_3	=	$1 - \Re^2$	0 to 1	no
B_1	=	$1 - S^2_{\hat{\theta}\|s;d}/P_d(1-P_d)$	0 to 1	no
B_2	=	$1 - 4S^2_{\hat{\theta}\|s;d}$	0 to 1	no
B_3	=	$1 - 2S_{\hat{\theta}\|s;d}$	0 to 1	no
Q_1	=	$S^2_{\hat{\phi}\|r;d}$	no limit	no
Q_2	=	$CV(\hat{\phi})$	no limit	no
H_1	=	$Q_2 \times R_{y,\boldsymbol{x}} \times \|R_{D,C}\|$	no limit	yes
H_2	=	$Q_2 \times R_{y,\boldsymbol{x}}$	no limit	yes
W	=	$\sqrt{1 - corr^2(\beta' X, Y)}\sqrt{1 - corr^2(\beta' X, I)}$	0 to 1	yes

3. Results

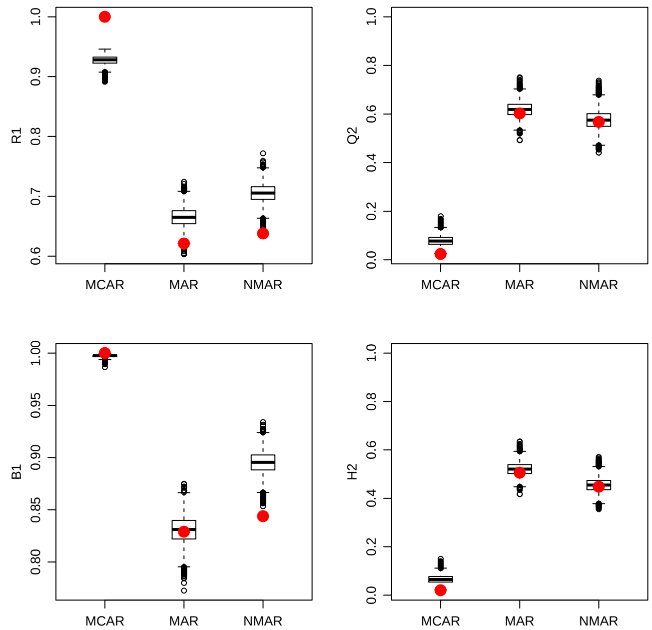
3.1 Sampling properties

Figure 1 illustrates selected sample indicators under three nonresponse mechanisms with a fixed response rate of 0.5. Each box-plot is a distribution of $10,000$ samples of size $1 : 10$ (that is, n=$2,885$). The red dots indicate the true population values for the corresponding indicators.

All indicators, except for Rs, closely correspond to the true population values under three missing data mechanism. Under MCAR, the true population R values are 1 for R_1 and R_2. The true population R_3 depends on the true response probability (θ) where the maximum true value is close to 1 when $\theta = 0.5$. Although R indicators slightly underestimate the true value under MCAR and consistently overestimate the true value under MAR and MNAR, they do differentiate various non-response mechanisms. This is not surprising, since R-indicators measure the departure from MCAR. Under MCAR, R-indicators give a value close to 1, whereas their values are lower than 0.8 under MAR and MNAR.

B-indicators very closely correspond to their true population values. In general these indicators have higher values higher than those of R-indicators. Similar to R-indicators, B-indicators very well differentiate various nonresponse mechanisms. As shown by the simulation results, under MCAR B-indicators display very small variation compared to the results obtained under MAR and MNAR.

Under the MCAR nonresponse mechanism and "correct model", Q-indicators closely correspond to their true population value except when response rate is 20%. Under MCAR, Q_1 is almost unbiased, especially with a response rate of 50% or higher. However, Q_2 tends to very slightly overestimate the true population value. On the other hand, both Q_1 and Q_2 are unbiased under MAR and MNAR invariant to the response rate. The variance of the Q-indicators decreases with increasing response rate.



(a) Response Rate = 0.5

Figure 1: Indicators and their population true value

The relationship between sample-based H-indicators and population H-indicators under MCAR is similar to that of the Q-indicators. H_1 and H_2 have similar sampling variation. Variation in these indicators decrease very quickly with increasing response rate. Under MCAR where there is no bias, the H_1 indicator closely reflects the true population value, even when the response rate is as low as 0.3. On the other hand, H_2 always slightly overestimates the true population value under MCAR.

Sampling variation among all indicators is small (mostly less than 0.05) and does not vary significantly across different response rates. R- and B-indicators have similar sampling variation, whereas Q- and H-indicators have similar sampling variation. One exception is Q_1 which has significantly larger variation when the response rate is 30% or less. The sampling variation for Q_1 under MCAR is larger than 0.15, more than 3-folds than that of other indicators. Under MAR and MNAR, then sampling variation for Q_1 reaches 0.5 when response rate is 30% or less. Similar sampling variation patterns are observed across different missing data mechanism and different bias models.

Sampling variation among all indicators, except for Q_1, increases very slightly (within 0.01 for R and B and within 0.04 for H and Q_2) with increasing bias of sample mean. Again, the variation of Q_1 has very sharp slope closely associated with the increasing bias. Furthermore, the patterns described in this section for all indicators in regard to both bias and variance are the same among three bias models.

In this study, the W-indicator appears to be close to a constant across all simulation scenarios. We look at its theoretical derivation. It is proportional to the product of two correlations. One is the correlation between Y and $\hat{Y}$, and the other is the correlation between Y and I, where I is the response indicator. Since we fixed the R^2 to be 0.7. This might explain why W does not vary much. We will omit it for further discussion in this document.

3.2 The effect of nonresponse mechanism

As mentioned previously, R-indicators very well differentiate various nonresponse mechanisms. Although both R_1 and R_2 reasonable corresponding to the true population value, R_3 in general departed from the true value significantly more than that of R_1 and R_2. This is, in part, attributed to the fact that R_3 has a very different formulation. Similar to R_1 and R_2, R_3 tends to overestimate the true value, with larger scale than R_1 and R_2. One exception where R_3 is almost unbiased to the true value is under MCAR and at the response rate of 0.5. The distribution of R_3 in terms of correspondence to the true value has an inverse U-shape across response rate which is true among all nonresponse mechanism. R_3 vastly overestimates the true value when sample response rate is away from 0.5 in either directions, and is unbiased at sample response rate of 0.5. This pattern is consistent across all nonresponse mechanism. On the other hand, R_1 and R_2 have very similar values under MCAR, their differences increase with similar scale under MAR and MNAR. These differences narrows down with increasing response rate.

B-indicators, as mentioned previously, have very similar profile as the R-indicators. Under MCAR, B-indicators do not vary considerably across different response rates; they range from 0.98 to 0.99, closely corresponding to the population value of 1 under MCAR condition. On the other hand, B-indicators consistently overestimate the population true value under MNAR invariant to the response rates. The variation of B-indicators are small across different nonresponse mechanism.

Under MCAR, Q_1 slightly overestimates the true population value in low response rate. With response rate of 0.4 or higher, Q-indicators are closely correspond to the true value under all nonresponse mechanism. Q-indicators are almost unbiased under MAR and MNAR regardless of the response rate. Q_2 in general is closer to the true value than Q_1, especially when response rates are lower than 0.7.

For H-indicators, H_1 is closer to its true value compared to H_2 under MCAR. Under MCAR, the deviation between H-indicators and their corresponding true values diminishes with increasing response rate. On the other hand, H-indicators are almost unbiased under MAR and MNAR regardless of the response rate. Similar to Q-indicators, H-indicators are not sensitive to different nonresponse mechanisms. The values of H-indicators are very similar between different nonresponse mechanism.

3.3 The effect of response rate

The value of R_1 and R_2 have a slight U-shape curvature where the lowest point is at a response rate of 0.5. The value of R_3 has an inverse U-shape that peaks at a response rate of 0.5. Similarly, the shape of B-indicators across different response rates is strikingly similar to that of R_1 and R_2. They remain almost constant across response rates under MCAR and have a slight U-shape curvature across response rate under MAR and MNAR. For both Q and H, the bias-reduction indicators, although in different magnitudes as compared to the R and B-indicators, their values decrease with increasing response rates. This is intuitive since bias decreases with increase response rates. Q and H are bias-indicators and hence decrease when the response rate is high. Figure 2 illustrates the profile of each indicator across response rates under MCAR and model 3. Figure 3 illustrates the profile of each indicator under MAR and model 3. To aid the visual comparison all graphs are on the same scale and therefore Q_1 is truncated for values large than 1.

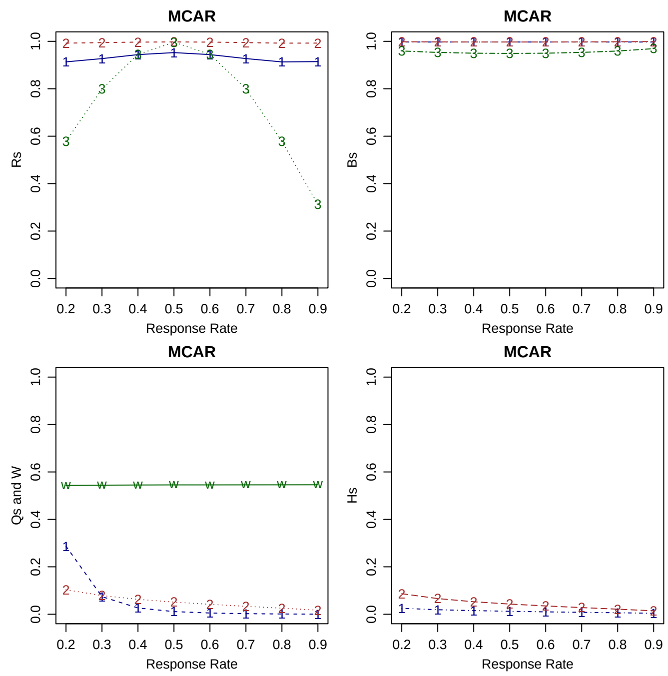


Figure 2: Indicators by response rate under MCAR

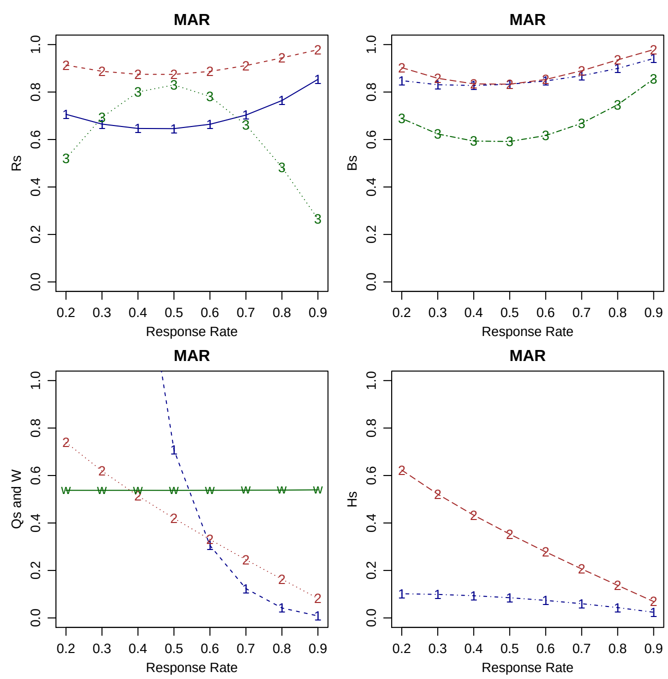


Figure 3: Indicators by response rate under MAR

3.4 The bias of the estimated mean

Recall that the $Bias = E(y_r) - \bar{Y}_U$. Figure 4 illustrates the relationship between sample indicator values and the sample relative bias under MNAR and model 3. Again, to aid the visual comparison all graphs are on the same scale and therefore Q_1 is truncated for values large than 1. One notices the similarity between figures 3 and 4. This is because bias and response are proportional to each other.

Both representativeness indicators and bias-reduction indicators closely correspond to the bias. However, the representativeness indicators do not have monotone relationships with the bias. The values of the representativeness indicator decrease in general with increasing bias, however, with a slight upward-curvature when the bias is large. The bias indicator Q and H demonstrate a good association with bias, larger values of the bias indicator corresponding to larger bias in the estimated mean. These patterns are consistent for both MAR and MNAR and for all three bias models.

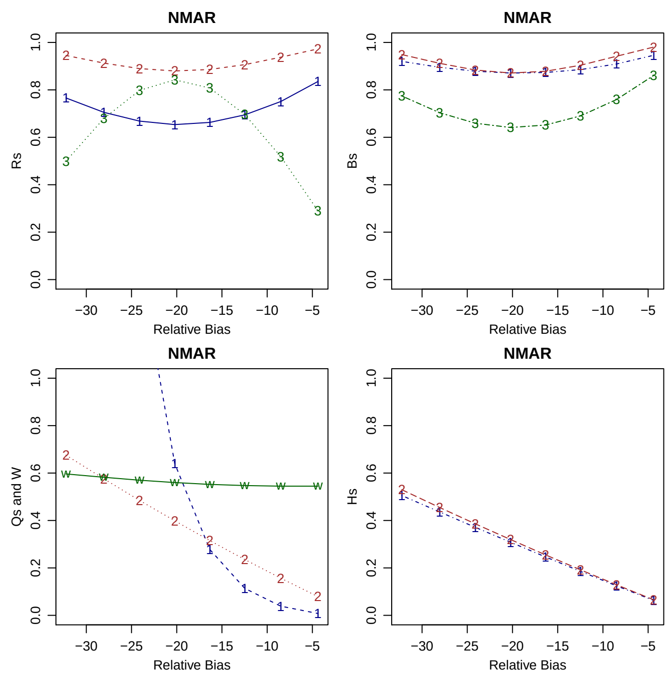


Figure 4: Indicators under various bias magnitude

4. Summary

Under MCAR, there is no bias and the response set is representative. R_1, R_2 and B-indicators have values that are close to 1, whereas bias indicators Q and H have values close to 0. This is expected since in a perfectly representative sample $R = 1$ and $B = 1$. Furthermore, a perfectly representative sample has no bias, hence Q and H are close to zero.

On the other hand, when the nonresponse is non-ignorable, this implies that the respondent set is biased. The values of R_1, R_2 and B-indicators depart from 1 and the bias indicators Q and H have larger values that correspond to the magnitude of bias.

We also observed that the variation of these indicators does not increase significantly with lower response rate and larger bias. One exception is Q_1 which has large sampling variation for a low response rate. Most of these indicators' values are not sensitive to the response rate and bias.

In general, the variation of these indicators decreases with the increasing response rate and decreasing bias. It is expected that the profile of R_1 and R_2 and B-indicators are similar under all investigated scenarios since they have very similar theoretical derivations. Similarly, the profile of H and Q are similar because they originate from the same framework. One exception is R_3, whose value has an inverse U-shape across response rates and is only unbiased at response rate of 0.5 and under MCAR.

Overall, the values of the bias indicators, H and Q, closely correspond to the magnitude of the estimator bias, especially Q_1, Q_2 and H_2. H_1 increases with the increasing value of bias, but with significantly less scale. On the other hand, the representative indicators R and B do not have monotone association with bias, although they do in general decrease (indicating that the respondents are less representative) with increasing bias magnitude.

From the perspective of a representative sample, the representativeness indicators do differentiate between nonresponse mechanisms. Lower values of R and B are observed under MAR and MNAR where R_1 and B_3 demonstrate a larger departure from 1 as compared to R_2, B_1 and B_2. A less representative sample implies a larger bias in the sample estimates. The bias indicators corresponds to bias in a monotone fashion.

In summary, the bias indicators are better indicators for assessing both representativeness and bias. However, in this study we investigated only a single survey variable. Nowadays most surveys are multi-purpose. When more than one survey variable exists, the bias indicators H_1 and H_2 that are Y-dependent cannot be used as survey-level representativeness indicators. Instead, Q-indicators that are Y-independent might be good candidates for evaluating both representativeness and bias. However, the drawback of this approach is that Q-indicators are not bounded which leads to difficulties in interpreting how representative a respondent set is for a Q value of, say, 0.8.

REFERENCES

[R] Schouten, Barry, Fannie Cobben, and Jelke G. Bethlehem. (2009), "Indicators for the representativeness of survey response." *Survey Methodology*, 35, 101–113.

[R] Cobben, Fannie, and Barry Schouten. (2005), "Bias measures for evaluating and facilitating flexible fieldwork strategies," Paper presented at 16th International Workshop on Household Survey Nonresponse, August 28–31, Tällberg, Sweden.

[B] Särndal, Carl-Erik, and Sixten Lundström. (2010), "Design for estimation: Identifying auxiliary vectors to reduce nonresponse bias." *Survey Methodology*, 36, 131–144

[Q] Särndal, Carl-Erik, and Sixten Lundström. (2008), "Assessing auxiliary vectors for control of nonresponse bias in the calibration estimator." *Journal of Official Statistics*, 24(2), 167–191.

[H] Särndal, Carl-Erik. (2011), "The 2010 Morris Hansen Lecture Dealing with Survey Nonresponse in Data Collection, in Estimation." *Journal of Official Statistics*, 27(1), 1–21. with Discussions by M. Brick and R. Tourangeau

[W] Schouten, Barry. (2007), "A selection strategy for weighting variables under a Missing-Not-at-Random response." *Journal of Official Statistics*, 23, 51–68.

Bethlehem, Jelke G. (1988), "Reduction of nonresponse bias through regression estimation." *Journal of Official Statistics*, 4(3), 251–260.

Kalton, Graham, and Ismael Flores-Cervantes. (2003), "Weighting methods." *Journal of Official Statistics*, 19(2), 81–97

Groves, Robert M. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly*, 70 (5): 646–675.

Schouten, Barry, and Fannie Cobben. (2007), "R-indexes for the comparison of different fieldwork strategies and data collection modes." Discussion paper 07002, CBS, Voorburg. Available at http://www.cbs.nl/nl-NL/menu/methoden/research/discussion-papers/archive/2007/default.htm.

Cobben, Fannie, and Barry Schouten. (2007), "An empirical validation of R-indicators." Discussion Paper 08006, Statistics Netherlands, Voorburg, The Netherlands. Available at http://www.risq-project.eu/papers/cobben-schouten-2008-a.pdf

Groves, Robert M., J.Michael Brick, Mick Couper, William Kalsbeek, Brian Harris-Kojetin, Frauke Kreuter, Beth-Ellen Pennell, Trivellore Raghunathan, Tom W. Smith, Roger Tourangeau, Ashley Bowers, Matthew Janes, Courtney Kennedy, Rachel Levenstein, Kristen Olson, Emilia Peytcheva, Sonja Ziniel, and James Wagner. (2008), "Issues Facing the Field: Alternative Practical Measures of Representativeness of Survey Respondent Pools." *Survey Practice*, October.

Schouten, Barry, Fannie Cobben, and Jelke G. Bethlehem. (2009), "Indicators for the representativeness of survey response." *Survey Methodology*, 35, 101–113.

Shlomo, Natalie, Chris Skinner, Barry Schouten, Jelke G. Bethlehem, Zhang, Li-Chun. (2009), "Statistical Properties of R-indicators." Work package 3, Deliverable 2.1, Version 2, RISQ project, 7^{th} Framework Programme (FP7) of the European Union, available at http://www.risq-project.eu/papers/RISQ-Deliverable-2-1-V2.pdf

Shlomo, Natalie., Chris Skinner, Barry Schouten, Vincent de Heij, Jelke G. Bethlehem, Pim Ouwehand. (2009), "Indicators for Representative Response Based on Population Totals." Work Package 3, Deliverable 2.2, RISQ project, 7^{th} Framework Programme (FP7) of the European Union, available at http://www.risq-project.eu/papers/RISQ-Deliverable-2-2-V1.pdf

Skinner, Chris, Natalie Shlomo, Barry Schouten, Li-Chun Zhang, Jelke G. Bethlehem. (2009), "Measuring Survey Quality Through Representativeness Indicators using Sample and Population Based Information," RISQ project, Paper presented at the NTTS Conference, 18-20 February 2009, Brussels, Belgium

Schouten, Barry, Mattijn Morren, Jelke G. Bethlehem, Natalie Shlomo, Chris Skinner. (2009), "How to use R-indicators?," Work Package 4, Deliverable 3, RISQ project, 7^{th} Framework Programme (FP7) of the European Union, available at http://www.risq-project.eu/papers/RISQ-Deliverable-3.pdf

Shlomo, Natalie, Barry Schouten, Chris Skinner. (2010), "Partial Indicators for Representative Response," Work package 5, Deliverable 4, Version 2, RISQ project, 7^{th} Framework Programme (FP7) of the European Union, available at http://www.risq-project.eu/papers/RISQ-Deliverable-4-V2.pdf

Schouten, Barry, Natalie Shlomo, Chris Skinner. (2011), "Indicators for Monitoring and Improving Representativeness of Response." *Journal of Official Statistics*, 27(2), 231–253