Alternative Indicators for the Risk of Nonresponse Bias: A Simulation Study

Raphael Nishimura[1], James Wagner[1], Michael R. Elliott[1,2]
[1]University of Michigan, Survey Methodology Program, Institute for Social Research, 426 Thompson St. Room 4050, Ann Arbor, MI 48104
[2]University of Michigan, Department of Biostatistics, School of Public Health, 1420 Washington Heights, Ann Arbor, MI 48109

Abstract
The growth of nonresponse rates for social science surveys has led to increased concern about the risk of nonresponse bias. Unfortunately, the nonresponse rate is a poor indicator of when nonresponse bias is likely to occur. We consider a set of alternative indicators - including the Fraction of Missing Information, R-Indicators, the coefficient of variation of subgroup response rates, and model fit statistics such as R-squared, pseudo R-squared, and the area under an ROC curve. A simulation study is used to explore how each of these indicators performs under a variety of circumstances. The simulations vary the missing data mechanism (MCAR, MAR, and NMAR), the strength of covariates in predicting response and survey outcome variables, and the impact of the misspecification of models. Finally, we discuss how these indicators can be used when creating a plausible account of the risks for nonresponse bias for a survey.

Key Words: Missing data, Nonresponse bias, Nonresponse indicators, Survey data quality measures

1. Introduction

It is well known that nonresponse is a major threat against quality of survey data. However, it is a source of error difficult to assess. This is especially true when no special studies are conducted to detect the risk of nonresponse bias, which is the case for most surveys in practice. Instead, other indirect methods are used. In that regard, in the absence of any other guidance, response rates have been widely used as a key measure for risk of nonresponse bias and survey data quality (Biemer and Lyberg, 2005).

The assumption underlying this practice is that a higher response rate indicates a lower risk of nonresponse bias. Because of this, the declining response rates over the last decades in surveys (de Leeuw and de Heer, 2002) have been leading increasing concern to practitioners about the risk of nonresponse bias. However, statistical theory shows that this underlying assumption is, at most, just partially true, since nonresponse bias is a function of both response rate (or response propensity) and the differences between respondents and nonrespondents in terms of the survey variable (or the association between the response propensity and the survey variable). In fact, a *reduction* on the response rate may lead to an *increase* in the nonresponse bias. If, for instance, the difference between respondents and nonrespondents is small, and an modification in the survey design to increase the response rate, such as providing incentives, enlarges this difference by bringing in only one sort of respondent, then an increase of response rate could potentially increase nonresponse bias.

Moreover, several recent empirical studies aimed to evaluate the response rate as a proxy measure for the risk of nonresponse bias have shown that higher response rates have not led to lower nonresponse bias (Keeter, et al., 2000; Curtin and Singer, 2000; Merkle and Edelman, 2002; Keeter, et al., 2006). Groves and Peytcheva (2008) summarized in a meta-analysis of 59 specialized nonresponse bias studies much of the empirical evidence in that regard. They found little association between nonresponse rates and nonresponse bias across many of the over 300 statistics produced by these studies.

These results give evidences that using response rates as the only tool for monitoring data collection or post-survey adjustments might be inefficient, bias the survey estimates or both. Therefore, more recently survey methodologists have been attempting to find new measures for the risk of nonresponse bias. However, very little research has been done regarding the usefulness of these alternative measures and under which conditions they may prove misleading. Moreover, one important question that survey practitioners often ask is whether a single or a set of these indicators can reliably indicate the risk of nonresponse bias.

This simulation study investigates the performance of a set of alternatives measures for the risk of nonresponse bias suggested in the literature under various conditions. After this brief introduction in section 1, the alternatives measures studied in this paper are presented in section 2. Then, details about the simulation settings are given in section 3, followed by its results in section 4. Finally, we discuss the results of these simulations and its practical implications in section 5.

2. Alternative measures for the risk of nonresponse bias

Here we define the alternative measures for the risk of nonresponse bias under investigation in this simulation study. Some of them are often used in practice in many surveys. Others, on the other hand, are still not much used, but they have been advocated by survey methodologists in the literature as strong candidates of indictors of nonresponse bias.

2.1 Response rate

Despite the evidences against its use to evaluate the risk of nonresponse bias, the response rate is included in this study to verify if there is any condition if it can be proved useful and also as a mean of comparison with the other measures. Assuming a sample of size n, let R_i be the response indicator for the i^{th} element in the sample, that is,

$$R_i = \begin{cases} 1, \text{ if i}^{\text{th}} \text{ element is a respondent} \\ 0, \text{ otherwise} \end{cases}$$

Then, the response rate, RR, is given by

$$RR = \frac{\sum_{i=1}^{n} R_i}{n} = \frac{n_r}{n}$$

2.2 Subgroup response rates

These are commonly used alternatives for the simple response rate. However, Peytcheva and Groves (2009) found that they are rarely predictive of nonresponse bias. The assumption underlying these measures is that differential response rates across different subgroups indicate a larger risk for nonresponse bias. This measure is computed as the response rate of subgroups formed by auxiliary information available for both respondents and nonrespondents, such as frame variables or paradata.

2.3 Coefficient of variation of subgroup response rates

This can be seems as a summary measure of the previous indicator. In that sense, it assumes that large values of coefficient of variation of subgroup response rates, that is, a higher level of differential response rate, indicate a larger risk for nonresponse bias. Let RR_k be the response rate of the k^{th} subgroup from a total of K subgroups, each with n_k elements. The coefficient of variation of subgroup response rates is given by

$$CV\left(RR_{sub}\right)=\frac{\sqrt{\frac{1}{K-1}\sum_{k=1}^{K}\left(RR_k-RR\right)^2}}{RR}$$

2.4 Variance of nonresponse weights

A usual method to deal with nonresponse is by weighting the respondents to compensate the nonrespondents within subgroups, also called nonresponse adjustment cells, formed by auxiliary information in the same fashion that is done for the two previous measures. For a given nonresponse adjustment cell, the nonresponse weight is the inverse of the response rate on that cell, that is, $w_{nr,k} = (RR_k)^{-1}$. Therefore, just as the previous measures, a higher level of variability of these weight could indicate a larger risk for nonresponse bias. Assuming a sample with n_r and K nonresponse adjustment cells, each with n_k respondents, the variance of the nonresponse weights is given by

$$Var\left(W_{nr}\right)=\frac{1}{n_r-1}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\left(w_{nr,k_i}-\bar{w}_{nr}\right)^2$$

where $\bar{w}_{nr}=\frac{1}{n_r}\sum_{k=1}^{K}\sum_{i=1}^{n_k}w_{nr,k_i}$.

2.5 Variance of poststratification weights

Another weighting method for nonresponse adjustment is poststratification. In this approach, the population sizes of nonresponse adjustment cells, N_k, are assumed to be known and the poststratification weight is given by $w_{ps,k} = N_k/n_k$, $k = 1,..., K$, where n_k is the number of respondents in the k^{th} adjustment cell. Although the relation is not as obvious as it is with the nonresponse weights, it is also assumed that the more variable these weight are, the larger is the risk for nonresponse bias. Hence, another indicator is the variance of such weights, given by

$$Var\left(W_{ps}\right)=\frac{1}{n-1}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\left(w_{ps,k_i}-\bar{w}_{ps}\right)^2$$

where $\bar{w}_{ps} = \frac{1}{n_r}\sum_{k=1}^{K}\sum_{i=1}^{n_k} w_{ps,k_i}$.

2.6 Correlation between nonresponse weights and survey variable

Little and Vartivarian (2005) showed that the effectiveness of a nonresponse adjustment depends both on the associations of the survey variable with the auxiliary variables used in the adjustments and also with the response propensities. The correlation between the nonresponse weights and the survey variable can be used as a proxy of the former. Let *Y* be the survey variable, then this indicator is given by

$$Corr\left(W_{nr}, Y\right) = \frac{\sum_{k=1}^{K}\sum_{i=1}^{n_k}\left(w_{nr,k_i} - \bar{w}_{nr}\right)\left(y_{k_i} - \bar{y}\right)}{\sum_{k=1}^{K}\sum_{i=1}^{n_k}\left(w_{nr,k_i} - \bar{w}_{nr}\right)^2 \sum_{k=1}^{K}\sum_{i=1}^{n_k}\left(y_{k_i} - \bar{y}\right)^2}$$

2.7 Area Under the Curve (AUC) of logistic regression predicting response propensity

The response rate or the subgroup response rates may be seem as crude estimates of the response propensity, that is, the probability of a selected element respond the survey request. Hence, a generalization of the nonresponse weights mentioned above is to use models for a binary outcome, such as a logistic regression, in order to predict the response propensity of the respondents, given a set of covariates. The inverse of such predicted response propensities could then be used as nonresponse weights. In that sense, the predictive power of the logistic regression model could also be used as an indicator for nonresponse bias. The Area Under the Curve (AUC), which is also the *C* statistic for binary outcomes, is one of such measures. It ranges from 0.5 to 1. Higher values of AUC indicate a better predictive model and, therefore, a higher risk for nonresponse bias.

2.8 Pseudo-R^2 (Nagelkerke) of logistic regression predicting response propensity

Similar to the AUC, the pseudo-R^2 could be seem as another measure of the predictive power of a logistic regression model for the response propensity. It varies between 0 and 1 and high values of the pseudo-R^2 also indicate a high predictive model.

2.9 R-Indicator

Schouten et al (2009) suggested using the variability of the predicted response propensities as a measure of survey quality. The underlying idea is that if the predicted response propensities don’t vary much, the association between response and characteristics that distinguish respondent and nonrespondents is low and, therefore, there is a low risk for nonresponse bias. This would be equivalent to a Missing Completely At Random (MCAR) or a Missing At Random (MAR) mechanism. On the other hand, if there is much variability among the predicted response propensities, then the association between responses and variables that discriminate respondents from nonrespondents might be high, which, in turn, might indicate a high risk for nonresponse bias. With these ideas, Schouten et al (2009) proposed the R-Indicator:

$$R\left(\rho\right) = 1 - 2S\left(\rho\right)$$

where $S(\rho)$ is the standard deviation of the response propensities. Therefore, it varies between 0 and 1, with higher values indicating a larger risk for nonresponse bias.

2.10 Fraction of Missing Information (FMI)

The concept of the Fraction of Missing Information (FMI) was developed in the missing data and multiple imputation literature (Dempster et al, 1977; Rubin 1987) as a measure of uncertainty about the values imputed for missing elements. More precisely, it is the proportion of the total variance of a survey estimate explained by the between-imputation variability. Wagner (2010) suggested using the FMI as a survey quality measure to monitor data collection. The underlying idea is that if the FMI is large, it means that there is much uncertainty on the imputed values of nonrespondents and, therefore, this indicates a large risk for nonresponse bias. The most straightforward method to estimate FMI is multiple imputing, say M times, the missing data for the nonrespondents under a model, estimating for each multiple imputed dataset the parameter θ by $\hat{\theta}_m$ and estimate the FMI by

$$\hat{\gamma} = \frac{\left(1+\frac{1}{M}\right)Var_B\left(\hat{\theta}\right)}{Var\left(\hat{\theta}\right)}$$

where $Var_B\left(\hat{\theta}\right)=\sum_{m=1}^{M}\left(\theta_m-\overline{\theta}_M\right)^2 \Big/ \left(M-1\right)$ is the between-imputation variance, $\overline{\theta}_M=\sum_{m=1}^{M}\hat{\theta}\Big/M$ is the average of the estimates using the M fully-imputed datasets, $Var\left(\hat{\theta}\right)=Var_W\left(\hat{\theta}\right)+\left(M-1\right)M^{-1}Var_B\left(\hat{\theta}\right)$ is the total variance of the estimate and $Var_W\left(\hat{\theta}\right)=\sum_{m=1}^{M}Var_m\left(\hat{\theta}\right)\Big/M$ is the within-imputation variance, which is the average of the M estimate's variances $Var_m\left(\hat{\theta}\right)$ computed using the M fully-imputed datasets.

3. Methods

Two simulation studies were conducted, each one using k = 1,000 simple random samples of size n = 1,000 to estimate a population mean $\overline{Y}$ with one observed explanatory variable X and another unobserved Z. In both studies it was varied:

- Missing mechanism
- Response rate
- Correlation between the explanatory and survey variables
- Correlation between the response propensities and the explanatory variables

In the first simulation study, a total of 1,083 different simulations were conducted with:

- 3 missing mechanisms: Missing Completely at Random (MCAR), Missing At Random (MAR) and Missing Not At Random (MNAR)
- 19 response rates varying from 5% to 95% with 5% increments
- 19 correlations between auxiliary variable (*X* or *Z*) and survey variable varying from 5% to 95% with 5% increments

For the MNAR mechanism in this study, only the unobserved variable *Z* was used to generate the missing pattern.

The focus of the second simulation study was the MNAR mechanism. In this case, the missing mechanism was generated using both the observed and unobserved variables *X* and *Z*. It was conducted 243 different simulations using:

- 3 response rates: 20%, 40% and 70%
- 3 correlations between the observed variable *X* and survey variable *Y*: low, medium and high
- 3 correlations between the unobserved variable *Z* and survey variable *Y*: low, medium and high
- 3 correlations between the response propensities and the observed variable X : low, medium and high
- 3 correlations between the response propensities and the unobserved variable Z : low, medium and high

The levels low, medium and high correspond, respectively, to 5%, 20% and 70% in the correlations mentioned above. The only exception is when both the correlations between the response propensities with the variables *X* and *Z* were high that, due to a restriction problem, they were set as approximately 54%, both.

The data and missing mechanism generation was done in the same way for both simulation studies. First, a sample of size n = 1,000 of a random vector (*Y*, *X*, *Z*) was generated with

$$\begin{pmatrix} Y_i \\ X_i \\ Z_i \end{pmatrix} \sim N_3\left(\begin{pmatrix} 100 \\ 10 \\ 10 \end{pmatrix}, \begin{pmatrix} 25 & \sigma_{yx} & \sigma_{yz} \\ \sigma_{xy} & 4 & 0 \\ \sigma_{zy} & 0 & 4 \end{pmatrix}\right), i = 1,\ldots,1000$$

The covariances σ_{yx} and σ_{yz} vary accordingly to the correlation levels stated above. Then, for each one of the 1,000 elements, it was computed a response propensity, p_i, using a logistic regression model, given by

$$\operatorname{logit}(p_i) = \beta_0 + \beta_1 x_i + \beta_2 z_i, i = 1,\ldots,1000$$

In the first simulation study, the coefficients β_0 , β_1 and β_2 are set accordingly to the response rates and the missing mechanism, as shown in the Table 1 below. In the second simulation study, the values of β_1 and β_2 varied according to the correlations between the response propensities and the observed and unobserved variables, *X* and *Z*; while the coefficient β_0 was set to adjust the overall response rate to 40%. These values are show in Table 2.

Table 1: Values of β_0, β_1 and β_2 for the first simulation study

RR	MCAR			MAR			MNAR		
	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
0.05	-2.94	0.00	0.00	-27.18	2.00	0.00	-27.18	0.00	2.00
0.10	-2.20	0.00	0.00	-25.64	2.00	0.00	-25.64	0.00	2.00
0.15	-1.73	0.00	0.00	-24.52	2.00	0.00	-24.52	0.00	2.00
0.20	-1.39	0.00	0.00	-23.69	2.00	0.00	-23.69	0.00	2.00
0.25	-1.10	0.00	0.00	-22.93	2.00	0.00	-22.93	0.00	2.00
0.30	-0.85	0.00	0.00	-22.25	2.00	0.00	-22.25	0.00	2.00
0.35	-0.62	0.00	0.00	-21.68	2.00	0.00	-21.68	0.00	2.00
0.40	-0.41	0.00	0.00	-21.06	2.00	0.00	-21.06	0.00	2.00
0.45	-0.20	0.00	0.00	-20.55	2.00	0.00	-20.55	0.00	2.00
0.50	0.00	0.00	0.00	-20.00	2.00	0.00	-20.00	0.00	2.00
0.55	0.20	0.00	0.00	-19.45	2.00	0.00	-19.45	0.00	2.00
0.60	0.41	0.00	0.00	-18.90	2.00	0.00	-18.90	0.00	2.00
0.65	0.62	0.00	0.00	-18.32	2.00	0.00	-18.32	0.00	2.00
0.70	0.85	0.00	0.00	-17.74	2.00	0.00	-17.74	0.00	2.00
0.75	1.10	0.00	0.00	-17.08	2.00	0.00	-17.08	0.00	2.00
0.80	1.39	0.00	0.00	-16.28	2.00	0.00	-16.28	0.00	2.00
0.85	1.73	0.00	0.00	-15.50	2.00	0.00	-15.50	0.00	2.00
0.90	2.20	0.00	0.00	-14.34	2.00	0.00	-14.34	0.00	2.00
0.95	2.94	0.00	0.00	-12.86	2.00	0.00	-12.86	0.00	2.00

For each one of the 1000 elements a random number u_i ~ Uniform[0,1] was generated and if $u_i < p_i$, then that element was classified as respondent. Otherwise, it was treated as a nonrespondent, that is, its value for the survey variable *Y* was treated as missing.

The subgroups used to compute the coefficient of variation of the subgroups response rates were formed by using the quintiles of the observed variable *X* as cut-off points. For the FMI indicator, it was used the multivariate imputation by chained equation model considering only the observed variable *X* as covariate and *M* = 10 multiple imputations. The simulations and analysis were performed in `R 2.13.2` with `survey`, `mice` and `Design` packages.

Table 2: Values of β_0, β_1 and β_2 for the second simulation study

Corr(*X*,*R*)	*Corr*(*Z*,*R*)	β_0	β_1	β_2
Low	Low	-1.40	0.05	0.05
	Medium	-3.20	0.06	0.22

	High	-21.35	0.13	1.90
Medium	Low	-3.20	0.22	0.06
	Medium	-5.05	0.23	0.23
	High	-30.90	0.64	2.32
High	Low	-21.35	0.13	1.90
	Medium	-30.90	2.32	0.64
	High*	-64.20	3.10	3.10

* In this case the high level of $Corr(X,R)$ and $Corr(Z,R)$ is High ≈ 0.54.

4. Results

4.1 Simulation Study I

The results of each indicator are shown in the graphs below. In this study, it was analyzed the behaviour of the indicators as functions of the response rates and the correlation between the observed and outcome variable, X and Y, respectively, under each of the three missing mechanism: MCAR, MAR and MNAR.

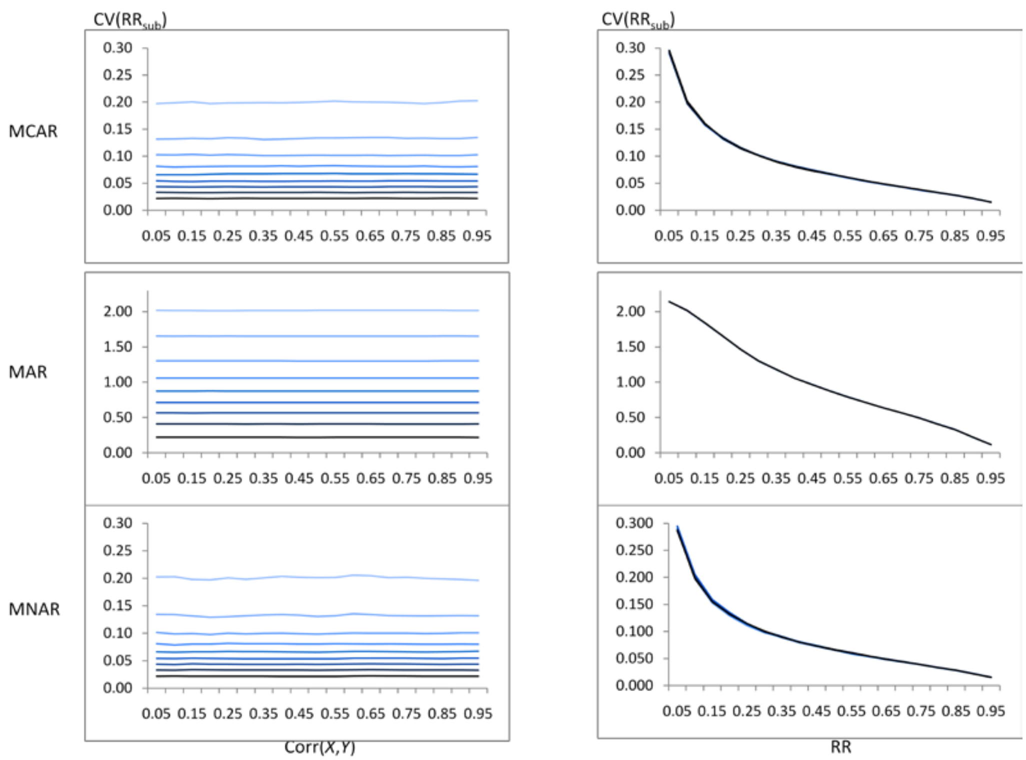

Figure 1: Coefficient of Variation of Subgroup Response Rates

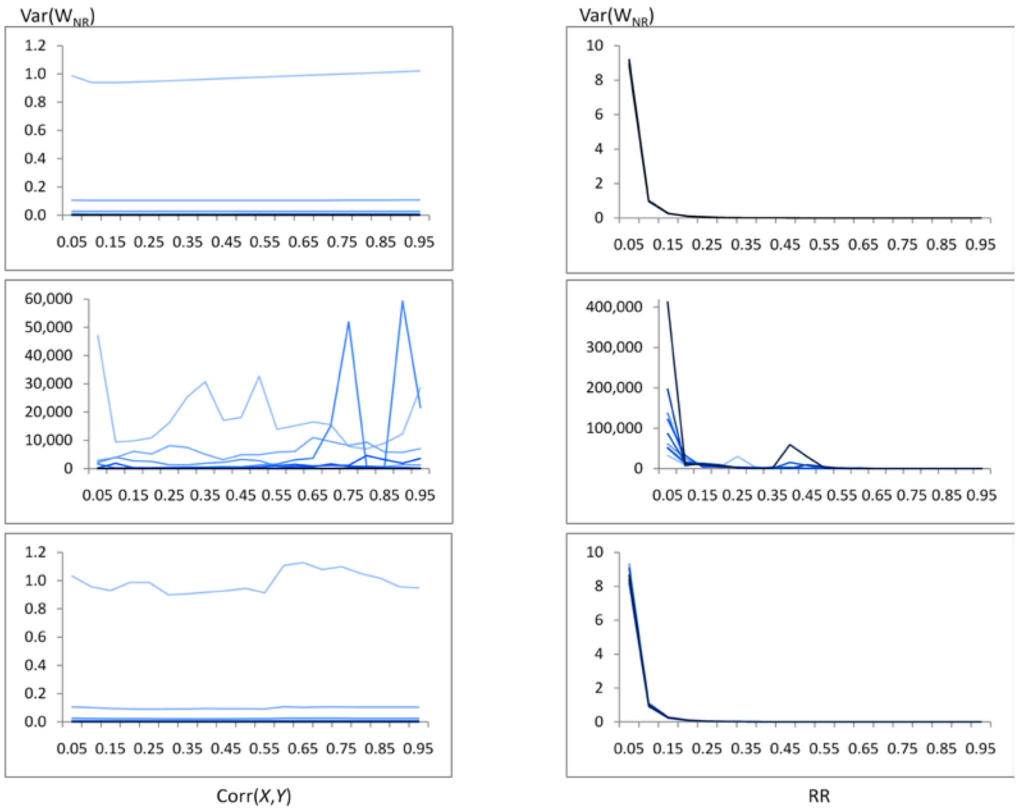

Figure 2: Variance of nonresponse weights

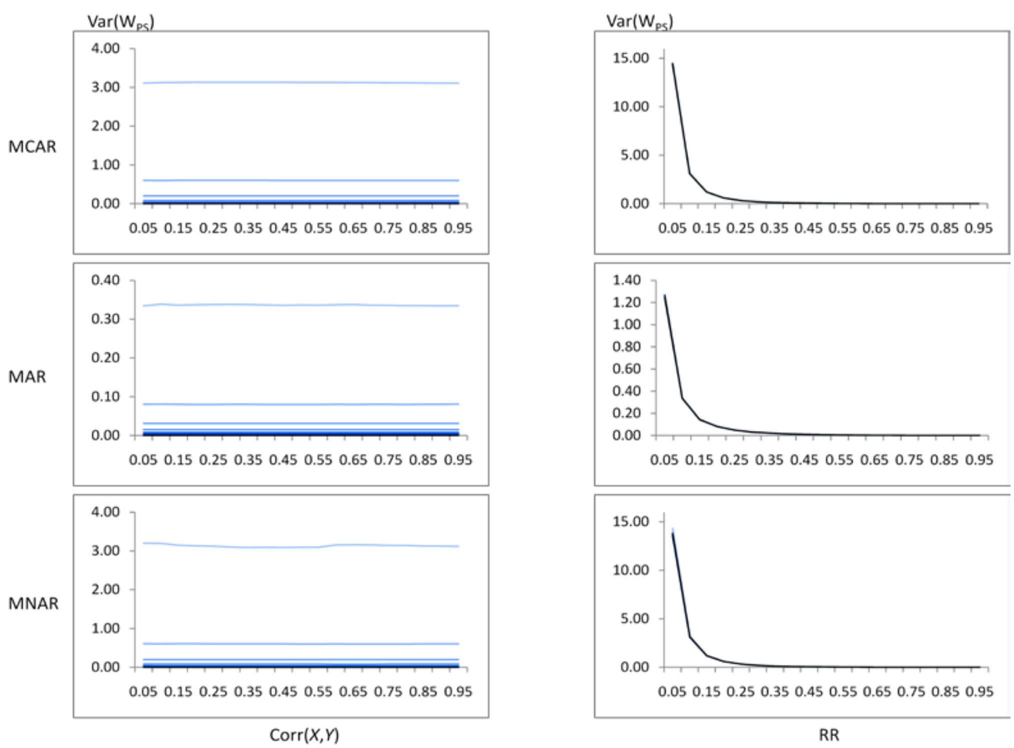

Figure 3: Variance of poststratification weights

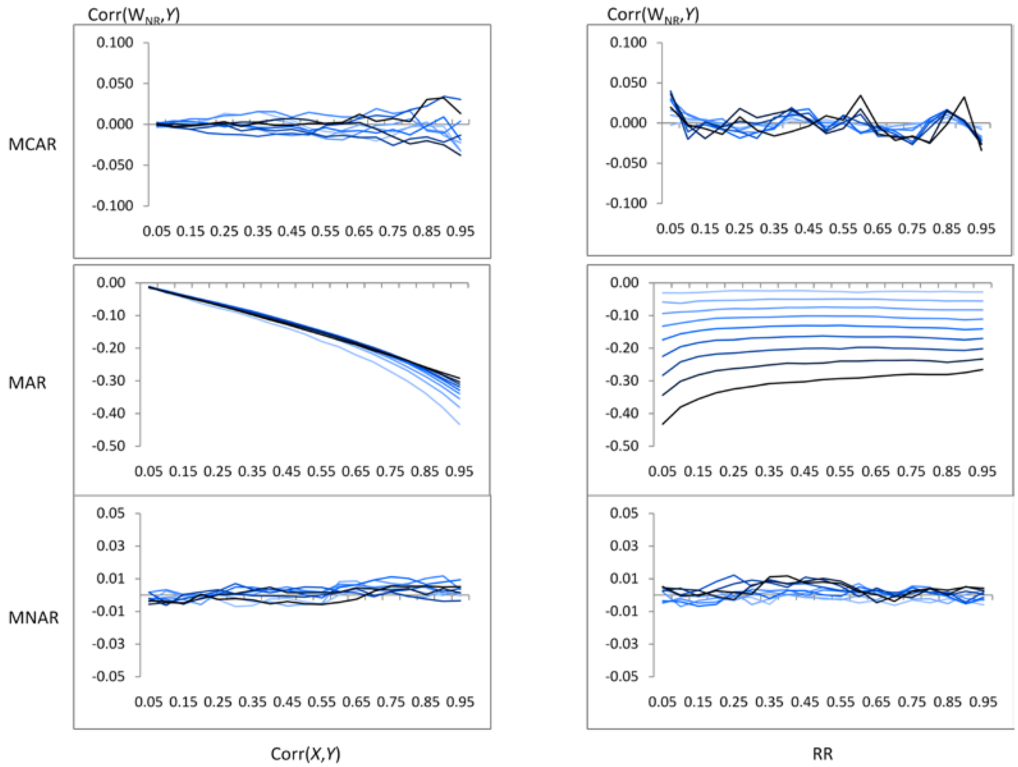

Figure 4: Correlation between nonresponse rates and survey variable

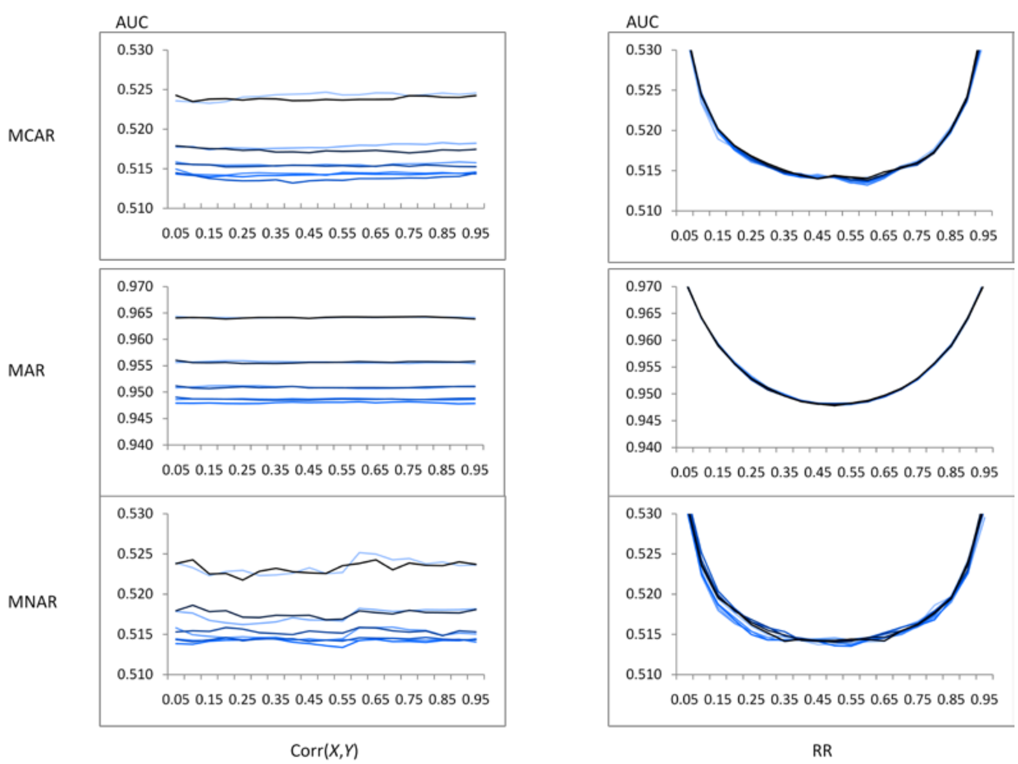

Figure 5: Area Under the Curve of logistic regression prediction response propensity

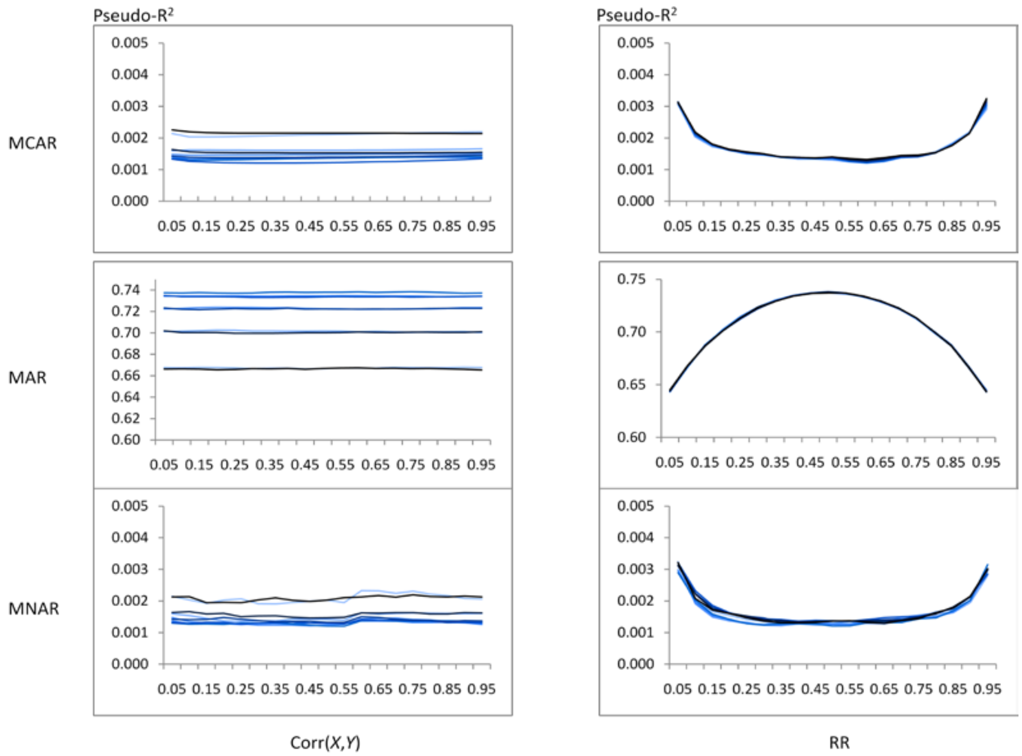

Figure 6: Pseudo-R2 (Nagelkerke) of logistic regression predicting response propensity

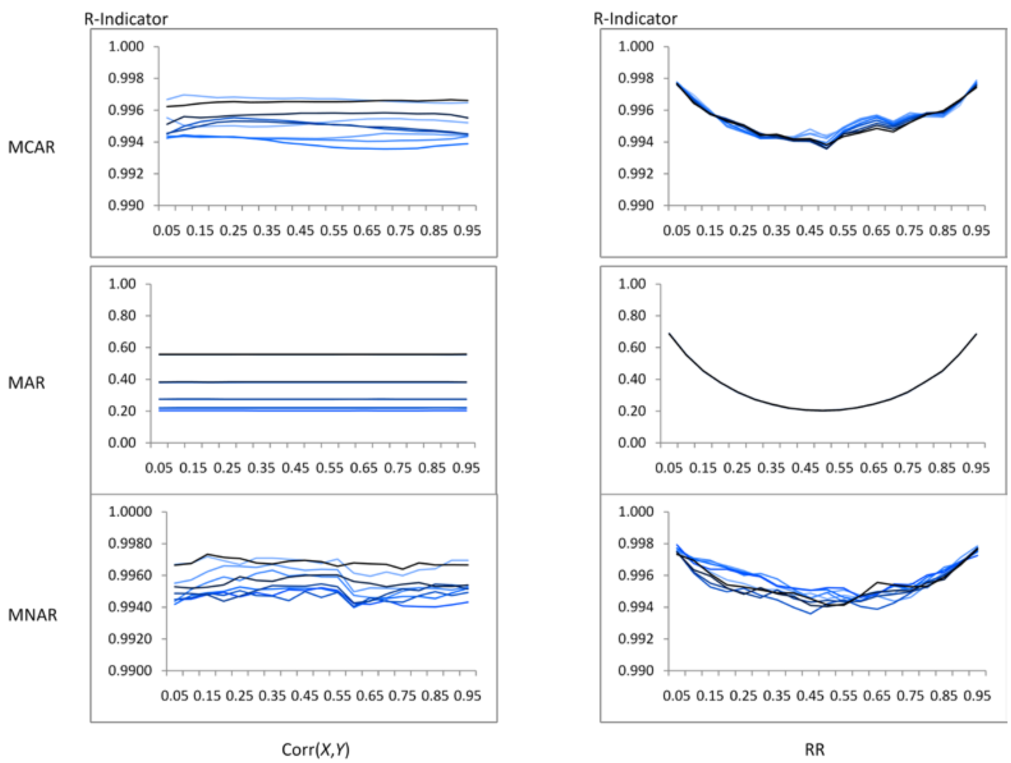

Figure 7: R-Indicator

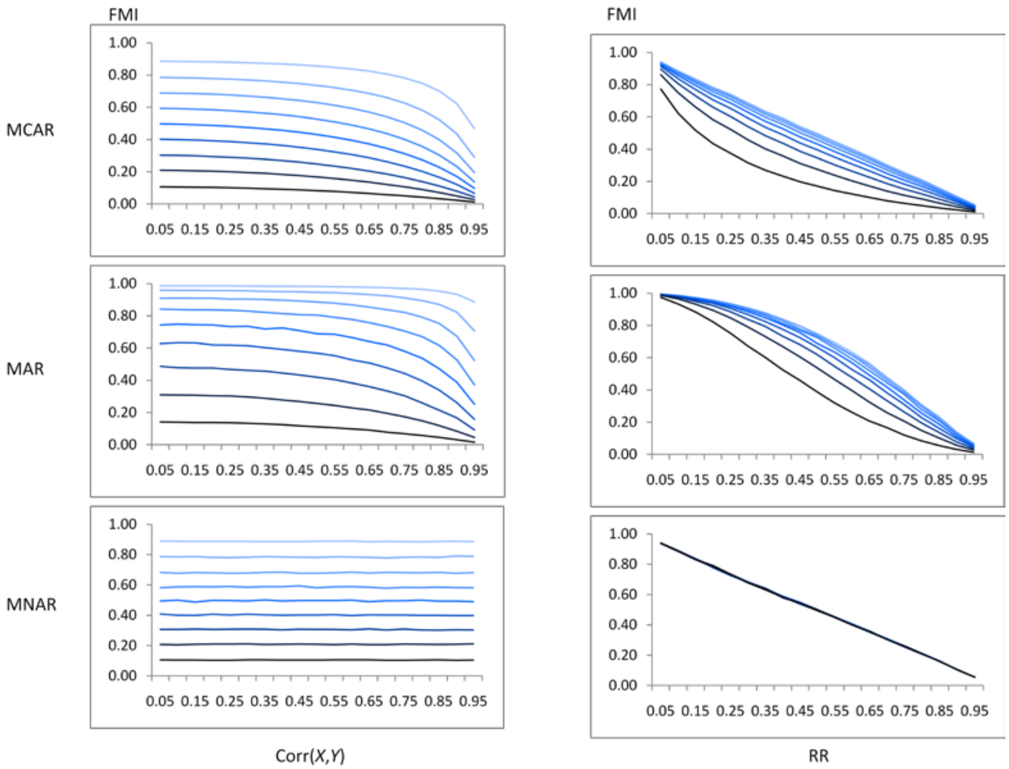

Figure 8: Fraction of Missing Information

4.2 Simulation Study II

In this study, it was analyzed the behaviour of the indicators as functions of the bias of the sample mean of the outcome variable *Y* under a MNAR missing mechanism and three different response rates: 20%, 40% and 70%. The results of each indicator are shown in the graphs below.

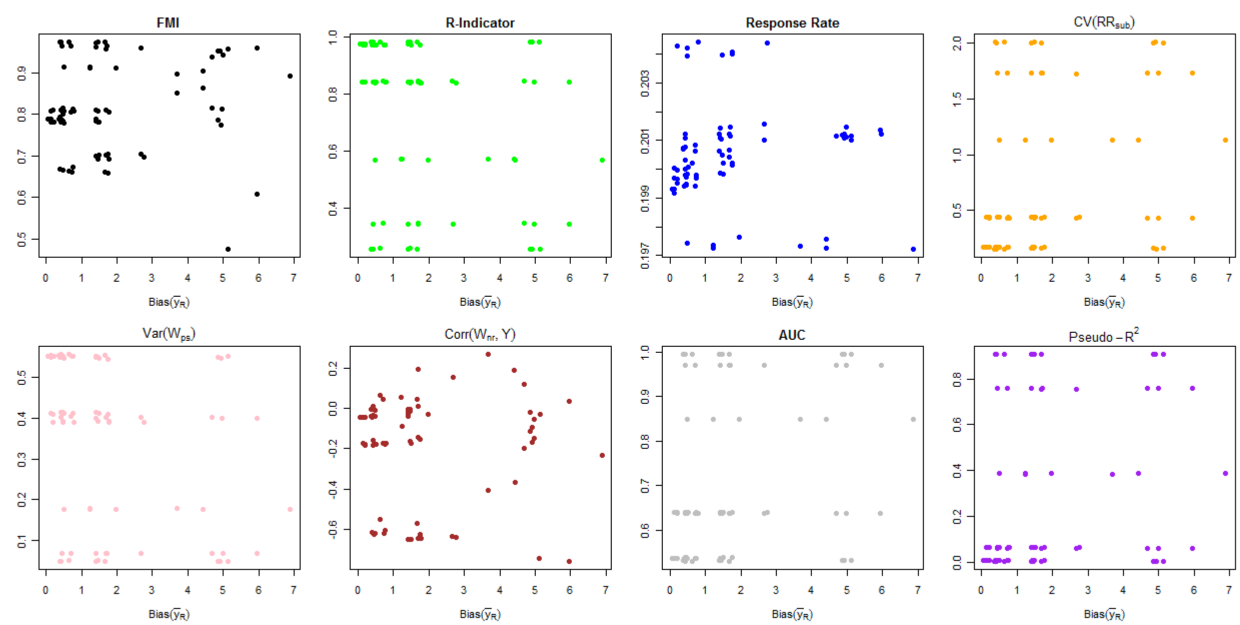

Figure 9: Indicators under MNAR mechanism and a 20% response rate

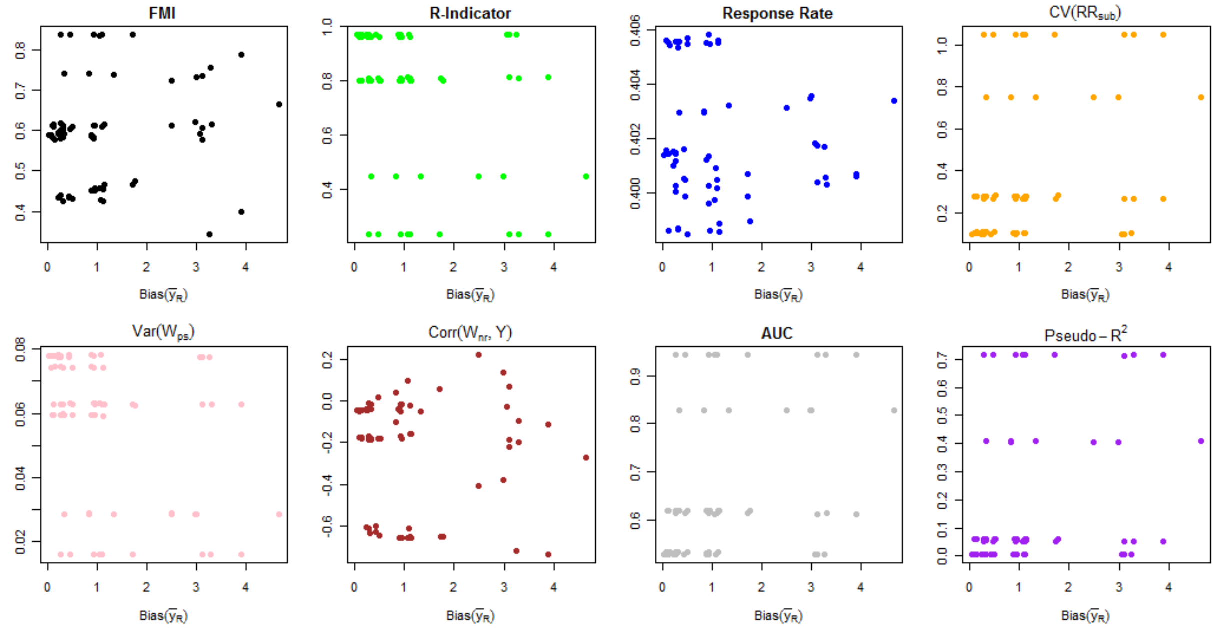

Figure 10: Indicators under MNAR mechanism and a 40% response rate

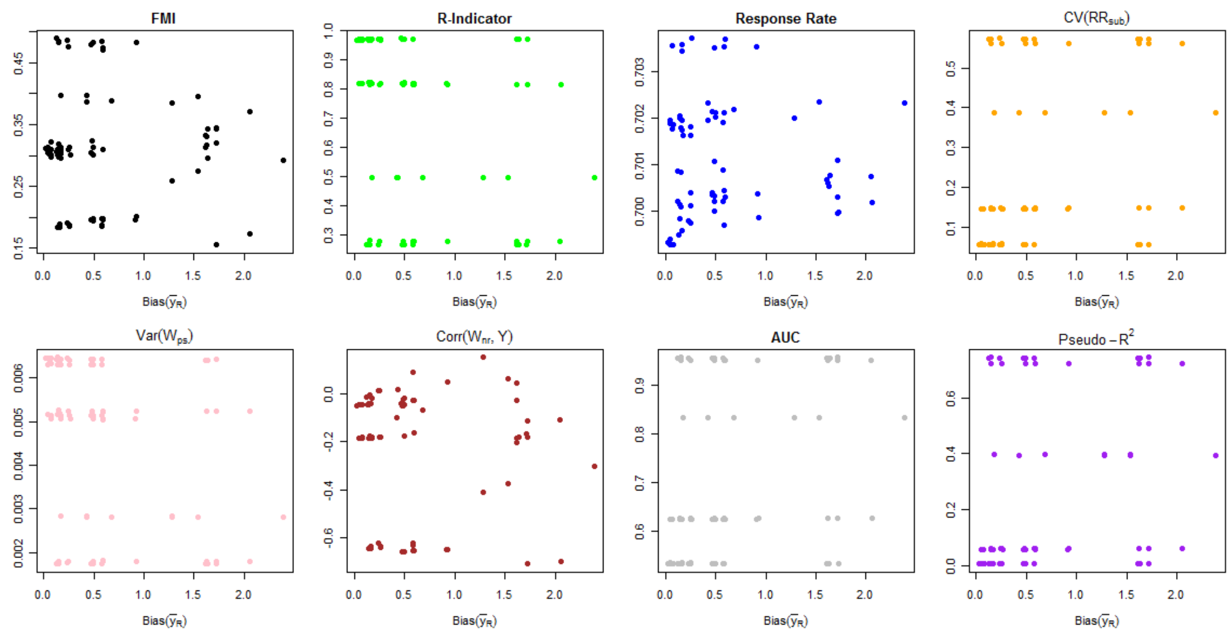

Figure 11: Indicators under MNAR mechanism and a 70% response rate

5. Discussion

It is observed in the first simulation study that most indicators are survey variable-independent, that is, their behaviour is the same no matter which outcome variable is analyzed. Despite this can show some advantages in terms of an indicator for the survey itself, it is well known that nonresponse potentially affect different survey variables in different ways. Therefore, a measure that can capture these differences would also be desirable. In that sense, FMI and the correlation between the nonresponse weights and the survey variable *Y* are the only indicator that show this property.

From the second simulation study, it can be noticed that none of the measure clearly show a pattern that would enable one to detect a risk of nonresponse bias under a MNAR mechanism in any of the response rates studied here. Moreover, it seems that, in general, not even a set of these measures is able to indicate a risk for nonresponse bias. This happens either because (1) there is no association with the indicators and the nonresponse bias or (2) it is not possible to distinguish the missing mechanisms between themselves, especially between MCAR and MNAR, since the indicators present similar behaviours in those situations.

It is important to analyze the same types of scenarios and indicators studied here using an approach that can pick the nonresponse bias more explicitly, using pattern-mixtures models, as it is done in Andridge and Little (2011), for example. This is one of the next steps that will be studies in this research. Another potential area of future work for these alternative indicators for the risk of nonresponse bias is how they can be used to prioritization of nonrespondents for the reduction of nonresponse error during the data collection stage of a survey.

Acknowledgements

This research is sponsored by The Eunice Kennedy Shriver National Institute of Child Health and Human Development, grant number R03HD070012-02

References

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

Biemer, P. P. and Lyberg, L. (2003). *Introduction to Survey Quality*. Hoboken, NJ: Wiley.

De Leeuw, E.D. and De Heer, W. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In: R.M. Groves, D.A. Dillman, J.L. Eltinge, & R.J.A. Little (Eds). *Survey nonresponse*. New York: Wiley, pp. 41-54.

Dempster, A. P., Laird N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B (Methodological).* 39(1):1–38.

Groves, R. M. and Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias – A Meta-Analysis. *Public Opinion Quarterly*. Vol 72, Issue 2, pp. 167-189.

Peytcheva, E., and Groves, R. M. (2009). Using Variation in Response Rates of Demographic Subgroups as Evidence of Nonresponse Bias in Survey Estimates. *Journal of Official Statistics*, 25(2): 193-201.

Keeter, S., Miller, C., Kohut, A., Groves, R. M. and Presser, S. (2000). Consequences of

Reducing Nonresponse in a Large National Telephone Survey. *Public Opinion Quarterly*, 64, pp. 125-48

Keeter, S., Kennedy, C., Dimock, M., Best, J. and Craighill, P. (2006). Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey. *Public Opinion Quarterly*, 70, pp. 759-779

Little, R. J., and Vartivarian, S. L. (2005). Does Weighting for Nonresponse Increase the Variance of Survey Means? *Survey Methodology*. 31, pp. 161-168.

Merkle, D. M. and Edelman, M. (2002). Nonresponse in Exit Polls: A Comprehensive Analysis. In *Survey Nonresponse*, ed. R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. A. Little, pp. 243-58. New York: Wiley.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley

Schouten, B., Cobben, F. and Bethlehem, J. G.. (2009). Indicators for the Representativeness of Survey Response. *Survey Methodology*. 35(1):101–13.

Wagner, J. (2010). The Fraction of Missing Information as a Monitoring Tool for Survey Data Quality. *Public Opinion Quarterly*. 74(2):223-243.