Effects of Unit Nonresponse on Estimating the Population Mean

Jessica Fox[1], Xiaodan Zhu[1], Muhammed Jawla[2], Evrim Oral[1*]

[1]LSUHSC School of Public Health Biostatistics Program, [2]LSUHSC School of Public Health Epidemiology Program

LSU Health NEW ORLEANS School of Public Health

Louisiana State University Health Sciences Center School of Public Health

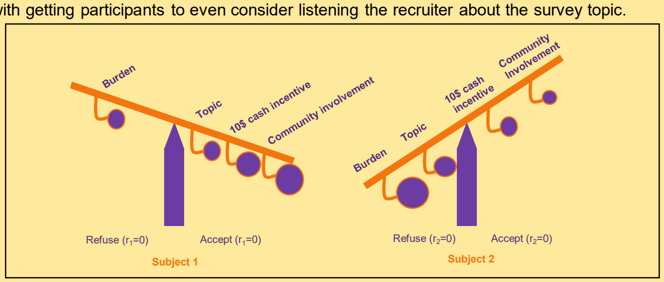
*Contact: eoral@lsuhsc.edu

Motivation and Objectives

There has been a rapid decline in response rates. There are two types of non-response; unit nonresponse, where a respondent does not respond to the survey at all, and item non-response, where a respondent fails to respond to a survey partially. In this study we focus on unit non-response. Unit non-response reduces sample size and study power. It has the potential to introduce non-response bias to estimates. In this study, using an extensive Monte Carlo Simulation, we evaluate the bias and empirical variance under different missingness mechanisms where the response propensity varies based on its dependence on the explanatory variables, outcome variables or survey design. For that purpose, we consider three types of non-response mechanisms: Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR) and search the effects of unit nonresponse on estimating the population mean.

Background

Leverage-Salience theory attempts to explain the causes of unit nonresponse: participants weigh the importance they put on an attribute in the decision process, the leverage, and how appealing the attribute is made during recruitment, the saliency. Cash incentives can be used to increase the salience, for example. Recently, populations growing distrust of their neighbors and increasing modern technologies, like cell phones, have created issues with getting participants to even consider listening the recruiter about the survey topic.



In this study we will consider three types of non-response (missingness) mechanisms):

MCAR occurs when the outcome does not depend on the explanatory variables, outcome variables, or the survey design:

$$f(M|D,\phi) = f(M|\phi) \ , \quad \forall\, D,\phi$$

This type of missingness can be ignored in the analyses. **MAR** occurs when the missing data depends on some observed data D:

$$f(M|D,\phi) = f(M|D_{obs},\phi) \ , \quad \forall D_{obs},\phi$$

in other words, when missingness depends on some observable data collected. This type of missingness can be corrected using weighting adjustment methods. **MNAR** occurs when the probability of nonresponse depends on the data missing:

$$f(M|D,\phi) = f(M|D_{obs},D_{mis},\phi) \ , \quad \forall\phi$$

We can not correct this type of missingness completely.

Simulation Study

To be able to compare the empirical biases and the variances of the sample mean estimator from MCAR, MAR and MNAR, we performed three separate Monte Carlo simulation studies using k=10,000 iterations as follows:

For MCAR, we assumed that the response propensity=p_i is equal for each subject i=1,2,…,n. For sample size=100, we considered that the relationship between the outcome Z and the covariate X is given with the equation

$$Z_i = \theta + \beta X_i + \epsilon_i \qquad (1)$$

where we generated X~N(0,1) and $\epsilon \sim N(0,1)$ and set $\theta = 10$. We defined

Simulation Study (cont.)

$$r_i = \begin{cases} 1, \text{if subject } i \text{ is respondent} \\ 0, \text{if subject } i \text{ is nonrespondent} \end{cases}$$

and generated r_i ~ Bernouilli(p_i). We considered various values for the parameter β changing from 0 to 5 in increments of 0.2. We calculated the empirical bias and empirical variance of the sample mean, as well as the response rate with the following formulas:

$$\bar{y}_k = \frac{\sum_{i=1}^{n} y_i}{n} \quad \text{where} \begin{cases} y_i = z_i \ , \ \text{if } r_i = 1 \\ y_i = \ . \ , \ \text{if } r_i = 0 \end{cases} \ , \quad m_k = \sum_{i=1}^{n} \frac{r_i}{n} \qquad (2)$$

$$Bias = \sum_{k=1}^{10,000} |\bar{y}_k - \bar{Z}_k|/10{,}000 \ , \quad \text{EVar} = \textstyle\sum_{k=1}^{10,000} (\bar{y}_k - z_k)^2/10{,}000 \qquad (3)$$

$$M = \textstyle\sum_{k=1}^{10,000} m_k/10{,}000 \ , \quad \bar{Z}_k = \sum_{i=1}^{n} (\theta + \beta x_i)/n \qquad (4)$$

For MAR, we assumed the same relationship (1) between the outcome and the covariate, but additionally we assumed that the relationship between the covariate and response propensity is as given below

$$\ln\left(\frac{p_i}{1-p_i}\right) = \alpha + \gamma X_i$$

We considered various values for the parameters $\alpha, \beta,$ and γ, changing from 0 to 5 in increments of 0.2. We set $\theta = 10$. MAR follows most of the equations of MCAR, but instead of assigning a fixed p for each subject, we calculated response propensity from

$$p_i = \frac{e^{\alpha+\gamma X_i}}{1+e^{\alpha+\gamma X_i}}$$

for i=1,2,…n. We used formulas (2)-(4) to calculate empirical biases and variances with the corresponding response rates.

For MNAR, we assumed the same relationship (1) between the outcome and the covariate, but we assumed that the relationship between the outcome and the response propensity is as given below

$$\ln\left(\frac{p_i}{1-p_i}\right) = \alpha + \gamma Z_i$$

We considered various values for the parameters $\alpha, \beta,$ and γ, changing from 0 to 5 in increments of 0.2. We set $\theta = 10$. As in MAR, MNAR follows most of the equations of MCAR but instead of assigning a fixed p for each subject, we calculated response propensity from

$$p_i = \frac{e^{\alpha+\gamma Z_i}}{1+e^{\alpha+\gamma Z_i}}$$

for i=1,2,…n. We used formulas (2)-(4) to calculate empirical biases and variances with the corresponding response rates.

Results

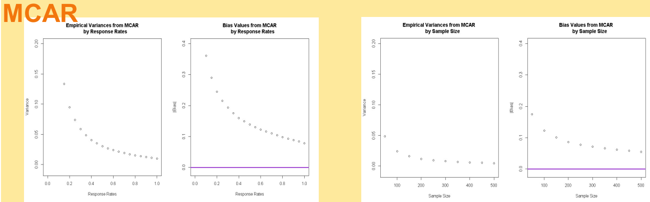
We present our results with the plots given below. In the first two plots, the empirical biases and variances are shown with respect to different response rates and sample sizes. When the missingness mechanism is **MCAR**, the response rate and sample size have similar effects on bias and empirical variance. As the response rate increases, given that all other parameters are fixed, both empirical bias and variance decrease, which is expected. Similarly, when the sample size increases, given that all other parameters are fixed, both empirical bias and variance decrease.

When the missingness mechanism is **MAR**, we see that as α increases, given that all other parameters are fixed, the response rate increase. The inverse is true of the variance and bias; as α increases the variance and bias decrease; this is because the response rate increases with α. Increase in β does not appear to effect the response rate. As β increases, given that all other parameters are fixed, the variance and bias increases. As γ increases, the response rate decrease slightly. As γ increases the variance and bias increase.

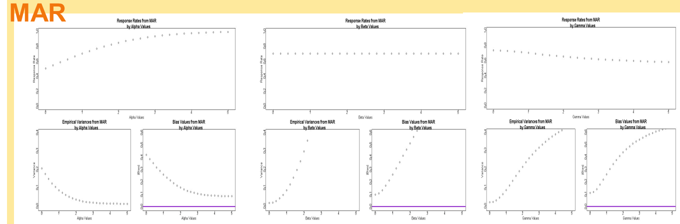
When the missingness mechanism is **MNAR**, we see that when α increases, so does the

Results (cont.)

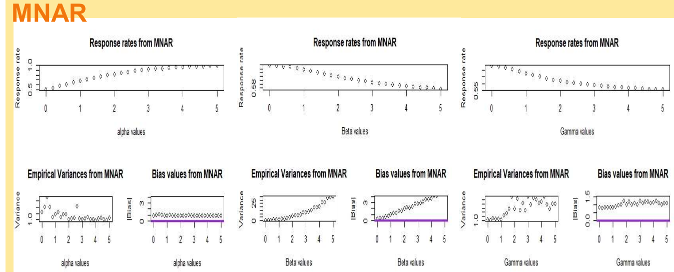
MCAR



MAR



MNAR



response rate. While there is a spike in the variance from α =0 to α =1, as α increases, there appears to be little to no effect on the variance. Similarly, increased α values have no effect on bias. The reason for stable bias and variance is due to the increased response rate. As β increases, the response rate decreases and the empirical variance and bias values increase. As γ increases the response rate decreases. The empirical variance increases as γ increases . There also appears to be a slight increase in the bias values when γ increases.

Conclusion

Using the results from our simulations, we assessed how empirical biases and variances are effected in relation to the different mechanisms of non-response. We showed that for MAR and MNAR, when the relationship between the outcome and the covariate increases, the bias and variance of the sample mean increases. Similarly for both MAR and MNAR, when the relationship between the outcome and response propensity increases the bias and variance of the sample mean increases. We suggest researchers to utilize adjustment or other correction procedures if they suspect they have MAR or MNAR in their data.

References

- Sharon L. Lohr, Sampling: Design and Analysis, 2010, Second Edition, Brooks/Cole, Cengage Learning. ISBN: 0495105279.
- Oral E, Simonsen N, Brennan C, Berken J, Su LJ, Mohler JL, et al. (2016) Unit Nonresponse in a Population-Based Study of Prostate Cancer. PloS ONE 11(12): e0168364. doi:10.1371/journal.pone.0168364
- Jelke Bethelem, Fannie Cobben and Barry Schouten, Handbook of Nonresponse in Household Surveys, 2011, Wiley. ISBN: 978-0-470-54279-8
- Brick JM, Williams D. Explaining rising nonresponse rates in cross-sectional surveys. Ann Am Acad Polit SS. 2013; 645: 36–59.