SUBSAMPLING NONRESPONDENTS: ISSUES OF CALCULATING RESPONSE RATES

Sonya Vartivarian, Donsig Jang, Sameena Salvucci, and Daniel Kasprzyk; Mathematica Policy Research, Inc. 600 Maryland Ave SW, Suite 550. Washington, DC 20024. Email: SVartivarian@mathematica-mpr.com

ABSTRACT

Resource considerations in relationship to data quality often contribute to the decision to subsample nonrespondents for follow-up. For example, the resources saved by limiting the number of nonrespondents to follow may allow for a more effective, and often more expensive, mode of data collection for those subsampled. We consider weighted response rates under subsampling schemes as a measure of data quality. When subsampling is used to follow-up nonresponse, the American Association for Public Opinion Research (AAPOR) guidelines suggest a weighted response rate that sets the nonsampled unit weights to zero and weights the subsampled unit weights by the inverse of the subsampling fraction. We provide simple examples of the calculation of weighted rates using AAPOR guidelines under different subsampling scenarios to illustrate the complexity involved in defining the quality of the data in terms of response rates and their interpretation.

KEY WORDS: Sampling weights, Nonresponse, Subsampling, Response Rate, Data Quality

1. INTRODUCTION

Nonresponse in surveys is an error of nonobservation that reflects an unsuccessful attempt to obtain the desired information from an eligible unit (U.S. Federal Committee on Statistical Methodology, 2001). There are two main types of nonresponse that can occur in a survey: unit nonresponse and item nonresponse. Unit nonresponse is a failure to obtain any data from a sample unit. Survey response rates are a measure of the level of unit response and are frequently used as an indirect indicator of the quality of the data. An unweighted response rate can be calculated as the number of interviews with reporting units divided by the number of eligible reporting units in the sample which is a useful indicator by which to evaluate field performance. For example, a survey may have a target response rate that can be compared to interim unweighted response rates during data collection and be utilized as an assessment of field performance.

In addition, weighting techniques are often used to minimize the effect of nonresponse error where the weight for each sample unit is calculated as the inverse of its probability of selection. Therefore, a weighted survey response rate differs from an unweighted response rate when unequal selection probabilities are used in selecting the sample and characterizes how completely the population was measured prior to adjustment for nonresponse.

Organizations often have internal guidelines on response rate reporting. Guidelines on response rate calculations are provided by the American Association for Public Opinion Research (AAPOR) and are viewed as an industry standard (AAPOR, 2006). National Center for Education Statistics (NCES) standards refer to the AAPOR guidelines and require a nonresponse bias analysis when the response rate suggests the potential for bias to occur, especially if the response rate is less than 85 percent (US Department of Education, 2003). NCES uses a weighted response rate when unequal selection probabilities are present and requires reporting of the weighted response rate for each stage of a survey. The US Census Bureau standards also refer to AAPOR (US Census, 2006). The Office of Management and Budget (OMB) highlights the importance of nonresponse bias studies and has recently put forth guidelines of response bias studies when response rates are below the level of 60% (OMB, 2006). OMB guidelines also recommend the reporting of a weighted response rate for all unequal weighted or complex surveys. Discussions of response rates often relate the response rate to the potential for bias. However, even when nonresponse does not result in bias, inefficiency due to sample loss is still of concern. This issue of loss of efficiency is not reflected in response rate calculations and is discussed in Section 3.

2. SUBSAMPLING FOR NONRESPONDENTS

While a survey with full response will not suffer from nonresponse bias, full response is an increasingly rare situation. The magnitude of nonresponse bias may be partially assessed by response rates, and a direct relationship between nonresponse bias and response rates exists (e.g. Little and Rubin, 2002). Therefore, reducing or eliminating nonresponse is desirable. Survey researchers primarily choose to subsample nonrespondents for follow-up for two reasons: (1) response rate is lower than anticipated and the available resources do not allow for nonresponse follow-up of the full sample; and (2) survey nonresponse is viewed as a potential source of bias and therefore researchers select a subsample of nonrespondents for further study. When compared to further attempts to reach all nonrespondents, subsampling can be viewed as a cost-saving measure since fewer are cases followed. Elliott, Little and Lewitzky (2000) describe examples of subsampling for follow-up including the National Survey of Family Growth, Cycle II (NSFG) and the American Community Survey (ACS). The NSFG is described as subsampling households in difficult-to-reach dwelling units, and the ACS subsamples nonrespondents after failed mail and telephone interviews attempts. In the case of subsampling nonrespondents, the AAPOR guidelines suggest “weighting up” the subsampled group and assigning zero weights for the non-selected group.

However, how should a subsample be designed? Should the subsample be a random sample from all nonrespondents? Or, should it be a random sample within strata, where more nonrespondents are selected from strata with higher (or lower) response rates? Should key domains or subgroups be sampled? Is the method of subsampling crucial, especially with respect to response rate calculations? Such questions require serious attention as illustrated in a simple example found in Table 1 and the resulting weighted response rate in Table 2. The unweighted response rate for this illustration is quite different from both the weighted response rate with subsampling and that without subsampling and demonstrates the need for caution in interpreting such response rates. Which is correct, and for what purpose? It is important to understand what the response rates convey, especially as they relate to data quality.

One aspect of improving data quality when nonresponse is present is bias reduction through the use of available information. If nonrespondents are randomly selected for follow-up, and all subsampled units respond, is it right to say that there is no longer any unit nonresponse, and therefore there is no unit nonresponse error? In an illustrative example, suppose N = 50,000 observations are generated for variables (Y_1, Y_2) that have a bivariate normal distribution with zero mean vector and variance covariance matrix:

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

where $\rho = 0.9$ or $\rho = 0.2$ and represent a high or low association, respectively. A binary stratifier Z is generated where Z=1 if Y_2 is less than or equal to 1, and Z = 2 otherwise. A stratified random sample of size n=1,000 is selected with equal sample allocation in each strata. One variable, Y_1, is subject to nonresponse, where the missing data is generated via a Bernoulli mechanism, and the probability of response depends on Y_2 through the following probit model:

$$P(R = 1 \mid Y_1, Y_2) = \Phi\left(0.1 * Y_2\right). \qquad (1)$$

Therefore, the data are missing at random given Y_2 since missingness of Y_1 depends only on the fully observed Y_2 (Little and Rubin, 2002). Five different subsampling rates are considered in this example: 0.2, 0.5, 0.7, 0.95, or a rate that equals the response probability defined in Equation (1), with a median response probability of approximately 0.5. Therefore, the 2 different populations determined by the correlations between Y_1 and Y_2 (high or low) and the 5 different subsampling rates form 10 different sample response scenarios. However, in this example, for each scenario, all subsampled units respond. As a result of the full response in the subsample, the AAPOR weighted response rate RR1 equals 1 for each of these scenarios, regardless of the rate of subsampling (see Table 3). Further, though the first five scenarios in Table 3 have a high correlation between Y_1 and Y_2, no assessment of the bias or differentiation between the case where the correlation between Y_1 and Y_2 is low is conveyed by the response rates. Certainly, it would be informative to know how useful the data are in addressing potential bias and loss of efficiency for each of these scenarios even though the weighted response rate with subsampling is 1 for all scenarios!

3. INTERPRETATION OF RESPONSE RATES WITH SUBSAMPLING FOR NONRESPONSE AND DATA QUALITY

That weighted and unweighted response rates may differ considerably, and that weighted response rates with and without subsampling, may also be quite different does not address a further concern: response rates do not quantify the quality of the data available for “recovery” of lost information. Perhaps an improved response rate is one that incorporates an indicator of nonresponse adjustment information. For example, in the bivariate normal data situation with one variable subject to nonresponse, the squared correlation between an observed and missing variable is related to the reduction in variance that is associated with taking the observed variable into account in analyses. For example, the simulation study in Vartivarian and Little

(2002) indicates that using a joint classification of the predictive mean in addition to the response propensity when forming nonresponse weighting adjustments offers between one-third and two-thirds of the efficiency achieved by using the mean based on data without nonresponse. By simply reporting a response rate, with or without subsampling, there is no indication of how useful (or not) the data are.

A related issue is the mean square error of estimates and is illustrated when subsampling results in 100% response rates of those subsampled such as in the example found in Table 3. In such a case, do we then have zero bias? And, does this imply that the mean square error is only (or mostly) a variance issue? Such a conclusion is probably not warranted, but further guidance in AAPOR and other guidelines would be helpful. Also, when nonrespondents are subsampled for a different mode of data collection, but the nonsampled nonrespondents are still followed up, the calculation and use of a weighted response rate is further complicated.

4. CONCLUSION

Response rates can be complex, especially with subsampling present. Response rates that incorporate information available for recovery may give a more accurate picture of data quality, and therefore suggest the need to develop such measures. Great progress has been made by AAPOR and others in a directed effort to standardize the computation of response rates contributing to well thought out general standards. These standards can be expanded with further attention to issues of response rate calculation and interpretation when subsampling nonrespondents.

REFERENCES

American Association for Public Opinion Research (2006). Standard Definitions: Final Disposition Codes and Outcome rates for Surveys. 4th edition. Lenexa, Kansas: AAPOR.

Elliott M.R., Little, R.J.A amd Lewitzky S. (2000). Subsampling Callbacks to Improve Survey Efficiency. *Journal of the American Statistical Association*, Vol. 95, No. 451, pp. 730-738

Little, R.J.A. & Rubin, D.B. (2002). *Statistical Analysis with Missing Data,* 2nd edition. Wiley, New York.

Office of Management and Budget (2006). Questions and Answers When Designing Surveys for Information Collections. Office of Information and Regulatory Affairs, January 2006.

Scheuren, F. (2004). What is a Survey? (http://www.whatisasurvey.info)

U.S. Census Bureau (2006). *Census Bureau Standard: Response Rate Definitions.* Issued March 9, 2006. (http://www.census.gov/qualityS180_v1.2_Response_Rate.htm)

U.S. Department of Education, National Center for Education Statistics (2003). *NCES Statistical Standards,* NCES 2003-601. Washington, DC: 20003.

U.S. Department of Education, National Center for Education Statistics (2002). Beginning Postsecondary. Students Longitudinal Study 1996-2001 (BPS:1996/2001) Methodology Report, NCES 2002-171, by Jennifer S. Wine, Ruth E. Heuer, Sara C. Wheeless, Talbric L. Francis, Jeff W. Franklin, and Kristin M. Dudley, RTI. Paula R. Knepper, Project Officer. Washington, DC: 2002.

U.S. Federal Committee on Statistical Methodology. 2001. Measuring and Reporting Sources of Error in Surveys. Washington, DC: U.S. Office of Management and Budget (Statistical Policy Working Paper 31).

Vartivarian, S. and Little, R.J.A. (2002). On the formation of weighting adjustment cells for unit nonresponse, *Proceedings of the Survey Research Methods Section, American Statistical Association 2002*

Table 1. Initial sample and subsample

		Initial Sample		Subsample Information	
Subsampling Strata	Sampled	Initial Responders	Respons e Rate	Subsampling Rate	Expected Respons e Rate
High	1000	700	0.7	0.8	0.5
Medium	2000	1000	0.5	0.7	0.3
Low	1500	150	0.1	0.1	0.1
Total	4500	1850			

Table 2. Weighted response rates with and without subsampling

AAPOR Response Rates (RR1)			
With Subsampling		Without Subsampling	
Weighted	Unweighted	Weighted	Unweighted
0.77	0.49	0.50	0.41

Table 3. An illustration of weighted response rates when all subsampled units respond

$\rho(Y_1, Y_2)$	Samples	Subsampling Rate	Initial Responders	Total Responders	AAPOR Response Rates (RR1) With Subsampling: Weighted	AAPOR Response Rates (RR1) Without Subsampling: Weighted	AAPOR Response Rates (RR1) Without Subsampling: Unweighted
High	1	0.2	543	632	1	0.520	0.543
	2	0.5	529	753	1	0.501	0.529
	3	0.7	546	867	1	0.515	0.546
	4	0.95	538	973	1	0.512	0.538
	5	Probit (mean 0.50)	501	761	1	0.469	0.501
Low	6	0.2	503	612	1	0.479	0.503
	7	0.5	506	743	1	0.476	0.506
	8	0.7	544	854	1	0.504	0.544
	9	0.95	510	977	1	0.474	0.510
	10	Probit (mean 0.50)	550	803	1	0.546	0.550

n = 1000