Journal of Official Statistics, Vol. 27, No. 1, 2011, pp. 29–32

Discussion

Roger Tourangeau[1]

1. Introduction

Until recently, the most widely used indicator of nonresponse bias was the response rate. It was apparent to most survey researchers that the response rate was an inadequate indicator of bias. This was clear, for example, from a commonly cited expression for the nonresponse bias in a mean or proportion (e.g., Lessler and Kalsbeek 1992):

$$
\begin{aligned}
E(\bar{y}_r - \bar{y}_n) &= \bar{Y}_R - [(1-P)\bar{Y}_N + P\bar{Y}_R] \\
&= (1-P)(\bar{Y}_R - \bar{Y}_N) \\
\bar{y}_r &= \frac{\sum_i^r y_{ri}}{r}
\end{aligned}
\tag{1.1}
$$

in which $\bar{y}_r$ and $\bar{y}_n$ represent the sample means for the respondents and nonrespondents on some survey variable, $\bar{Y}_R$ and $\bar{Y}_N$ refer to the corresponding populations means and P is the expected response rate for a survey. Although the nonresponse rate is related to the risk of bias, it is not itself an estimate of the bias for any given mean or proportion. It is also apparent from 1.1 that the nonresponse *rate* will only be highly correlated with the nonresponse *bias* when the expected difference between respondents and nonrespondents does not vary much across variables or across surveys. This is clearly an implausible assumption in most situations.

This theoretical shortcoming of the nonresponse rate as an indicator of nonresponse error has been demonstrated empirically a number of times. Perhaps the most striking evidence for the low correlation between nonresponse rates and nonresponse errors was reported by Merkle and Edelman (2002), who examined response rates in exit polls and the error in estimated vote shares at the precinct level. Table 1 shows the correlations between the two for four recent U.S. elections; the sample sizes in the tables are the numbers of precincts in the analyses. None of the correlations is very impressive and most of them do not differ significantly from zero. Several other studies report similarly weak relationships between nonresponse rates and nonresponse errors. For example, Groves and Peytcheva (2008) report a correlation of about .20 between the two.

2. Imbalance Indicators

Thus, one of the many contributions of Särndal's lecture is that it presents three alternative bias indicators (see Equations 7.1, 7.2, and 7.3). All three of these measures are functions

[1] Joint Program in Survey Methodology, University of Maryland, Survey Research Center, University of Michigan, 1218 LeFrak Hall, College Park, MD 20742, U.S.A. Email: rtourang@survey.umd.edu

Table 1. Correlation between Precinct Response Rate and Signed and Unsigned Errors in the Estimated Vote Share, by Election

Election	Signed errors	Unsigned errors	n
1992	.10	−.13	1,005
1994	.00	−.07	885
1996	−.01	−.04	1,205
1998	.01	−.07	894

Source: Merkle and Edelman (2002).

of discrepancies between the full sample and the responding sample on some set of auxiliary variables, $\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}$, where the elements of $\mathbf{D}$ are the differences between the means for the respondents and for the full sample on the auxiliary variable x (that is, $\mathbf{D} = \bar{\mathbf{x}}_{r;d} - \bar{\mathbf{x}}_{s;d}$) and $\boldsymbol{\Sigma}_s$ is a cross-products matrix for the auxiliaries. Thus, the core of the balance indicators is a standardized measure of the distance between the respondents and the full sample on a set of variables available for both respondents and nonrespondents, typically frame variables.

The balance indicators are, however, useful only to the extent that there are useful auxiliary variables available, so that $\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}$ is itself related to the nonresponse bias. The concern is not whether the responding members of the sample resemble the full sample on some set of covariates, but whether they represent the full sample on the actual survey variables. Of course, if the vector of auxiliaries ($\mathbf{x}_s$) is highly related to the vector of survey variables ($\mathbf{y}_s$), then the balance indicators that Särndal defines will be very helpful proxies for the nonresponse error. This condition may be met in many establishment surveys, where a fair amount of useful information is often available for all of the establishments, but, in the U.S. at least, it is unlikely to be met in most household surveys. With U.S. surveys, weak relationships between frame variables and the survey variables seem to be the rule rather than the exception.

3. Implications for Responsive Design

Before the recent renewed attention to nonresponse error, the usual goal for two-phase sampling (in which a sample of initial nonrespondents was selected for further follow-up field work) was to reduce bias by collecting survey data on a representative sample of the late or difficult sample members, the ones that required special refusal conversion efforts or many additional contact attempts. The danger to this strategy was that the respondents brought in by these second-phase efforts would simply exacerbate the imbalances already present in the first phase. If the second-phase response propensities are proportional to the first-phase propensities ($\theta_{2j} \propto \theta_{1j}$), then the second-phase of data collection will not reduce the nonresponse bias.

As Särndal makes clear (see, for example, Equation 6.2), it is the variance in the estimated overall response propensities ($\hat{\theta}$) that is the key; this variance is proportional to the central imbalance measure $\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}$:

$$S^2_{\hat{\theta}|s,d} = P^2 \times \mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}$$

where P is the overall response rate. To put it another way, to the extent that everyone has the same response propensity, nonresponse is not a problem and does not produce imbalance. As a practical matter, this means that a successful "responsive design" (Groves and Heeringa 2006) is one in which the response propensities in the second phase of data collection compensate for any differences in the response propensities in the first phase:

$$\theta_{2j} \propto 1/\theta_{1j} \tag{2.1}$$

Still, there are a number of practical problems with implementing a strategy that meets the requirements of (2.1). First of all, we never really have the response propensities themselves, only estimates of them, and if the estimates are poor, then we cannot target the right cases for the enhanced follow-up efforts. In addition, the estimates of the response propensities will be based on the same auxiliary variables from which the imbalance measures are derived (perhaps supplemented by paradata) and these may have low predictive power. If the fit of the model for predicting the response propensities is poor (that is, if the model for predicting response propensities has a low r^2), then the fitted propensities will be regressed to the mean and their variance will underestimate the actual variance of the propensities. In a situation in which the response propensities are hard to predict (which is perhaps the usual situation), the fitted propensities will not vary much and the imbalance indicators will give an overly optimistic picture of the nonresponse bias. As with the imbalance measures, the estimated response propensities are only as good as the auxiliary variables they are derived from.

4. Conclusions

The approach taken in Särndal's article is an extension of his earlier very important work on model-assisted and related calibration methods. These strategies necessarily rely on whatever auxiliary variables happen to be available. They raise the issue of how far can we get with approaches that are model-based (or model-assisted) but theory-free. That is, how well can we compensate for nonresponse error without *any* understanding of why people respond (or fail to respond) to a given survey? How far can purely statistical models of response propensities take us? It seems to me that we would be in a much better position if we had some well-validated theory that would allow us to estimate the proportion of sample members with response propensities of zero (for an affordable level of effort) for a given survey or one that gave us some reasonable basis for identifying the survey variables likely to be related to the response propensities. Unfortunately, we seem to be a long way from having such a theory. Until then, we will have to fall back on purely statistical indicators and adjustments.

Despite the skeptical tone of some of my remarks, I see Särndal's Hansen Lecture as having made an extremely valuable contribution to the literature. It clarifies the role of variation in response propensities in producing nonresponse bias and it thus underscores the importance of this statistic. It introduces and systematizes several potential measures of sample imbalance. All of them are much better proxy measures for nonresponse error than the response rates than are now routinely reported. I look forward to the day when Särndal's imbalance measures are reported just as routinely as

AAPOR response rates are now. Finally, the article demonstrates the utility of new weighting procedures when the auxiliary variables do have predictive power. It was an honor to be selected to offer comments on such a valuable article.

5. References

Groves, R. and Heeringa, S. (2006). Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs. Journal of the Royal Statistical Society, Series A (Statistics in Society), 169, 439–457.

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

Lessler, J. and Kalsbeek, W. (1992). Nonsampling Error in Surveys. New York: John Wiley.

Merkle, D. and Edelman, M. (2002). Nonresponse in Exit Polls: A Comprehensive Analysis. In Survey Nonresponse, R. Groves et al. (eds). New York: John Wiley, 243–258.

Received December 2010