Journal of Official Statistics, Vol. 27, No. 1, 2011, pp. 23–28

Discussion

J. Michael Brick[1]

This year is the centennial of Morris Hansen's birth in 1910, and I would like to begin with a few remarks about the person to whose memory this lecture is dedicated. With the passage of time the number of people in the profession who had personal experience with Morris is dwindling. Those who did not have such experience rely largely on his written contributions to judge why he was called the most influential statistician in the evolution of survey methodology in the 20th century. The memoir of Hansen by Waksberg and Goldfield (1996) provides some background to justify this acclamation.

Morris was unlike any other statistician I have ever encountered. He was prolific and his contributions were seminal, and while he was making these contributions he was also developing and directing a world-class bureaucracy at the U.S. Census Bureau. He was principled and relied on theory to guide survey designs, but he also painstakingly considered the practical aspects of the data collection. He was insightful and tackled tough issues like nonsampling errors that were difficult to even conceptualize. Finally, he was a great colleague who was humble, considerate, and willing to share his knowledge and wisdom with struggling new statisticians like me who were interested in learning.

In this special year, it is very appropriate that Professor Särndal is the Hansen Lecturer. He too has made innumerable important contributions to survey research. As I had come to expect from his previous research, Särndal has thought deeply about the issue of nonresponse and provides new and profound insights into the linkage between key components of surveys that determine the quality of the estimates. I would like to thank the Hansen Lecture Committee for giving me the opportunity to comment on this important work.

1. Linking Sampling, Data Collection, and Estimation

The main focus of Professor Särndal's article is the linkage between three distinct components of a survey: sample design, data collection, and estimation. This work makes it clear that these three components need to be considered simultaneously. The best design without proper data collection or estimation will not lead to high-quality estimates. In retrospect, I believe that ignoring data collection was a major shortcoming of the debate that followed the work of Royall (1970) and his colleagues concerning the appropriate method of inference in sample surveys.

[1] Westat, Inc., 1650 Research Boulevard, Rockville, MD 20850-3195, U.S.A. and Joint Program in Survey Methodology, University of Michigan, 1218 Lefrak Hall, Ann Arbor, MI 48106-1248, U.S.A. Email: mikebrick@westat.com

The article restates the importance of data collection as an equal partner with sampling and estimation. Survey sampling is unique in the field of statistics at least partially because it recognizes that data collection plays such an integral role in making inferences about characteristics of finite populations. Other areas of statistics may consider one or two aspects of data collection, but none treat data collection as a key feature that must be addressed at all stages of the statistical investigation.

Särndal utilizes the concept of balance to link sampling, data collection, and estimation. Balance in sampling was used by model-dependent theorists to achieve robustness in the inferences without requiring probability sampling. In this context, balance has a very specific definition that requires that the sample have the same mean (or other distributional characteristic) as the population for specified auxiliary variables. Deville and Tillé (2004) introduced the cube method as a way of producing a balanced sample that is also a probability sample.

Särndal extends the concept of balance into the data collection phase by introducing what he aptly calls a balanced response set. While a balanced sample is useful, inferences flow from data reported by respondents. We must be at least as concerned about this part of the process as about the design. Särndal discusses responsive design as one data collection approach that might be used to achieve a more balanced response set. Other data collection methods could also be used to attain this objective; planned activities such as switching modes or sending incentives to nonrespondents are not responsive designs but accomplish the same purpose (Brick et al. 2005, is one example).

The linkage is completed by recognizing that all data collection methods are limited and a perfectly balanced response set cannot be realized (and even if it could it would not eliminate bias for all survey estimates). Estimation methods are proposed to further reduce the bias due to nonresponse. Särndal suggests calibration, a technique he and others have long studied in the context of nonresponse, as the method of bringing auxiliary data to bear at the estimation phase.

In sampling, data collection, and estimation, powerful auxiliaries are essential to reducing nonresponse bias. The setting Särndal chooses as an example is one in which such auxiliaries are available from registers. Without powerful auxiliaries, the methods outlined will not be very effective in reducing bias. Unfortunately, the lack of powerful auxiliaries is a common phenomenon in the United States, where the frames for household surveys are not very rich. New ideas and methods are being formulated and tested to address this limitation. One idea that has garnered some attention is collecting paradata to supplement frame data. Clearly, the creation of powerful auxiliaries for reducing bias is an area that is ripe for continued development.

2. Balance Measures and Nonresponse Bias

Three balance measures are introduced and proposed as indicators of the representativeness of the estimates produced from a survey. In one sense, these measures can be considered replacements for the response rate, which has little power for predicting bias (see Groves et al. 2008). As such they join other measures such as those of Särndal and Lundström (2005) and Schouten (2007). However, Särndal goes much further in his article in utilizing balance measures to motivate data collection strategies and estimation techniques.

The quantity that plays a central role in all of the balance measures presented by Särndal is $\mathbf{D}'\mathbf{\Sigma}_s^{-1}\mathbf{D}$ or $\mathbf{D}'\mathbf{\Sigma}_r^{-1}\mathbf{D}$. When there is a single auxiliary variable, this quantity is directly related to a well-known expression for nonresponse bias for a mean since

$$\mathbf{D}'\mathbf{\Sigma}_s^{-1}\mathbf{D} = (1-P)^2(\bar{x}_{r,d} - \bar{x}_{s,d})^2\mathbf{\Sigma}_s^{-1} \propto (1-P)^2(\bar{x}_{r,d} - \bar{x}_{s,d})^2.$$

Notice that the positive square root of the last quantity, $(1-P)(\bar{x}_{r,d} - \bar{x}_{s,d})$, is the deterministic expression for nonresponse bias estimated for a given sample based on a particular auxiliary variable. This relationship is very reassuring and helps justify the critical role of this quantity in all of the balance measures.

Särndal then derives the standardized adjustment, the "bias" defined as the difference between the unadjusted and calibrated estimate in standardized units, as the product of three factors. The first factor is $\mathbf{D}'\mathbf{\Sigma}_r^{-1}\mathbf{D}$. The second factor is R_{yx}, the correlation between the y variable and the auxiliary vector from a design-weighted multiple regression fit. The third factor, R_{DC}, is an unusual one in that it is the correlation between differences, $D_j = (\bar{x}_{j|r;d} - \bar{x}_{j|s;d})$, and covariances, $C_j = Cov(x_j, y)$. The correlation in the third factor is based on weighted least squares fit of the regression through the origin of D_j on C_j.

The standardized adjustment is a new expression. Its components make sense heuristically. The first component is the contribution to bias due to differences in means of auxiliary variables between the sampled and responding units. The product of the second component and the third component is the reduction in "bias" associated with the auxiliary vector, R_{DC} is large when big differences in the auxiliaries covary with the correlation between the auxiliary variables and y. R_{yx} is large when there is strong linear relationship between y and x.

At first, I thought the standardized adjustment might be interpreted as an extension of the deterministic nonresponse bias expression from the simple Horvitz-Thompson to the GREG estimator, but this is not the case. For example, $R_{yx} = 0$ does not imply there is no bias; it means that the auxiliaries are not related to the y, and thus do not reduce bias. The issue is that the "bias" is being measured as the distance between the unadjusted estimate and the calibrated estimate, as Särndal clearly states. The hope is that this approximates the bias, but it only does so when the auxiliaries are powerful predictors for specific y's.

It is interesting that the article does not discuss response propensity weighting and the stochastic view of nonresponse bias. Brick and Jones (2008) take this perspective and provide a representation for nonresponse bias for the poststratified estimator that, for a sample mean, is approximately proportional to the sum of the within poststratum covariances between the response propensities and the y's. The ability of the auxiliaries to reduce bias is due to their ability to partition the population into cells (poststrata) in which the y's are not related to the response propensities. It might be possible to extend these types of relationships to measures that, like the standardized adjustment in the article, are computable from the sample. Similarly, deterministic nonresponse bias expressions at the population level for simple calibration estimators like those for the stochastic approach could be informative. These are areas in need of more development.

3. Response Enhancement in Practice

As we contemplate changing data collection strategies using techniques such as responsive designs, we need to be cautious of the effects such changes may have on the quality of the data. Some of the common strategies for enhancing data quality are using incentives, changing modes of data collection, reducing the instrument length, trying to convert reluctant respondents, and modifying the information given about sponsorship or apparent content of the survey. All of these methods have been shown, at least in some circumstances, to increase response rates. However, they do not always decrease nonresponse bias. Furthermore, there are other effects that need to be considered. Some examples are given to provide some perspective.

Mohadjer et al. (1997) describe an experiment in a National Adult Literacy Survey field test of using incentives that seems to have nearly all the desirable types of effects. Not only were the overall response rates increased, but the incentive induced greater response from the lowest education and minority adults who were underrepresented without an incentive. As a result, key outcome estimates like literacy scores by race and education level were different, and apparently less biased. In addition, the incentives lowered the overall data collection costs.

This experiment is one of the few I have encountered that exhibits all these desired results. By far, most other manipulations show that efforts that are effective in increasing response rates have little or no effect on nonresponse bias. Keeter et al. (2000) is one such example; in this study they greatly increased response rates in a telephone survey by increasing the level of effort only to find that it had virtually no effect on the bias for a wide variety of statistics. There are many other examples with similar results that used different methods to increase response rates. Some recent studies are by Haring et al. (2009), Wetzels et al. (2008), and Ingen et al. (2009).

Even more discouraging are surveys that have implemented procedures that increased response rates but found that the higher response rates were accompanied by larger rather than smaller nonresponse bias. Merkle et al. (1998) give an example that involves the use of incentives alone; Schmeets (2010) describes a more complete revision in data collection procedures to increase response rates.

Efforts aimed at increasing response rates may also introduce other nonsampling errors. For example, the American Community Survey in the United States switches modes from mail to telephone and then face-to-face to increase its response rates. Those rates do increase substantially. The switch also admits new errors such as mode effects and interviewer effects that must be evaluated against any reductions in nonresponse bias. The trade-offs are complex and formal evaluations are rarely done.

Särndal's use of balance to link data collection and estimation also clarifies what can be expected from the estimation stage. Just as enhanced data collection does not always reduce nonresponse bias, adding auxiliaries in the estimation stage may have little effect on nonresponse bias. Unfortunately, the standardized adjustment is dependent on the assumption that the bias decreases. This situation is in some sense unavoidable because we do not observe the characteristics of the nonrespondents and are forced to rely on some model assumptions.

4. Further Challenges

Professor Särndal has once again provided us with new insights and approaches for addressing nonresponse in surveys, along with much to ponder in this struggle. His proposed measures of balance appear to be appropriate for surveys such as those that he uses as examples – those with powerful auxiliaries. Surveys that are not so fortunate must look for other solutions.

After decades of work, we still do not have a strong theoretical basis for understanding the causes of nonresponse and the types of estimates that will be most biased from that nonresponse. The literature is sprinkled with examples, but theories that are consistent at predicting bias are restricted to the simplest situations. Similarly, stepwise regression and search algorithms are still used to choose which auxiliary variables should be included in the calibration step to reduce nonresponse bias because theoretical guidance is weak.

Without a solid theoretical basis we are left to speculate whether enhanced data collection methods that improve response rates will reduce bias. Are the costs of enhanced methods worthwhile? Does the use of auxiliaries reduce nonresponse bias in this case? The article shows that we have much work left to do on nonresponse.

5. References

Brick, J.M., Montaquila, J., Hagedorn, M.C., Brock Roth, S., and Chapman, C. (2005). Implications for RDD Design from an Incentive Experiment. Journal of Official Statistics, 21, 571–589.

Groves, R.M., Brick, J.M., Couper, M., Kalsbeek, W., Harris-Kojetin, B., Kreuter, F., Pennell, B., Raghunathan, T., Smith, T., Tourangeau, R., Bowers, A., Jans, M., Kennedy, C., Levenstein, R., Olson, K., Peytcheva, E., Ziniel, S., and Wagner, J. (2008). Issues Facing the Field: Alternative Practical Measures of Representativeness of Survey Respondent Pools. Survey Practice, October.

Haring, R., Alte, D., Völzkea, H., Sauer, S., Wallaschofski, H., John, U., and Schmidt, C. (2009). Extended Recruitment Efforts Minimize Attrition but not Necessarily Bias. Journal of Clinical Epidemiology, 62, 252–260.

Ingen, E., Stoop, I., and Breedveld, K. (2009). Nonresponse in the Dutch Time Use Survey: Strategies for Response Enhancement and Bias Reduction. Field Methods, 21, 69–90.

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, 125–148.

Merkle, D., Edelman, M., Dykeman, K., and Brogan, C. (1998). An Experimental Study of Ways to Increase Exit Poll Response Rates and Reduce Survey Error. Paper presented at the Annual Conference of the American Association of Public Opinion Research, St. Louis, MO.

Mohadjer, L., Berlin, M., Rieger, S., Waksberg, J., Rock, D., Yamamoto, K., Kirsch, I., and Kolstad, A. (1997). The Role of Incentives in Literacy Survey Research. Chapter 10 in Adult Basic Skills: Innovations in Measurement and Policy Analysis, A. Tuijnman, I. Kirsch, and D. Wagner (eds). Creskill: Hampton.

Schmeets, H. (2010). Increasing Response Rates and the Consequences in the Dutch Parliamentary Election Study 2006. Field Methods, 22, 391–412.

Waksberg, J. and Goldfield E. (1996). Biographical Memoirs: Morris Howard Hansen, December 15, 1910–October 9, 1990. National Academy Press, Washington D.C. Available at www.nap.edu/readingroom.php?book = biomems&page = mhansen.html

Wetzels, W., Schmeets, H., Van den Brakel, J., and Feskens, R. (2008). Impact of Prepaid Incentives in Face-to-face Surveys: A Large-scale Experiment with Postage Stamps. International Journal of Public Opinion Research, 20, 507–516.

Received December 2010