VERSITA

Journal of Official Statistics, Vol. 29, No. 3, 2013, pp. 355–358, DOI: 10.2478/jos-2013-0027

Discussion

Olena Kaminska[1]

1. Introduction

The article by Michael Brick comes at a time when the survey methodology field is actively looking for solutions to constantly decreasing response rates. After a number of decades developing design features for achieving higher response rates, and therefore unintentionally educating our clients and funders that response rates are important, we are now struggling to explain the importance of nonresponse bias. But what is more challenging is to understand ourselves how we can deal with nonresponse bias in the best way.

I found the article to be a much needed reminder to the field of the gaps in our knowledge about nonresponse bias today, and how much is to be developed in order to identify best practice in dealing with nonresponse. The work is both comprehensive and current with a historical overview of research into nonresponse, and identification of areas with unanswered issues, and areas with the potential to answer pressing questions.

I enjoyed reading about recent developments in the field of nonresponse that directly refer to nonresponse bias, instead of response rate. Brick first reviews adaptive or responsive design that tailors data collection in order to decrease nonresponse bias. One attraction of such designs is the idea of tailoring fieldwork procedures in response to information obtained before or during the fieldwork. Yet to me the biggest value of such an approach is that for the first time we are developing design with an explicit aim of decreasing nonresponse bias. Adaptive and responsive designs do not have to be the only designs with such an aim; and as the author suggested, we should review already developed design features with respect to their influence on nonresponse bias. We know that incentives increase response rates (e.g., Singer et al. 2000; Singer et al. 1999), but do they also decrease nonresponse bias? We know that mentioning a salient topic of the survey may increase response rate (e.g., Groves et al. 2004), but does this decrease nonresponse bias? Questions like these require answers in order to tailor our practice to decreasing nonresponse bias directly, rather than through increasing response rate alone.

Another important development mentioned is the collection of new paradata which should give stronger predictors for the adjustment stage. While weighting for nonresponse is hoped to be a ‘solution’ to nonresponse bias, it largely depends on good correlates of nonresponse and of *y*-variables (more precisely, of estimates of substantive interest). Often little information is available on both respondents and nonrespondents; and gathering additional information that can be used in nonresponse adjustment models and that is

[1] ISER, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK. Email: olena@essex.ac.uk

tailored to important y-variables has direct impact on the quality of nonresponse adjustment. While little resources tend to be put into collecting paradata in comparison to large resources for converting reluctant respondents, it is possible that the reverse would be most beneficial for reducing nonresponse bias in final estimates.

A more complex development suggested by the author is an integration of three practices that largely have been developing autonomously so far: research into causes of nonresponse, development of design features to decrease nonresponse bias, and adjusting for nonresponse. For example, from a fieldwork perspective, responsive design is a very attractive set of procedures which in the end should result in minimal nonresponse bias on selected variables. Yet such a design, having differential selection and nonresponse probabilities, may lead to an increase in standard errors of estimates which can outweigh the gains from bias. While this is theoretically possible, little is known about such interaction at the moment. Thinking about both design features and nonresponse adjustment in this example would pose these questions earlier, and will challenge the development of designs that optimize collection and adjustment simultaneously.

With the above said, I feel that the literature on survey weighting is particularly in need of development in order to answer the questions being raised by the innovations in data collection procedures. Weighting has largely developed in the previous century for a one-time cross-sectional study of one population and for one survey protocol. Michael Brick's article is one of very few attempts today to develop the best weighting approach for a situation which differs from that above: a situation where the survey protocol changes during data collection. This includes two-stage design, where only some nonrespondents are attempted in the second stage, responsive design, or a design with increasing incentives in the later stages of the fieldwork. In my discussion, I comment on response probabilities in such situations and point out an alternative weighting procedure to account for selection probabilities and nonresponse.

2. Do Response Probabilities Change with Fieldwork?

This is one of the questions raised by Michael Brick in the article (Section 6). In my opinion, the answer to the above question is yes and no – and both perspectives are useful. When we think of fieldwork and design procedures to convert reluctant sample members, we aim to change reluctant sample members' probabilities conditional on not having yet participated. We do this either by sending reminders, issuing another call, offering higher incentive, sending more experienced interviewer and so on – each of these with one aim: to increase the chance of response of those who have not responded yet. The idea that *conditional* response probabilities are constant and cannot be changed over the fieldwork period is not practical in such a situation as it would imply that whatever we do – we cannot help bringing more respondents through design. In this situation researchers are interested in response probability at a particular call – and it is useful to treat such conditional probabilities as prone to manipulation via survey design features.

Nonetheless I share the opinion of the author (Section 6) that the above perspective of changing probabilities over time is not useful in all contexts; in particular, weighting adjustment should estimate *final* probabilities, that is, total, cumulative probabilities over all stages of survey fieldwork. This is because at the end of the fieldwork period we aim to extrapolate the information from final respondents to the whole sample (or population). It is therefore important to know the final response probability for each sample member. From this perspective final probabilities under the same protocol and in the same survey situation (population, topic of the survey, etc.) are constant, and do not vary over fieldwork time (unlike the conditional probabilities discussed above). I understand that in the discussion of Figure 1 in the article when describing RHGs Michael Brick talks about final probabilities to respond.

3. Weighting for a Two-Stage Design

One of the contributions of Michael Brick's article is the discussion of weighting for a two-stage design, where some respondents participate in a survey in the first stage, and at the second stage all or a subsample of nonrespondents is followed, some of whom also provide interviews. The author suggests two ways of developing weights in this situation, Method A and Method B. I believe that both methods are unbiased under specific assumptions. Method B is unbiased under MAR assumption that all respondents are different from nonrespondents only on variables in the nonresponse adjustment model. Method A is unbiased under MAR assumption that Stage 2 respondents are different from nonrespondents only on variables in the nonresponse adjustment model. I agree with the author that Method A corrects for nonresponse bias better than Method B when Stage 2 respondents are more similar to final nonrespondents in comparison to Stage 1 respondents.

I would like to suggest Method C for weighting correction in a two-stage design, which not only recognizes the two stages of design, but also recognizes that each respondent has a chance to respond at either (but not both) of the two stages. The discussion from the previous section becomes useful here: at both stages of the design respondents have probabilities to respond – the probability of responding in the second stage is conditional on not responding in the first stage; the total probability is the combination of these two probabilities. Thus, the total response probability can be expressed as

$$p_{\text{total}} = p_1 + (1 - p_1)^* p_2$$

where p_1 is the probability to respond at the first stage and p_2 is the conditional probability to respond at the second stage. $(1 - p_1)$ expression reflects a chance of a sample member being issued into Stage 2, which is conditional on nonresponse in Stage 1.

In the design where second stage nonrespondents are subsampled, a probability of selection (p_{sel}) should be included in the expression:

$$p_{\text{total}} = p_1 + (1 - p_1)^* p_{\text{sel}}^* p_2$$

The important point here is that every selected sampling unit has a value for each probability. In other words, respondents, who are observed to have responded in Stage 1, had a chance to not respond in Stage 1. In this situation they would have a chance to be selected into Stage 2, and a conditional chance to respond in Stage 2.

While the formulae make sense theoretically, estimating these probabilities in practice is challenging given that we do not observe a Stage 2 response outcome for those not selected into Stage 2 (either because of subselection or because they have responded in

Stage 1 already). Such calculation is nevertheless possible and can follow an approach similar to the one in Kaminska and Lynn (2012). First, p_1 is estimated in the usual way using predictors available for respondents and nonrespondents. Selection probability p_{sel} is known by design. Next, p_2 is estimated only for those who were issued into Stage 2, drawing upon the same pool of auxiliary variables as in the above model. Given the model for p_2, we can now estimate p_2 for all respondents, including respondents from Stage 1. This is possible because the same auxiliary variables are available for all respondents. This way we estimate response probability in Stage 1, p_1, and conditional response probability in Stage 2, p_2, for each respondent, regardless of the stage at which they participated. This provides us with all the components required for the nonresponse correction.

One advantage of this approach over methods A and B, described by Michael Brick, is that it estimates response probabilities at each stage empirically and independently of each other, thus avoiding the unnecessary assumptions.

4. Conclusion

It has been an honour to be a discussant of such an interesting, comprehensive, current and innovative article. There are many more thoughts and ideas in the article worth discussion and further development. I feel we are at the turning point of understanding nonresponse and I look forward to future developments in this field.

5. References

Groves, R.M., Presser, S., and Dipko, S. (2004). The Role of Topic Interest in Survey Participation Decisions. Public Opinion Quarterly, 68, 2–31.

Kaminska, O. and Lynn, P. (2012). Combining Refreshment or Boost Samples with an Existing Panel Sample: Challenges and Solutions. Proceedings of the International Panel Survey Methods Workshop.

Singer, E., van Hoewyk, J., and Maher, M.P. (2000). Experiments with Incentives in Telephone Surveys. Public Opinion Quarterly, 64, 171–188.

Singer, E., van Hoewyk, J., Gebler, N., Raghunathan, T., and McGonagle, K. (1999). The Effect of Incentives on Response Rates in Interviewer-Mediated Surveys. Journal of Official Statistics, 15, 217–230.