Strategies for Selecting a Follow-up Sample of Nonrespondents in Business Surveys

Elisabeth Neusy[1], Wesley Yung[1], Jean-François Beaumont[1], Mike Hidiroglou[1], David Haziza[2]
[1] Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario, Canada K1A 0T6
[2] Département de mathématiques et de statistique, Université de Montréal, Montréal, Québec, Canada H3C 3J7

Abstract
Follow-up of nonrespondents in business surveys is a time and resource intensive activity. Unlike household surveys where respondents are more-or-less equivalent, business surveys have influential units for whom a response is absolutely necessary in order to not adversely affect level estimates. An extensive simulation study was undertaken to compare different methods of performing follow-up for business surveys. For a fixed budget, is it best to follow-up all nonrespondents or only a sample of nonrespondents? If it is better to follow-up a sample of nonrespondents, what is the best way to select this sample? Should the sample be selected using simple random sampling (SRS), stratified SRS or probability proportional to size (PPS) sampling? The different methods were evaluated via a simulation study by comparing their Monte Carlo mean square errors.

Key Words: Nonresponse, Follow-up

1. Introduction

Collection research is currently a very hot topic amongst national statistical agencies looking to increase response rates and/or reduce collection costs. With the extremely high costs of collecting survey data, even a small increase in efficiency can translate into significant monetary savings. Recognizing this opportunity, much research has been done in recent years on how collection costs can be managed despite the growing reluctance of businesses and individuals to respond to survey requests. Falling response rates imply that there is increased concern for the potential of nonresponse bias.

Much of the research done to date has been on adaptive collection designs (also called adaptive survey designs, responsive collection designs, responsive survey designs or simply responsive designs in the literature). Groves and Heeringa (2006) define a responsive survey design as one that uses paradata, or process data, to guide changes in the features of data collection in order to achieve higher quality estimates per unit cost. Beaumont, Bocci and Haziza (2014) note that the literature on adaptive collection designs has mainly focussed on developing procedures that aim at reducing the nonresponse bias of an estimator which is not adjusted for nonresponse (see for example Schouten, Cobben and Bethlehem, 2009, and Peytchev, Riley, Rosen, Murphy and Lindblad, 2010). Beaumont, Bocci and Haziza (2014) argue that any information (e.g., auxiliary data,

paradata) that can be used during collection to reduce nonresponse bias can also be used at the estimation stage. In other words, the nonresponse bias that can be removed at the collection stage through an adaptive collection procedure, could also be removed at the estimation stage through appropriate nonresponse weight adjustments. They suggest that adaptive collection procedures, such as call prioritization, cannot reduce the nonresponse bias to a greater extent than a proper nonresponse weight adjustment.

In one of the first papers to discuss nonresponse, Hansen and Hurwitz (1946) propose subsampling nonrespondents to reduce nonresponse bias. They consider the case where questionnaires are mailed out and after a certain period of time, a subsample of nonrespondents is randomly selected for follow-up by personal interviewers to obtain their responses. They show that the nonresponse bias can be completely eliminated if all the units selected in the subsample respond. The latter condition is not realistic in practice. However, subsampling nonrespondents may achieve some nonresponse bias reduction even if there is some residual nonresponse in the follow-up subsample.

This paper describes a simulation study that was undertaken to examine different strategies of subsampling nonrespondents in the context of business surveys. Data collection for business surveys at Statistics Canada typically begins with the mailout of a questionnaire to the businesses selected in the sample. The nonrespondents to the mailout are followed up by telephone, which is a time and resource intensive activity. The current study examines, for a fixed budget, whether it is preferable to follow-up all the nonrespondents or only a sample of the nonrespondents. It compares different sample designs for the follow-up: simple random sampling (SRS), stratified SRS, and probability proportional to size (PPS) sampling. The different strategies of selecting the follow-up sample are evaluated by comparing their Monte Carlo mean square errors.

The paper is organized as follows: Section 2 provides a description of the simulation, Section 3 summarizes the results of the simulation and Section 4 gives a few conclusions.

2. Methodology of the Simulation

This section describes each of the steps involved in the simulation.

2.1 Data used for the Simulation

The data used for the simulation was a sample from an actual business survey: Statistics Canada's Monthly Survey of Food Services and Drinking Places (MFS). Two variables were used for the study: 'Revenue' and 'Sales'. The first variable, Revenue, comes from the frame, and is present for all units selected in the MFS sample. The second variable, Sales, is one of the variables collected by the survey. Both unit and item nonresponse were handled by imputation. Overall, Sales was imputed for 15% of the units in the sample. The correlation between Revenue and Sales is about 83%, with and without the imputed data.

As is typical for business surveys, the MFS uses a stratified random sample design with 'Take All' and 'Take Some' strata. The 'Take All' strata consist of the large and important businesses that are selected with certainty in the sample. The take-all units were excluded from the study because they are all followed up - subsampling them for follow-up is not an option. A total of 2,375 sample units were included in the study, from 63 'Take Some' strata (province by industry as defined by the North American Industry

Classification System). The set of sample units included in the study are denoted as S_1. The subset of S_1 falling into strata h, h=1,...,63, is denoted as S_{1h},.

2.2 Mailout Collection

Data collection for business surveys typically begins with a mailout of questionnaires to the units selected in the sample. The mailout involves the mailing of a paper questionnaire, or the mailing or e-mailing of a letter with a link to an electronic questionnaire. Some units do not respond to the mailout. The set of respondents to the mailout is denoted as S_{1R}, and the set of nonrespondents as S_{1NR}. Nonresponse to the mailout is simulated as follows. Before the start of the simulation each unit is assigned a probability of response, denoted as p_{1hi} for unit i in stratum h. Nonresponse is randomly generated based on the values of p_{1hi}.

Two different response scenarios were considered:

1) Uniform: p_{1hi} = 50% for all units, regardless of the characteristics of the units. Under this scenario, the expected number of respondents to the mailout is 1187.5 (=2,375/2).

2) Correlated to Sales (the variable of interest): p_{1hi} is based on the logistic function,

$$\text{Ln}\left(\frac{p_{1hi}}{1-p_{1hi}}\right) = -0.31 + 0.000004\, y_{hi},$$

where y_{hi} is the value of the Sales variable. The constants, -0.31 and 0.000004, were chosen so that the expected number of respondents to the mailout under this scenario is also approximately half of the initial sample. Note that under this model, there is a 97% correlation between p_{1hi} and Sales.

2.3 Follow-up Sample

The next step in the simulation is to select a follow-up sample from the mailout nonrespondents (S_{1NR}). The follow-up sample is denoted as S_2. The following designs were considered for the follow-up sample:

1) Follow-up of all mailout nonrespondents is carried out.
2) A simple random sample is selected from S_{1NR} (the original stratification is ignored).
3) A stratified SRS is selected from S_{1NR} using the original stratification. Sample allocation is proportional to the number of mailout nonrespondents.
4) A sample is selected systematically from S_{1NR} with probability proportional to the frame variable, Revenue. The original stratification is ignored.
5) A sample is selected systematically from S_{1NR} with probability proportional to Revenue multiplied by the design weight. The original stratification is ignored.

Note that the size variables used for the PPS sampling were trimmed from below to the 5th percentile to remove zeros and some extremely small values which were causing instability.

On average, there are 1,188 nonrespondents to the mailout. For the first design, all 1,188 nonrespondents are followed up. For the remaining four designs, the follow-up sample sizes used for the simulation were chosen as 100, 200, 300, 400, 500, 700, and 900.

2.4 Follow-up Collection

Follow-up of nonresponding units for business surveys is typically done by telephone. Multiple phone call attempts are sometimes necessary to reach and get a response from a unit. The collection process is simulated at the call attempt level. Each call attempt is randomly assigned one of the following possible outcomes:

1) Response: a response is obtained from the unit. The unit is removed from the calling queue so that it does not get called again.
2) Final Nonresponse: the unit is finalized as a nonrespondent and removed from the calling queue. An example of this outcome is a final refusal.
3) Still in-progress: the unit needs to be called again; it is returned to the calling queue. An example of this outcome is an attempt where no contact is made (i.e., no answer), or an attempt where an appointment is made for a call back.

The 'response' outcome and 'final nonresponse' outcome are both final outcomes, in the sense that the unit is removed from the calling queue and from collection. This is in contrast to the 'still in-progress' outcome where the unit is returned to the calling queue so that it can be called again.

For each unit, the probability of each of the three outcomes is assigned before the start of the simulation. The probabilities are denoted as $P_{2hi}^{(1)}$, $P_{2hi}^{(2)}$, and $P_{2hi}^{(3)}$ for unit i in stratum h, for outcomes 'response', 'final nonresponse' and 'still in-progress' respectively.

Two different response scenarios were considered:

1) Uniform: $P_{2hi}^{(1)}$ = 25%, $P_{2hi}^{(2)}$ = 5%, $P_{2hi}^{(3)}$ = 70% for all units, regardless of the characteristics of the units. This set of call outcomes and probabilities follow the set up of Xie, Godbout, Youn and Lavallée (2011).

2) Correlated to Sales (the variable of interest): The probability of a 'response' is based on the logistic function,

$$\text{Ln}\left(\frac{P_{2hi}^{(1)}}{1-P_{2hi}^{(1)}}\right) = -1.29 + 0.000002\, y_{hi} + 0.3\, Z,$$

where y_{hi} denotes Sales and Z is Normal (0,1). The constants, -1.29, 0.000002, and 0.3 were chosen so that the expected number of units with outcome 'response' is approximately 25% of the sample when the expectation is calculated over the entire sample. The other two probabilities are defined as $P_{2hi}^{(2)} = \frac{0.05}{0.75}\left(1 - P_{2hi}^{(1)}\right)$, and $P_{2hi}^{(3)} = \frac{0.70}{0.75}\left(1 - P_{2hi}^{(1)}\right)$. Note that there is a 61% correlation between $P_{2hi}^{(1)}$ and Sales, due to our choice of constants and the addition of the random noise variable, Z.

Some units are called several times. For each call, the same set of probabilities, $P_{2hi}^{(1)}$, $P_{2hi}^{(2)}$, and $P_{2hi}^{(3)}$, are used to randomly generate the outcome of the call. The unit remains in the calling queue until it is finalized with an outcome of 'response' or 'final nonresponse', or until collection ends. There is no cap on the number of call attempts made to the same unit.

Note that for the second response scenario, where the probabilities of response are not the same for all units, the units that respond to the first call attempt tend to be units that have a higher probability of response. This implies that the units that remain in the calling queue for the second attempt tend to be units with a lower probability of response. Consequently, the proportion of units that respond in the second attempt is lower than in the first attempt. Similarly, the proportion of units that respond in the third attempt is lower than in the second attempt, and so on. The proportion of units that respond decreases with each call attempt, as the units that remain in the calling queue are those with a lower probability of response.

The simulation takes cost into account by assigning a cost for each call attempt. The amount charged depends on the outcome of the attempt: a 'response' outcome has a cost of 5, a 'final nonresponse' outcome has a cost of 2, and a 'still in-progress' outcome has a cost of 1. In practice, cost is related to the length of the phone call. The simulation assumes that phone calls with a 'response' outcome tend to be longest, and phone calls with a 'still in-progress' outcome tend to be shortest. The total collection budget is fixed at 3,000. Collection ends when the budget runs out, or when there are no more units left in the calling queue (i.e., all units are finalized with an outcome of 'response' or 'final nonresponse'), whichever occurs first.

At the end of data collection, there are two types of nonrespondents: the units that were finalized with a 'final nonresponse' outcome, and the units that remain in the calling queue at the end of collection with a 'still in-progress' outcome. Both types of nonrespondents are taken into account through weighting adjustments. The set of nonrespondents is denoted as S_{2NR} whereas the set of respondents is denoted as S_{2R}.

2.5 Estimation

The final step of the simulation is to compute weights and produce estimates. Given that the parameter of interest was the total, the estimator used in the study is given by:

$$\hat{Y} = \sum_h \sum_{i \in s_{1hR}} w_{1hi} y_{hi} + \sum_h \sum_{i \in s_{2hR}} \tilde{w}_{2hi} y_{hi} ,$$

where y_{hi} is the variable of interest (Sales) for unit i in stratum h, S_{1hR} is the set of units in stratum h that responded to the mailout, S_{2hR} is the set of units in stratum h that responded to the follow-up. The weights, w_{1hi} and $\tilde{w}_{2hi}$, respectively represent the weights for the mailout respondents and the adjusted weight for the follow-up respondents. The weight, w_{1hi}, is the design weight for the original sample; it is given by $w_{1hi} = 1/\pi_{1hi}$, where π_{1hi} is the inclusion probability associated with the selection of the original sample. The adjusted weight for the follow-up respondents, $\tilde{w}_{2hi}$, is of the form $\tilde{w}_{2hi} = w_{2hi} \times a_{2hi}$, where $w_{2hi} = 1/\pi_{1hi} \times 1/\pi_{2hi}$ is the design weight for the follow-up sample, π_{2hi} is the inclusion probability associated with the selection of the follow-up

sample, and a_{2hi} is a global nonresponse adjustment for the follow-up nonrespondents. This nonresponse weight adjustment is calculated as follows:

$$a_{2hi} = \frac{\sum_h \sum_{i \in s_{2h}} w_{2hi}}{\sum_h \sum_{i \in s_{2hR}} w_{2hi}},$$

where S_{2h} is the set of units in stratum h selected for the follow-up. Note that this is a very simplistic method of adjusting for nonresponse. The Revenue variable from the frame, which is correlated with Sales, could have been used to improve the nonresponse adjustment. The Revenue variable was not used so that the effect of subsampling the nonrespondents could be studied more easily.

3. Simulation Results

The simulation was performed for four response scenarios:

1) The probability of response is uniform for both the mailout and the follow-up. This serves as a baseline scenario with which to compare the other scenarios.
2) The probability of response is correlated to Sales for the mailout, and uniform for the follow-up.
3) The probability of response is uniform for the mailout, and correlated to Sales for the follow-up.
4) The probability of response is correlated to Sales for both the mailout and the follow-up. This scenario is probably more realistic.

For each scenario, the simulation was performed with the follow-up sample designs described in Section 2.3. The simulation was repeated 1,000 times for each response scenario and design, and an estimate based on the estimator given in Section 2.5 was computed. The follow-up sample designs were evaluated by comparing their Monte Carlo relative bias (RB) and relative root mean square error (RRMSE), which were calculated as follows:

$$\text{RB} = \frac{\frac{\sum_{r=1}^{R} \tilde{Y}_r}{R} - \tilde{Y}}{\tilde{Y}} \times 100\%, \text{ and RRMSE} = \frac{\sqrt{\frac{\sum_{r=1}^{R} (\tilde{Y}_r - \tilde{Y})^2}{R}}}{\tilde{Y}} \times 100\%,$$

where $\tilde{Y}$ is the estimate for Sales based on the original sample and the original weights, $\tilde{Y}_r$ is the estimate based on the rth replicate of the simulation, and R is the number of replicates in the simulation (R=1000).

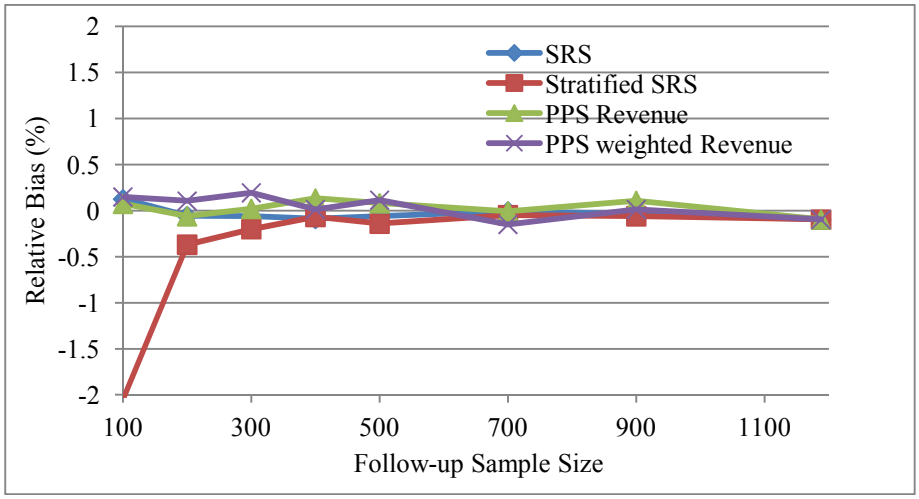
We observed that for follow-up sample sizes of 100 to 400, all the units are finalized with an outcome of 'response' or 'final nonresponse' before the end of collection. For sample sizes of 500 or over, the collection budget runs out before all the units are finalized. More specifically, we observed that at the end of collection, on average approximately 440 cases are finalized and the other units remain in the calling queue with an outcome of 'still in-progress'. The collection budget used for the simulation was just large enough to finalize approximately 440 units. This is a key to understanding the results of the study.

The above observation implies the following. As the follow-up sample size increases from 100 to 400, the number of follow-up respondents increases. On the other hand, as the sample size increases from 500 to 1188, the number of respondents remains roughly

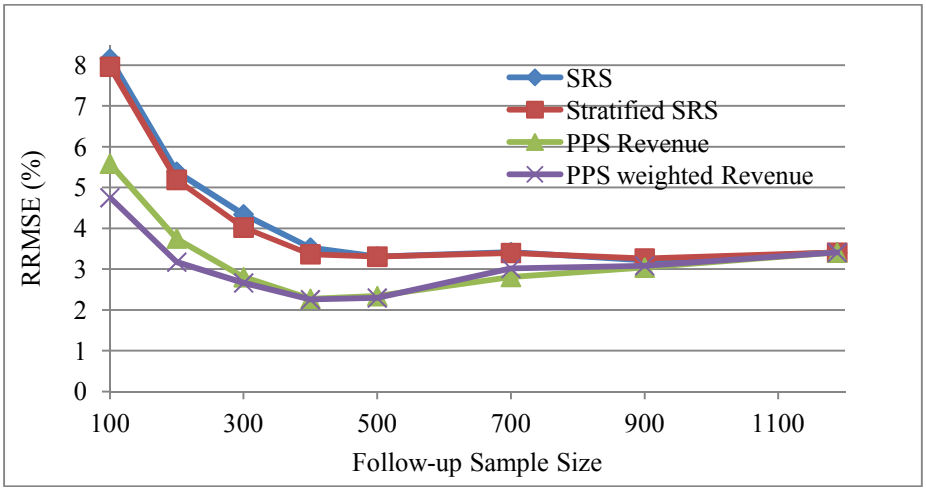
constant while the number of nonrespondents increases, which implies that the response rate to the follow-up decreases. The increase in the number of nonrespondents can be explained by a smaller average number of call attempts per sample unit as the sample size increases.

3.1 Results for Response Scenario 1

This section provides results for the scenario where the probability of response is uniform for both the mailout and the follow-up. In other words, the probability of response is independent of the characteristics of the units. Graph 3.1 is the graph of the relative bias versus the follow-up sample size for the follow-up sample designs described in Section 2.3. Graph 3.2 is the graph of the RRMSE versus the follow-up sample size. Note that the results for the follow-up of all mailout nonrespondents are given by the last point on the graphs (i.e., sample size of 1,188).



Graph 3.1: Relative Bias versus Follow-up Sample Size for Scenario 1



Graph 3.2: RRMSE versus Follow-up Sample Size for Scenario 1

Graphs 3.1 and 3.2 show the following:

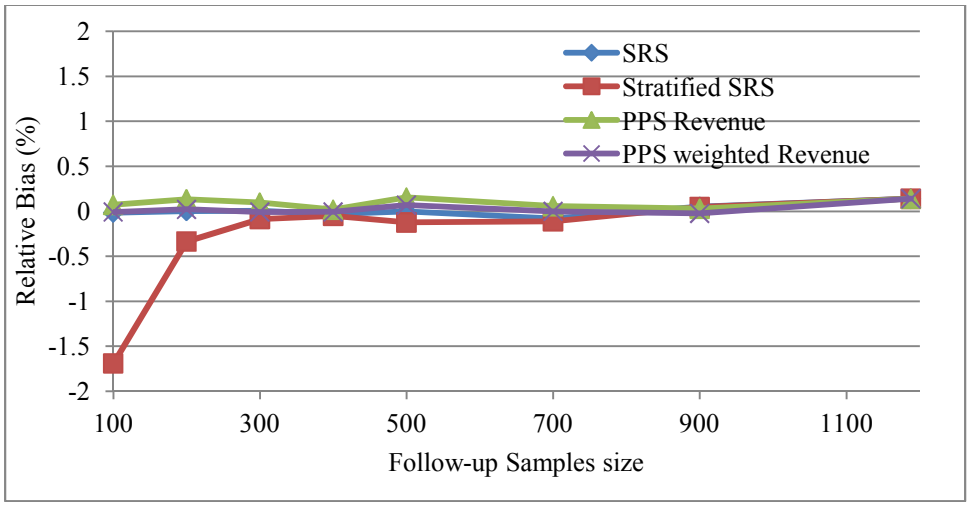
- The RB is approximately zero for all follow-up sample sizes and designs. The only exception is stratified SRS with a follow-up sample size of 100. The allocation

strategy for the follow-up sample does not ensure that at least one unit is selected from each stratum, and therefore, for smaller follow-up sample sizes (e.g., 100), some strata end up with no follow-up sample. This causes a negative bias because the mailout nonrespondents are not represented in strata with no follow-up sample.

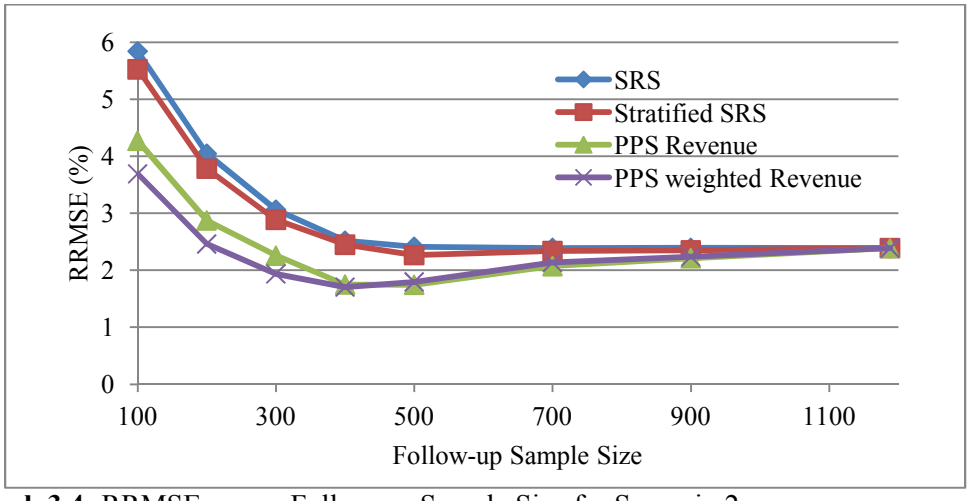
- As the sample size increases from 100 to 400, the number of follow-up respondents increases, and it follows that the RRMSE decreases.
- For sample sizes greater than 400, the number of respondents remains constant so the RRMSE remains constant for the SRS and stratified SRS designs.
- The PPS designs seem to be more efficient than the SRS and stratified SRS designs. However, for sample sizes greater than 400, the gains in efficiency diminish as the sample size increases.

3.2 Results for Response Scenario 2

This section provides results for the scenario where the probability of response is correlated to Sales for the mailout, and uniform for the follow-up. Graph 3.3 is the graph of the relative bias and Graph 3.4 is the graph of the RRMSE.



Graph 3.3: Relative Bias versus Follow-up Sample Size for Scenario 2



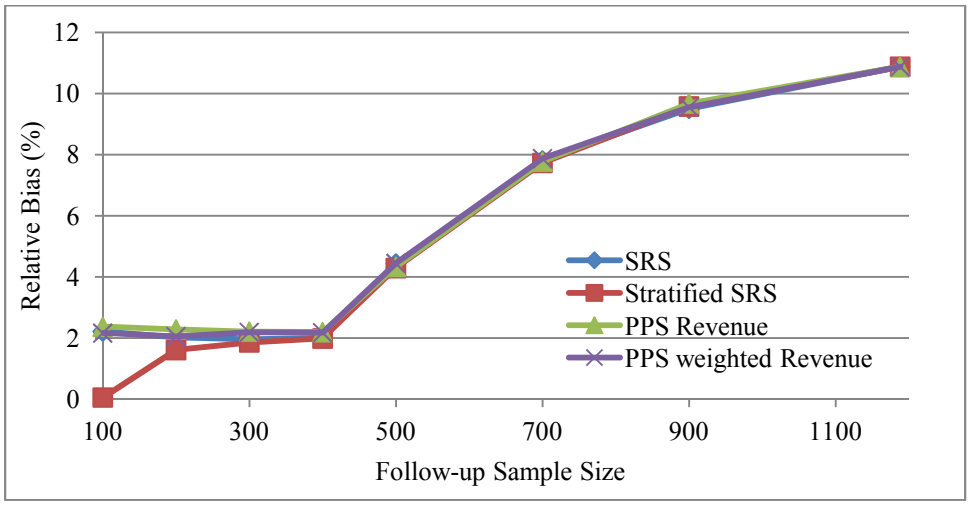
Graph 3.4: RRMSE versus Follow-up Sample Size for Scenario 2

Graphs 3.3 and 3.4 show the following:

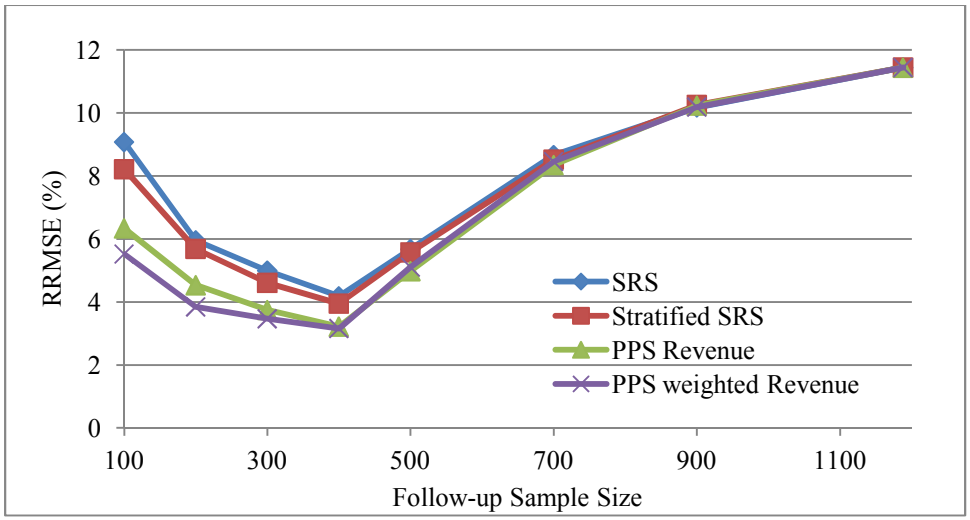
- The results show that if the mailout response probability is correlated to Sales, but the follow-up response probability is uniform, then the bias can be nearly eliminated through the follow-up.
- The results given for Scenario 1 in Section 3.1 apply to Scenario 2 as well.

3.3 Results for Response Scenario 3

This section provides results for the scenario where the probability of response is uniform for the mailout, and correlated to Sales for the follow-up. Graph 3.5 is the graph of the relative bias and Graph 3.6 is the graph of the RRMSE.



Graph 3.5: Relative Bias versus Follow-up Sample Size for Scenario 3



Graph 3.6: RRMSE versus Follow-up Sample Size for Scenario 3

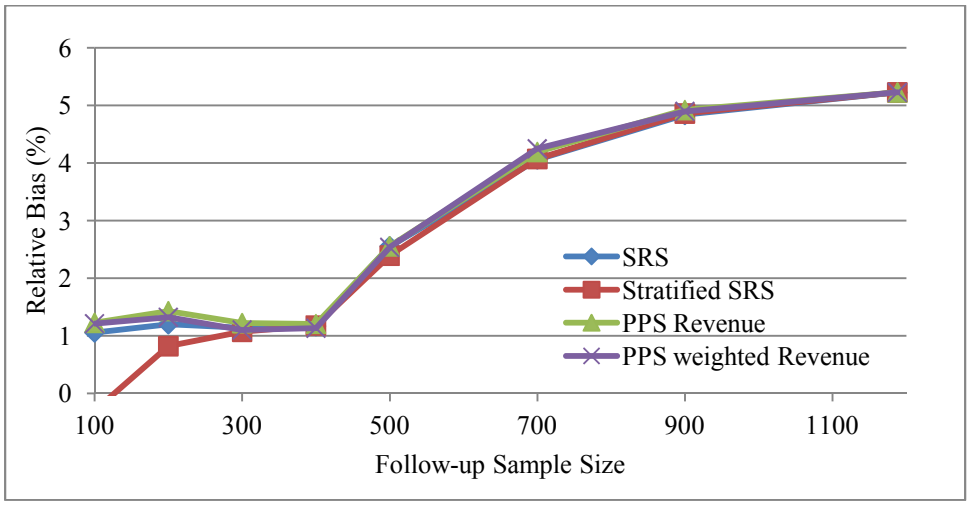
Graphs 3.5 and 3.6 show the following:

- The RB is lowest for sample sizes less than 400, where the units are all finalized before the budget runs out. The lower RB for stratified SRS with a follow-up sample size of 100 is due to strata with no follow-up sample (see Section 3.1).
- The RRMSE is minimized for a sample size of 400.

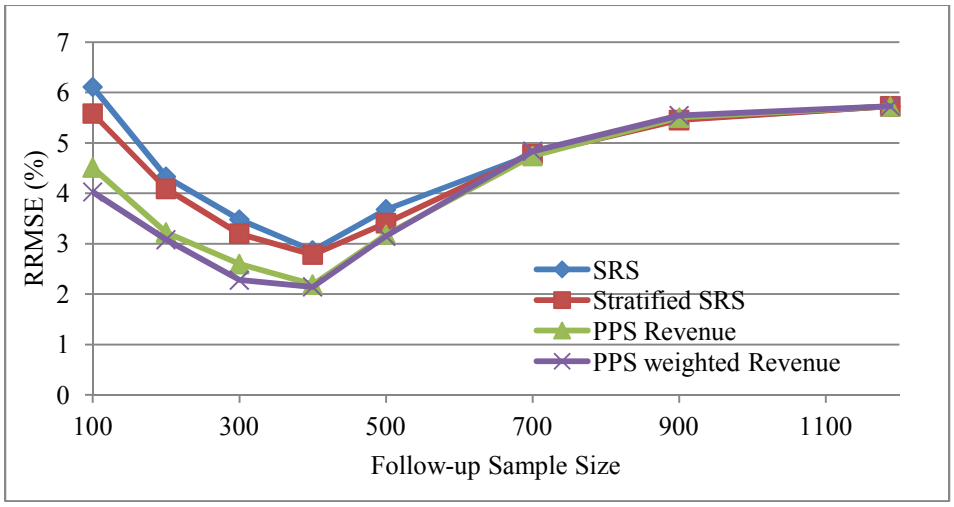
- For sample sizes greater than 400, increasing the sample size increases the nonresponse rate. This explains why the RB and RRMSE increase.
- The PPS designs seem to be more efficient than the SRS and stratified SRS designs. However, for sample sizes greater than 400, the gains in efficiency diminish as the sample size increases.

3.4 Results for Response Scenario 4

This section provides results for the scenario where the probability of response is correlated to Sales for both the mailout and the follow-up. Graph 3.7 is the graph of the relative bias and Graph 3.8 is the graph of the RRMSE.



Graph 3.7: Relative Bias versus Follow-up Sample Size for Scenario 4



Graph 3.8: RRMSE versus Follow-up Sample Size for Scenario 4

Graphs 3.7 and 3.8 are similar to Graphs 3.5 and 3.6. The results given for Scenario 3 in Section 3.3 apply to Scenario 4 as well.

4. Conclusions

The following conclusions can be drawn from the simulation study.

Given the estimator that was used for the simulation, it is the follow-up response mechanism that seems to determine the shape of the RRMSE curve and whether there is a bias in the estimates. This means, for example, that if the mailout response probability is correlated to the variable of interest and the follow-up response probability is uniform, the mailout nonresponse bias is eliminated through the follow-up.

Irrespective of whether the mailout response probability is correlated to the variable of interest, we noted the following:

1) When the follow-up response probability is uniform, the bias is zero. As well, the RRMSE is reduced if the follow-up sample size is large enough that the fixed collection budget is expended.

2) When the follow-up response probability is correlated to the variable of interest, the bias is minimized if the follow-up sample size is small enough that all the follow-up units are finalized at the end of collection. As well, the RRMSE is minimized if the follow-up sample size is small enough that all the follow-up units are finalized at the end of collection and large enough that the collection budget is expended.

The PPS designs appear to perform better than the SRS and stratified SRS designs. However, no attempt was made to optimize the stratification or allocation of the stratified SRS design. Perhaps the performance of the latter design could have been improved.

The above results were based on the choice of arbitrary parameters. If the parameters were changed, we would expect the graphs of the RB and RRMSE to have the same general shape as those obtained in the study.

In practice, the study shows that the nonresponse bias is minimized if sample sizes are such that there is sufficient collection budget to obtain responses from all units, even those that are harder to reach. Business surveys at Statistics Canada currently send all the mailout nonrespondents for follow-up collection and an adaptive collection procedure is used to prioritize cases. We believe that the nonresponse bias could be reduced if subsampling were used in situations where the follow-up collection budget is insufficient to properly handle the volume of mailout nonrespondents obtained. The adaptive collection procedure currently in place could continue to be used to manage the collection of the subsample.

Acknowledgements

The authors would like to thank Joseph Duggan and Joanne Leung for reviewing the paper and providing valuable comments and suggestions.

References

Beaumont, J.-F., Bocci, C., and Haziza, D. (2014). An Adaptive Data Collection Procedure for Call Prioritization. To appear in the Journal of Official Statistics.

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

Hansen, M.H., and Hurwitz, W.N. (1946). The Problem of Non-response in Sample Surveys. Journal of the American Statistical Association, 41, 517-529.

Peytchev, A., Riley, S., Rosen, J., Murphy, J. And Lindblad, M. (2010). Reduction of Nonresponse Bias in Surveys through Case Prioritization. Survey Research Methods, 4, 21-29.

Schouten, B., Cobben, F. and Bethlehem, J. (2009). Indicators for the representativeness of survey response. Survey Methodology, 35(1), 101-113.

Xie, H., Godbout, S., Youn, S. and Lavallée, P. (2011). Collection Follow-up Operation Using Priority Scores For Business Surveys. Conference of European Statisticians, United Nations Statistical Commission and Economic Commission for Europe.