Australian Bureau of Statistics

Use of Paradata to Develop a Framework for Predicting Cost and Statistical Outcomes of Changing Follow -Up Procedures : Case Studies in ABS Business and Household Surveys

Paper for ISDC Workshop Sydney June 2010
Adapted from Internal Papers

Melanie Black and Geoffrey Brent
Operations Research and Process Improvement
Australian Bureau of Statistics

Abstract

This paper contains a summary of work done by Operations Research and Process Improvement (ORPI) on supporting the drive to identify efficiency savings, bu developing a framework for evaluating potential follow-up procedure changes on cost, response and statistical quality. We have used available paradata to develop and link separate statistical models of cost, response, and statistical quality to produce an integrated model to be used for decision-making support in survey operations. This paper provides a summary of the framework developed and it's uses, as well as it's application to two ABS survey case studies.

1. Introduction

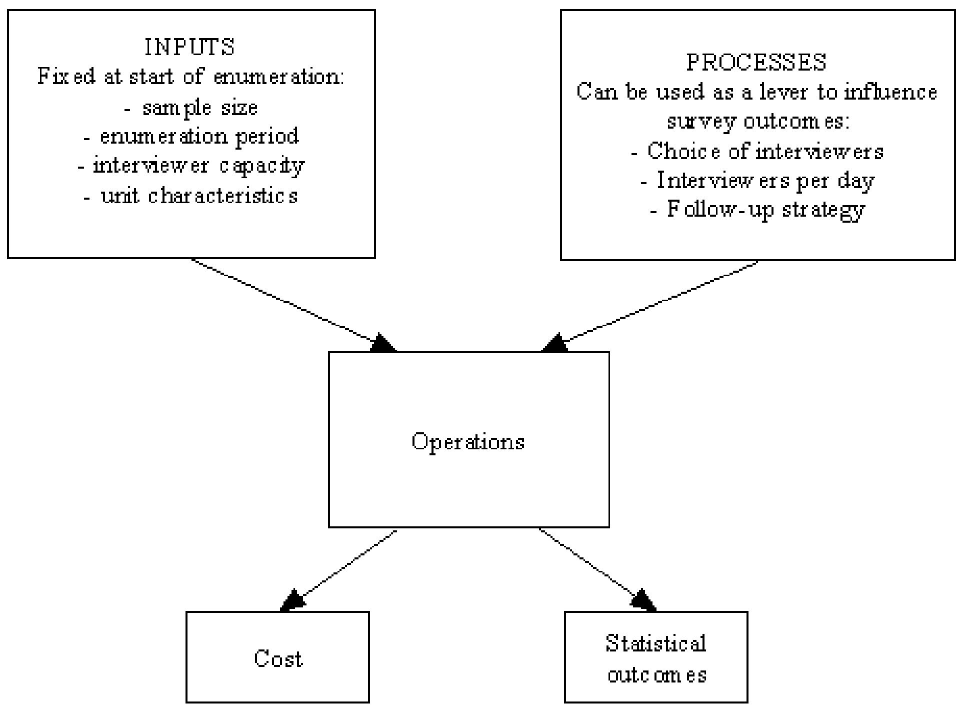
When conducting official surveys, a considerable amount of time and effort is spent in obtaining responses from selected survey units, particularly using interviewers to enumerate such selections or to follow up providers who do not return their survey form by the due date. The cost of obtaining responses from survey respondents is expensive, and therefore strategies are needed to optimise the approach to collecting such information. As it has become increasingly feasible to collect operational *paradata* about statistical data collection activities in real time, analytical and operations research methods are being increasingly used to improve the efficiency and effectiveness of statistical collections. There is evidence to suggest that there has been a general decrease in response rates over the last few decades, and where the response rates have been maintained this has been through significant additional cost and effort. Therefore, an understanding of the trade-off between cost and response / statistical quality is increasingly needed in order to make informed decisions about the follow up processes.

2. A Framework for Cost and Statistical Outcomes

With increasing difficulty in contacting respondents (often leading to increasing costs or decreasing response rates, or both), a better understanding of what drives respondent behaviour, and at what cost, is needed in order to assist informed decision making. We have addressed this key need via the development of statistical models with the ability to forecasting response outcomes, as well as cost, resulting from various allocations of effort. In order to make use of these models transparent and accessible, we have developed a cost (effort) / response / quality (measured by estimates of bias) framework to understand the relationships between these different aspects of survey operations.

Our aim has been to establish an interactive tool where the impact of changing operational procedures on costs, response rates and survey output can be investigated on an ongoing basis. Ultimately, it was hoped that this tool would enable the optimisation of operational procedures to minimise costs, while ensuring an acceptable level of response and statistical quality. The recent availability of paradata, including timing of calls/visits and response, for ABS surveys has allowed us to undertake better analyses of costs required to achieve a particular statistical outcome. This richer set of information about survey inputs which can help inform process decisions in a more scientific manner. To do this, a cost/response/quality framework has been developed:

Figure 1: Cost/response/quality framework diagram



The use of such paradata allows an analysis of changes to costs and statistical outcomes (response and bias) without disrupting the real time business operations, as well as understanding what further information could be sought to better inform future decisions.

The focus of developing such a framework is ensuring sophisticated models exist to reliably estimate the outputs so that questions, such as below, can be asked:

- What if I conducted part of my survey as a telephone interview and part of it face to face?
- How would my costs and output change if I were to increase or decrease my target response rate?
- How should I run my survey to achieve the results that I want for the least cost?

Questions such as these are not readily answerable from the individual cost models themselves. Respondents are not all the same and neither are interviewers and these factors play a significant part in the tension between cost and statistical outcomes such as response rates and quality of estimates.

Related topics are also of ongoing importance to the planning and conduct of both ABS household and business surveys - for instance, the effects of workload geography (clustering) and interviewer distance from the workload on survey costs and accuracy, or the way in which increased load from one survey can affect the costs of another survey by reducing interviewer efficiency.

The aim of this work has been to:

1. To develop a mathematical model for costs which
   - accurately reflects costs of operational procedures involved in data collection
   - is transparent
   - enables 'what if' scenarios to be assessed, specifically including different field procedures
   - can be readily maintained and be used for monitoring costs
2. To develop measures of estimate quality ('bias') for various types of estimates.
3. To understand the relationship between operational procedures and fitness for purpose.

Although this generic framework applies to both household and economic surveys, the models themselves are quite different owing to different cost structures, follow-up procedures, respondent characteristics, and so on. Details of the models themselves are not included in this paper (as some are still being finalised), but can be provided upon request.

To complement the model framework, real time experimental trials will be conducted, where a subset of the providers would receive the new procedures, and these would then be compared to the results using the current procedures. The use of trials would be a way of demonstrating whether the anticipated savings are actually realised, and if the changes are practical and implementable. A by-product of running a number of trials is that the predictions from the framework can be confirmed, or alternatively highlight deficiencies in the framework which will guide updates to the models.

Another use of the trials would be to collect information (or paradata) that isn't currently collected for the current set of follow-up procedures. For example, we don't currently have any cost or response information available on the use of the internet for sending reminder letters. This information could then be extrapolated where possible and added into our current framework for future operational change considerations.

The potential real-time field test scenarios include:
Business Surveys
- applying a best time of day, and best day of week to call, or otherwise targetting calls more effectively
- assessing reminder letters, including changes to the seriousness of the wording, assessing calls earlier in the enumeration period in conjunction with sending a reminder letter for previous non-responders, and the optimal length of time between reminders, to maximise their effectiveness and eliminate waste
- introducing a "call capping" strategy, to save resources

Household Surveys
- Modifying callback strategy based on unit characteristics, to allocate effort where it will be most productive.
- Changing contact strategy (e.g. time of day) to maximise effectiveness
- Modifying choice of interviewer according to distance to workload to reduce travel costs.

The following two sections outline examples of applying the linked model to ABS Business and Household collections respectively. The results below are intended only as demonstrations of how the models work and not as accurate results, as they are based on experimental versions of the models.

3. Case Study - Business Surveys

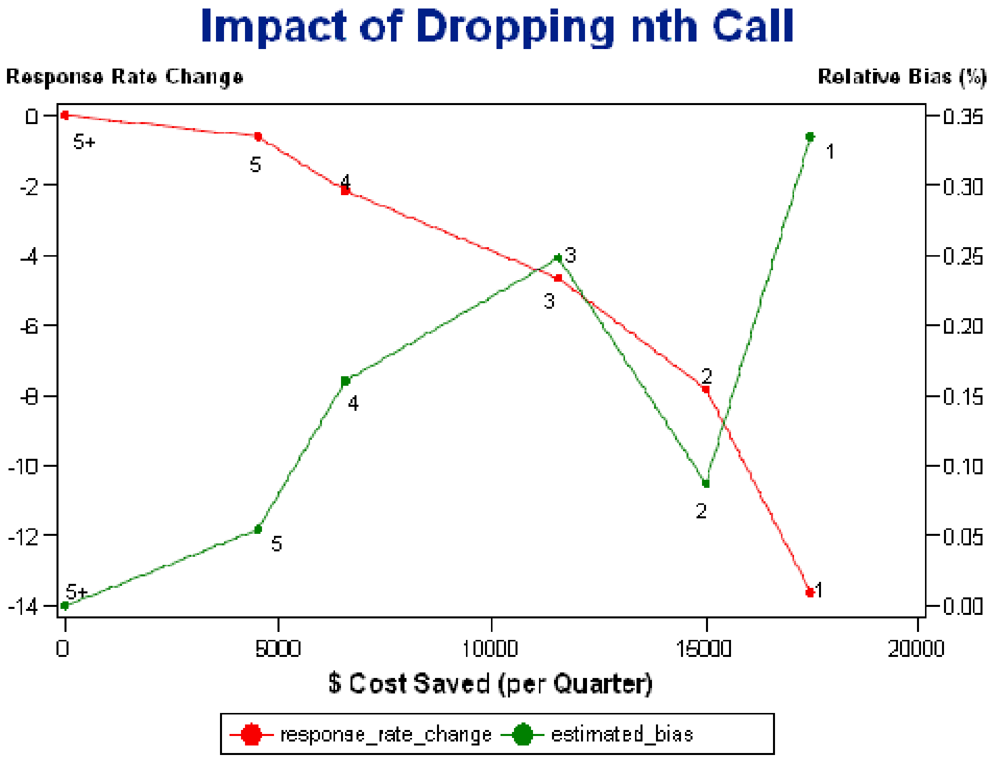
ABS business surveys are predominately mail-out mail-back collections supplemented by follow-up procedures involving both reminder letters and telephone follow-up. Paradata is available on the number and timing of callbacks and other interactions between the ABS and each respondent, and on response status and timing. This paradata has been used to develop models as discussed above, which can then be used to predict the outcome of changing follow-up procedures. This section outlines a case study of the application of these models.

The scenario being considered here is the impact of dropping the nth call, as it is known that despite calls generally become decreasingly effective as more are made, some providers receive large numbers of calls (and still do not respond). Using the probability of response model, the average (over the 4 quarters) predicted response rate was calculated with the number of outbound calls is capped at 1, 2, 3 etc per provider. 'Crucial' providers (as identified by subject matter areas) for whom follow-up would not be ceased without a response did not have their calls capped and still receive their usual number of calls. This response rate is then used as an input into the cost model to obtain a yearly cost of the scenario, so that the cost savings obtained by capping the calls can be estimated.

The quality of the estimate in this case is the 'relative bias', which is calculated as the % difference between the average four quarter estimate over a sample with the given response rate and the average four quarter estimate over the full sample. For the scenario where no calls are dropped, the relative bias is set to zero, since it is the basis for comparisons. It should be noted that the bias given here is only indicative as it is based on an earlier version of the model. (The updated version includes not only the expected change in the estimate resulting from a strategy change, but also the change in the standard error).

Figure 2 shows the impact of capping at the nth call (dropping calls above n). The current scenario ('5+') is the basis for comparison, that is, the response rate change is 0, and the 'relative bias' is also 0. If calls were capped at a maximum of 5, then this linked model would forecast a response rate drop of 0.6% and a relative bias of the estimate of 0.05%, and the resulting cost saving (for telephone operations only) would be about $4,500 per quarter. As the maximum number of calls declines, the response rate declines more and more rapidly. The bias fluctuates considerably, with a 'chance' low bias outcome at 2 calls (expected to be fixed in improved versions of the model), but generally a fairly small but steady increase in bias as the number of calls declined. Using this kind of integrated model, a decision-maker could start with a requirement on any of the three axes, and determine (from this class of IFU strategies) what procedure to persue. For example, if a cost saving of $10,000 was required, a call cap could be set at 3 (perhaps allowing a small number of "special" providers to receive more, since the saving at this point would be slightly higher) with an expected decline in response of 4.7% and an increase in relative bias of 0.24% (i.e. a small drop in estimate quality).

Figure 2. Impact of dropping calls on cost, response and bias.



This kind of approach could be used to evaluate more complex scenarios, for example, capping calls at different values for different kinds of providers so that significant units will be followed-up more heavily than less significant units, or to examine the joint impacts of reminder letters and telephone follow-up for different types of providers.

4. Case Study - Household Surveys

The Monthly Population Survey (MPS) is the main vehicle for ABS household surveys. Selected households remain in the MPS sample for eight months. For the first month in sample (FMIS), households are typically approached and interviewed face-to-face. Non-first-month (NFMIS) households are generally interviewed by telephone, but some require a face-to-face approach, either by request of the householder or because it's not possible to contact them over the phone.

Interviewers are paid for their time and mileage (MVA), so attempting face-to-face households contributes a large part of MPS operational costs. If a household doesn't respond after repeated contact attempts, it will eventually be abandoned as 'nonresponse' for that month. One important question is how many contact attempts should be made before abandoning a household, since additional contact attempts will be increasingly costly but decreasingly effective in achieving response. At present, the contact requirements are the same for FMIS and NFMIS households, but it might be appropriate to treat them differently.

To explore this, the linked model was used to make predictions for several possible variations to MPS callback procedures. For comparative purposes, the first scenario shown is a baseline scenario similar to current MPS practice; others modify callback efforts for face-to-face households.

Table 1: callback scenarios

Description	Max calls before abandonment (ex-met FMIS)	Max calls before abandonment (ex-met NFMIS)	Max calls before abandonment (metro FMIS)	Max calls before abandonment (metro NFMIS)	Identifier (graph)
Baseline scenario	5	5	8	8	Base
Reduce NFMIS callbacks by 1 call	5	4	8	7	-1N
Increase NFMIS callbacks by 1 call	5	6	8	9	+1N
Reduce FMIS callbacks by 2 calls	3	5	6	8	-2F
...by 4 calls	1	5	4	8	-4F
Increase FMIS callbacks by 2 calls	7	5	10	8	+2F
...by 4 calls	9	5	12	8	+4F

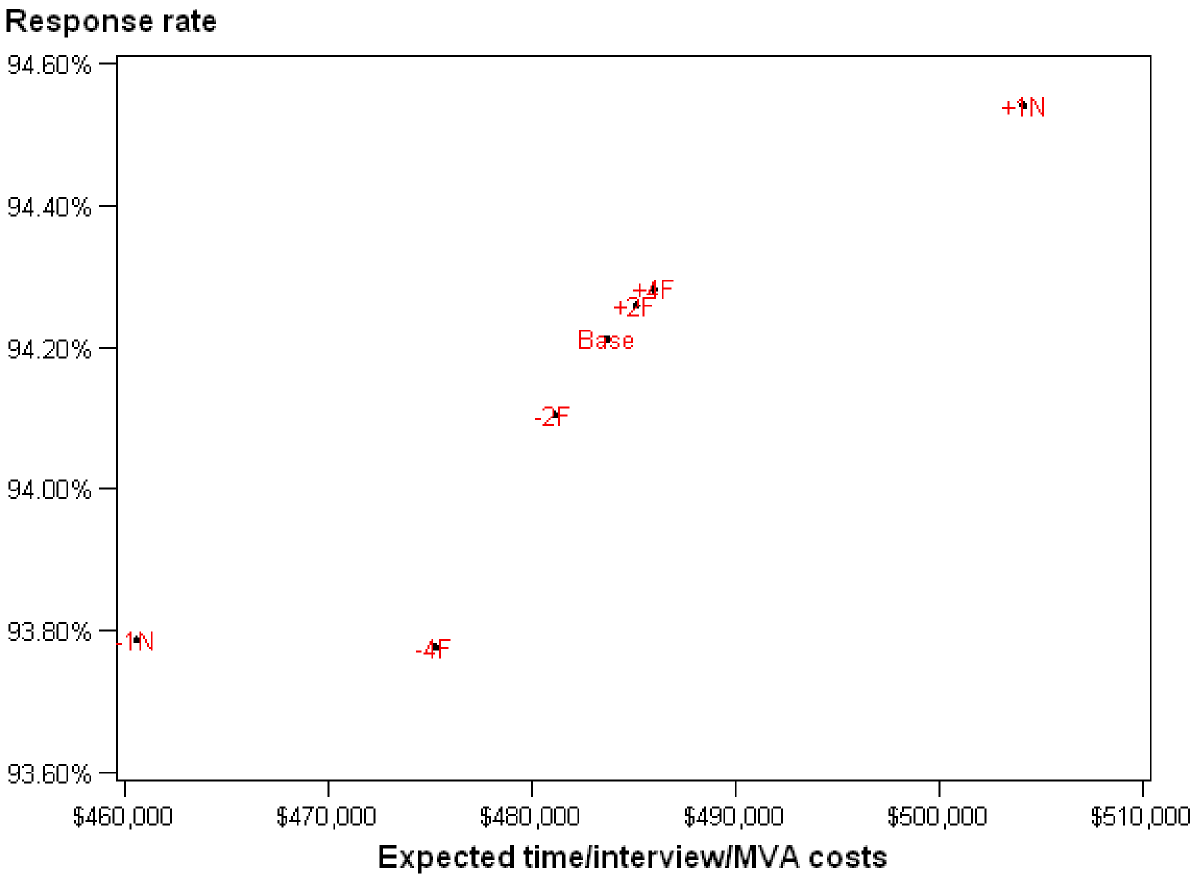
The model produces predictions of cost, response, and experimental estimates of bias in employment estimates for each scenario over one month. Fine-level figures are available but for brevity, only national-level results are shown here:

Table 2: Response, costs, and quality(bias) for callback scenarios

Scenario	Response rate	Expected bias	Expected costs
Base	94.21%	0.0023%	$483,681
-1N	93.79%	-0.0038%	$460,567
+1N	94.54%	0.0023%	$504,106
-2F	94.10%	0.0021%	$481,137
+2F	94.26%	0.0017%	$485,102
+4F	94.28%	0.0017%	$485,991
-4F	93.78%	0.0012%	$475,253

It can be seen from table 2 that changing operational practices gives differing outputs, and trade-offs between cost, response and bias need to be considered carefully when deciding on the best approach.

Figure 3: Response vs costs for callback scenarios

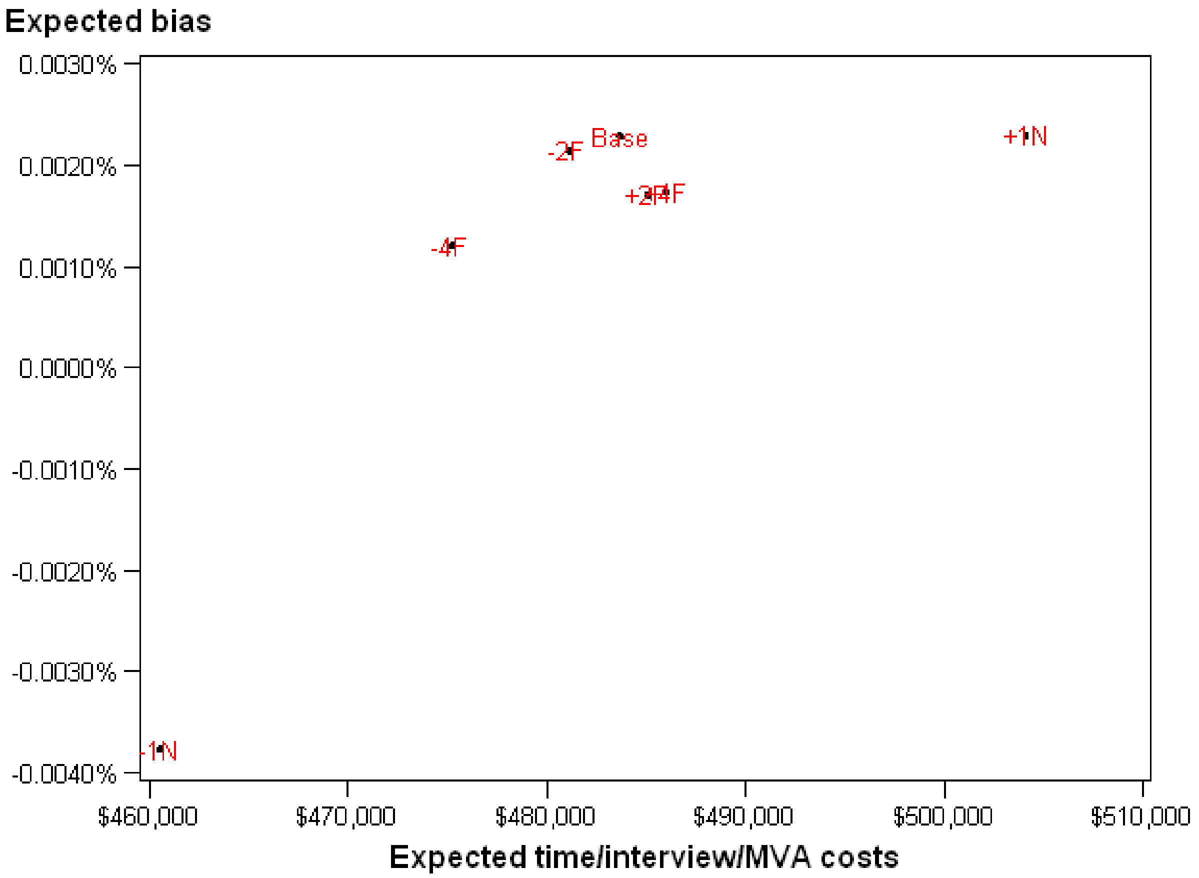


The linked model shows that FMIS and NFMIS calls have quite different cost/benefit ratios. For example, if we were to cut back NFMIS callbacks by one call (scenario -1N), we would expect a 0.42% drop in response rate relative to the base case, with a cost saving of $23,114. By contrast, although severe FMIS cutbacks (scenario -4F) produce a similar response rate, the cost saving is much smaller ($8428). In terms of response rate per dollar spent, FMIS callbacks are giving good value for money. Based on this analysis, it may be desirable to slightly reduce NFMIS contact effort in order to devote more effort to FMIS households.

This represents a combination of two effects: it's more expensive to approach NFMIS households (due to a lower level of clustering), and those that are approached have poor response rates (the "easy" households are shifted to telephone interview after their first month).

Expected bias can also be considered. The following plot shows predictions of bias for a particular estimate for the same scenarios considered above:

Figure 4: Bias and response vs costs for callback scenarios



In this example, although scenarios -1N and -4F had similar response rates, response rate alone doesn't guarantee data quality - the absolute bias for -1N is substantially larger. The bias model used for this analysis was an experimental 'placeholder' that is not expected to be accurate, so it is not appropriate to draw operational conclusions from this plot - but clearly, with a more reliable bias model, this sort of analysis would be highly useful in deciding between possible scenarios.

5. Conclusions

In conclusion, a model framework has been developed for assessing the impacts of changes to various follow-up strategies on operational costs as well as on the quality of statistical outcomes including response rates and bias. Through case studies for both business and household collections, this framework has been proven to be an effective tool for making informed decisions before embarking on real time changes to operations, so that the anticipated impacts are known and confidence in these impacts to the survey output estimates can be understood by decision makers. This framework will be used in guiding changes to follow-up procedures, in particular trials of alternative processes, in the future. The framework will also be updated as survey operations change to reflect emerging technologies and the opportunities they bring. More broadly, using scientific methods to increase our understanding of how changes in operational procedures impact on costs and statistical outcomes like response rates and the quality of survey outputs, allow survey managers or other senior executive staff in the ABS, to be more comfortable in making informed decisions about operational efficiencies in IFU practices.