Increasing Data Quality for Business Surveys without Impacting Collection Effort

Jessica Andrews, Matei Mireuta, Loïc Muhirwa and Pierre Daoust.

Statistics Canada, Business Survey Methods Division, RHC 22, 100 Tunney's Pasture Driveway, Ottawa, Ontario, K1A 0T6. Jessica.Andrews@canada.ca, Matei.Mireuta@canada.ca, Loic.Muhirwa@canada.ca, Pierre.Daoust@canada.ca

Proceedings of the 2018 Federal Committee on Statistical Methodology (FCSM) Research Conference

Abstract: In 2016 Statistics Canada implemented a dynamic collection methodology for the prioritization of collection follow-up activities, needed to address non-response and failed edits issues, for a number of business surveys. The Quality Indicators/Measures of Impact (QIMI) methodology reprioritizes units during active collection in order to concentrate efforts on domains where the quality of estimates is insufficient, where quality can be measured in terms of coefficient of variation (CV) or a variety of other indicators. This prioritization leads to collection effort being directed towards the units whose information will have the most impact on the quality of the survey estimates and reduces the number of contacts for respondents with less impact. In the future the potential improvement to QIMI due to using collection paradata related to the progress of respondents through the surveys' electronic questionnaires should improve the targeting of collection effort.

1. Introduction

Statistics Canada has recently introduced the Integrated Business Statistics Program (IBSP). The aim of this program is to achieve greater efficiency in processing Statistics Canada's business surveys while maintaining or improving the quality of data. The IBSP will standardize processes and methods for over a hundred and twenty surveys, of which approximately ninety have currently been integrated or are in the process of being integrated. A key component of the IBSP is the active collection management tool which is used to provide regular updates to collection on follow-up priorities for both non-response and failed edit cases to improve surveys' data quality. This tool is used for all surveys with large enough samples to benefit from collection prioritization, currently approximately sixty surveys. The tool creates priority lists by assessing data quality for key estimates and prioritizing the collection of units with the greatest potential for improving the quality of these estimates. This paper will discuss the impact that this tool has had on survey quality.

2. Active Collection Management Tool

The IBSP system differs from a standard survey processing system in that it is non-linear and uses a circular rolling estimate approach as explained in Turmelle et al (2012). Due to this rolling estimate approach, estimates for the survey are created during active collection which allows the active collection tool, QIMI (Quality Indicators/Measures of Impact), to assess current data quality as well as to predict which units would most improve the survey quality. A thorough explanation of the QIMI process can be found in Andrews et al (2016), Godbout et al (2011), and Turmelle et al (2014).

To run QIMI, each survey identifies key estimates, choses quality measures, and sets quality targets which are used to assess survey data quality throughout collection. More specifically, with each rolling estimate all collected data are fully integrated, edits are run, non-respondents and partial respondents are imputed, and estimates are created as if the survey collection had been completed. At this point, the active collection tool is used to assess the quality of estimates by comparing quality indicators to the pre-set quality targets. Quality in QIMI is currently measured using two indicators. The first is a key variable weighted response rate, where each variable's response rate is weighted by its own values and the final estimation weights (see Andrews et al 2016) and the second is the relative deviation from predicted value indicator (see Mireuta et al 2017) which assesses the impact of the difference between a unit's responses and predicted values.

Where a key estimate's quality has not met its pre-set quality target, unit measures of impact are calculated for each quality measure. The exact formulas can be found in Andrews et al (2016) and Mireuta et al (2017). Essentially, the measure of impact yields the change in quality which would be expected at the key estimate level should the unit be treated, in this case either collected for the key variable weighted response rate or corrected in the case of the relative deviation from predicted value. The measures of impact are then combined across the survey into a global measure of impact for each unit and quality measure, using the relative importance of the key estimates and the distance from the quality target to weight key estimates.

Finally, the global measures of impact for the quality measures are combined into a single priority list which is used for collection purposes. This list is broken down into six groups with one group kept for units that are either predicted to have no impact on any key estimates which have not met their quality targets or are ineligible for collection due to either being fully collected or having special collection instructions. The remaining units are given a priority from 1 to 5, with units in priority groups 1-3 receiving the most collection effort.

The data from QIMI is also made available to analysts at each rolling estimates run. This enables them to quickly start analysis of any units that are ineligible for collection which are shown to be problematic by one of the measures of quality. Analysts can then assess which domains have sufficient quality to begin their analysis prior to the end of the collection period.

3. Impact

3.1: Conceptual

With moving to QIMI, there is a large change in how quality is measured during the collection period. The standard measure of quality previously used for most of the business surveys was a weighted response rate indicator where a single weight (often, expected revenue multiplied by the design weight) was used to assess if collection had been sufficiently successful. QIMI, however, as explained in the last section, allows the assessment of multiple key estimates using more than one quality indicator. Most surveys identified their key estimates by choosing four to six important survey variables which are used for the survey's chosen key domains based on a few classification variables (for example geography and industry). As such, a considerable number of key estimates are tracked during collection for surveys, and QIMI is prioritizing follow-up activities to achieve quality targets for all these key estimates. Once a target is met then collection effort is redirected towards domains and units which will most improve quality where targets are yet to be met. Thus, as the emphasis is no longer focussed on a single variable, usually revenue, it would be expected that there would be an improvement in data quality across the entire survey, thereby improving the survey relevance to users. In addition, the automated QIMI process makes the active collection process much more effective in detecting and responding to quality issues for the survey than the previously used process.

3.2: Response rates

An original concern with running QIMI for surveys was that reprioritizing collection towards smaller domains might have a negative impact on overall response rates for the survey and units important to data quality at the national level might be missed. For almost all surveys in the study, improvements were found not only in terms of quality of key estimates but in the single measure previously used as the main gauge of collection success. In the few cases where results with QIMI were not better, response rates were close to those achieved previously. Figures 1 to 4 compare three collection years. In 2015, all but three of the first wave of thirty-six surveys used only a prioritisation list created before the start of collection based on Canada level estimates. The other three surveys ran a pilot version of QIMI for 6 weeks using only the key variable weighted response rate quality measure. In collection year 2016, all surveys began by using a prioritisation list derived prior to collection at the province and industry levels and began their dynamic use of QIMI about 3 months into their collection period. Collection year 2016 used both the key variable weighted response rate indicator and a relative deviation to predicted value indicator. In collection year 2017, all surveys used QIMI dynamically a month into their collection period at the start of active collection. Though the same two quality indicators were again used, the formula was improved for the relative deviation to predicted value.

Figure 1 compares economically weighted response rates between 2015 and 2017 for a representative survey. In the graph, the starting point for 2016 is the point at which QIMI runs started and in 2015 it is the point at which Rolling estimates began. As is clearly shown in the graph, the weighted response rates were significantly better for the two later years when QIMI was used. In collection year 2017, active collection was ended earlier for the survey, hence a slightly lower weighted response rate was achieved. However, at the end of collection the survey was ahead of where it had been at the same point for collection year 2016. Similar results were found for the rest of the first group of QIMI surveys, with some variation in whether the 2017 response rate was slightly higher or lower than that of 2016. As this group of surveys share a collection budget, there are some weeks where certain surveys would have received no collection effort, as can be seen in figure 2.

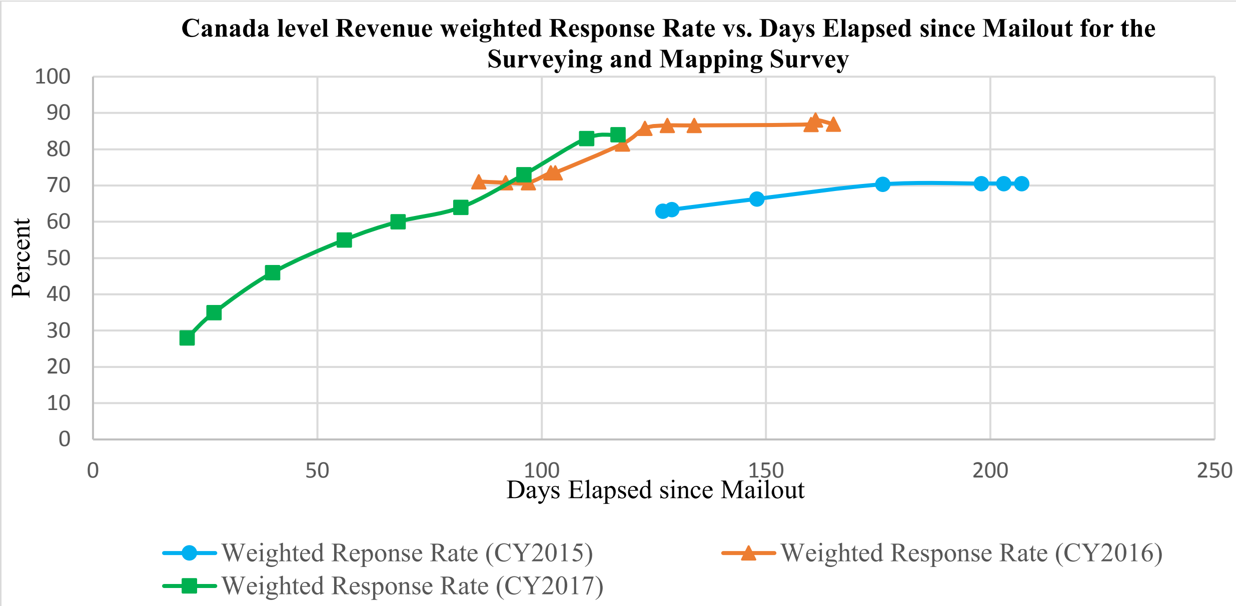

Figure 1. Comparison of Canada-level revenue-weighted response rates in collection years 2015 through 2017.

Figure 2 compares the percentage of collection targets for the key variable weighted response rate which were met across the three collection years, for the same representative survey as Figure 1. In collection year 2015, there was no QIMI for this survey and fewer than sixty percent of targets (as set in 2016) were met. In 2016 the full QIMI methodology was used, however no updates were made to prioritisation lists until the three month mark. In 2017, quality targets were increased for this survey, so although the percentage met was very slightly lower at the end of active collection, the actual quality demanded and achieved was slightly higher.

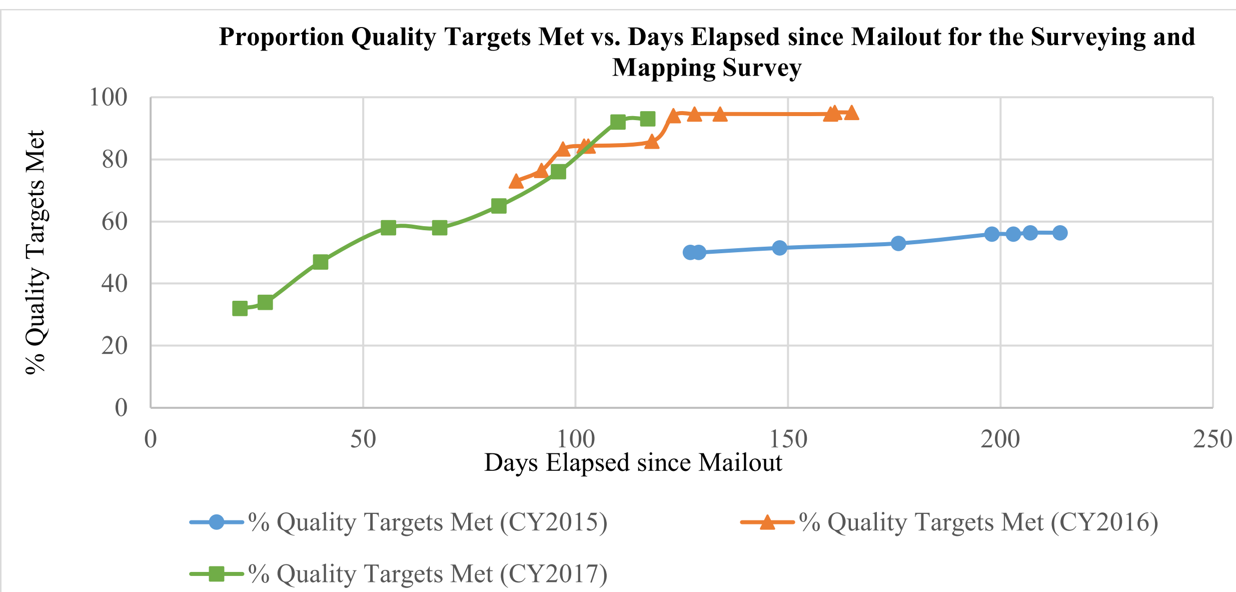

Figure 2. Comparison of Percentage of Quality targets met between collection years 2015 through 2017

Figure 3 compares mean local quality distances for 25 surveys for which sufficiently detailed data was saved for dates close to the end of active collection in 2015. The local quality distance was calculated for a survey as the average relative distance between the value of the key variable weighted response rate and the response rate target. For 2015 the 2016 targets and key estimates were used to judge the local quality distance. For all but three surveys this mean value was lower for collection year 2017 than in 2015, even though many quality targets were higher than those used for the 2015 calculations and active collection finished earlier.

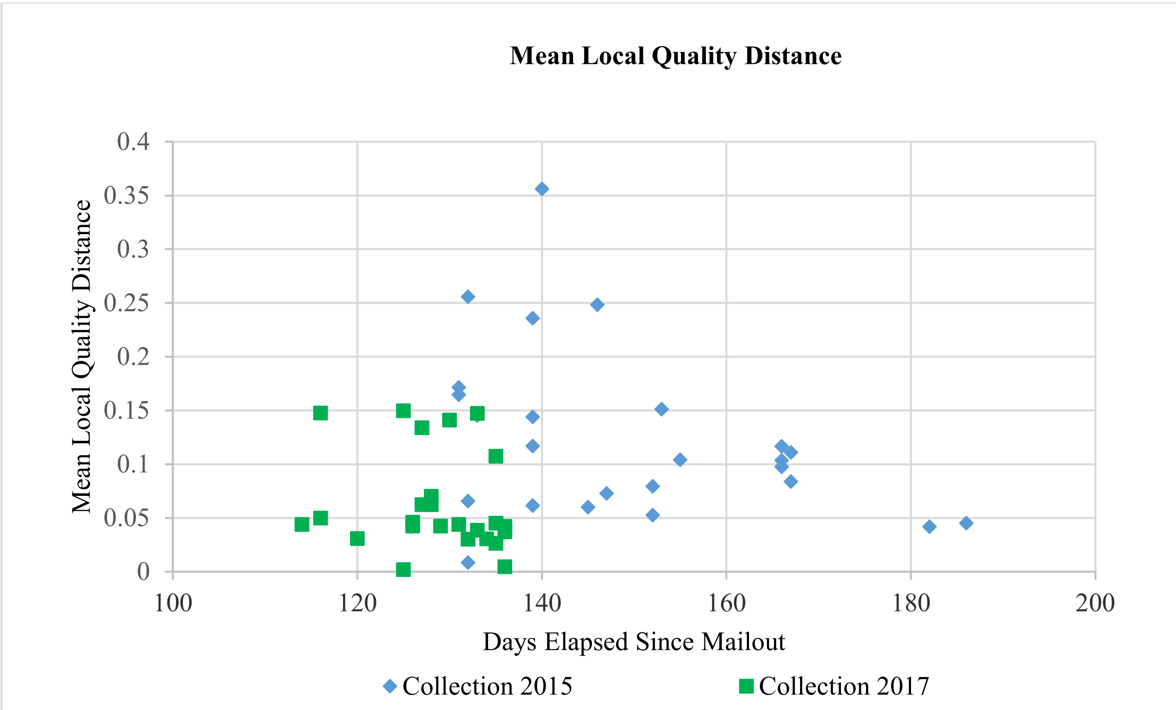

Figure 3. Comparison of Mean Local Quality Distances for key variable weighted response rates for 25 surveys between collection year 2015 and collection year 2017.

In this fourth figure for a different representative survey from Figures 1 and 2, it is seen that in terms of cumulative interviewer effort (time in minutes), there was a significant increase in quality for the two years where QIMI was used. Again, this was a survey where collection finished earlier in 2017 and quality targets were raised from the previous year. Thus, improvements in quality for the surveys are not due to an increase in effort and show that the level of quality achieved in 2015 could have been reached in subsequent years with expending less effort for most surveys. It is, however, possible that the increase is due in some part to respondent's increased level of comfort with electronic questionnaires and not just the improved prioritisation due to QIMI.

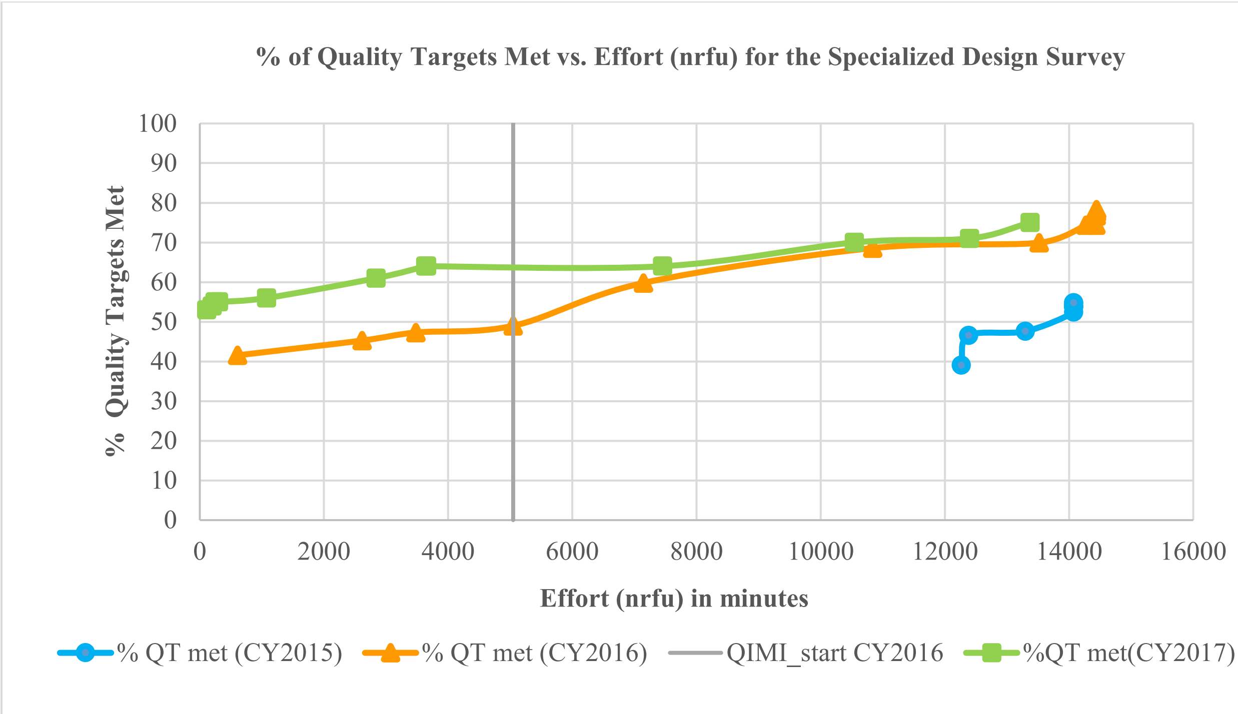

Figure 4. Comparison of quality targets met versus the cumulative non-response follow up effort (nrfu) used for collection years 2015 through 2017

3.3 Timeliness

Another pressing concern with respect to quality at Statistics Canada is the timeliness of the data produced. For the group of 36 surveys considered in this paper, all but one in collection year 2015 published more than 12 months after the end of the reference year 2014. For these surveys, it is desirable that data be published within a year of the end of the reference year and a study was conducted aiming at decreasing the collection period in both 2016 and then significantly in 2017 in order to meet this goal. In 2017, for reference year 2016, a third of these surveys were able to meet this target. The improved prioritisation of units due to QIMI combined with shifts in timing of collection effort led to a significant shortening of the active collection period for these surveys. As indicated in the previous section, the quality achieved for surveys did not suffer with the shortening of collection periods and many of the surveys could have stopped collection even earlier and achieved the previous year's quality. Thus, the aim in the future is to shorten collection periods further when a sufficient percentage of QIMI targets are met or a quality distance (calculated using QIMI local quality indicators and targets) is small enough. This shortening of collection, with the resultant decrease in collection effort, will not only improve timeliness of surveys but also reduce collection costs.

4. Future improvements

QIMI continues to evolve as there are constantly new surveys that are being added to the IBSP with new features requiring development. In 2018, there are plans to add three new indicators to QIMI which can be used in combination with those currently in use depending on survey requirements.

For surveys where the data will be used to update the population frame or for modelling purposes, there is a requirement that a sufficient number of small respondents are collected. This has led to the introduction of two new response rate indicators. The first of these is a weighted response rate indicator where the weight is the final estimation weight of the unit. This indicator will lead to the prioritisation of units that are representative of a large number of other units. When used in a balanced way with the key variable weighted response rate, collection priorities will be the units responsible for the largest percentages of estimates and those representing the most units. A second indicator which will be added for some surveys is an unweighted response rate, an indicator that gives every unit in a domain the same weight. These indicators, when balanced with a key variable weighted response rate, will ensure that a sufficient number of units in a domain are targeted as well as a sufficiently high proportion of the estimate. For a census, this second indicator is the same as the weighted response rate indicator.

The third indicator that will be implemented in the system is an indicator of total coefficient of variation(CV) of the estimate. This will be possible with the addition to the IBSP system of a functionality to calculate the variance due to imputation (Beaumont and Bissonnette 2011). This addition to the IBSP system will allow QIMI to prioritize active collection processes based on a more holistic quality measure for the key estimates. Using this total CV indicator will allow QIMI to target units for which the imputation model has high variability. This may mean that some smaller units may well be prioritised along with the largest units but should reduce the total CV of estimates and should increase the reliability of survey results. It may also mean that some of the units that are currently prioritised due to their large contributions to key estimates may have low priority for collection if they can be accurately imputed, for example if strong auxiliary data is available for the units.

The other aspect which should improve QIMI is using paradata on respondent behaviour. There are currently two sources of available paradata on respondents: one which provides all contacts with a respondent as well as their outcomes, and the other, which logs all times spent on the electronic questionnaires. Using these data sources to add to the QIMI methods for prioritisation it should be possible to increase or decrease priority as suggested by the paradata. For instance, a unit that has previously needed several phone contacts before replying could have its priority increased, compared to units with similar contributions to survey quality. On the other hand, a company that has spent a significant period of time accessing their electronic questionnaire and visited all key pages might be deprioritized as it would be likely that much of the necessary information had already been entered. It might also be possible to identify units which were stuck in the questionnaire, for whom a phone contact might make the difference between a reply and getting very little data.

5. Conclusions

This paper has shown that the introduction of QIMI as part of the IBSP system at Statistics Canada is already having a positive impact on surveys. It is helping to obtain higher response rates, improve data quality and also publish earlier. The system is, however, constantly being improved and it is expected that further gains in data quality and timeliness as well as reductions in collection effort and cost can be made over the next few years.

6. Acknowledgements

The authors would like to thank Matthew Brodka for his contributions to creating the graphs in this paper. We would also like to thank Wesley Benjamin and John Stardom for reviewing this paper and their helpful comments.

6. References

Andrews, J., P. Daoust, and M. Mireuta (2016). Small Domains and Sparse Variables: Challenges with creating an Active Collection Strategy. Proceedings of the Survey Methods Section, SSC Annual Meeting, June 2016.

Beaumont, Jean-François and Joël Bissonnette (2011). Variance Estimation under Composite Imputation: The Methodology behind SEVANI. Survey Methodology, December 2011 171, V0l 37, No. 2, pp. 171-179. Statistics Canada, Catalogue No. 12-001-x.

Godbout, S., Y. Beaucage, and C. Turmelle. (2011). Achieving Quality and Efficiency Using a Top-Down Approach in the Canadian Integrated Business Statistics Program. *Conference of European Statisticians*. Ljubljana, Slovenia.

Mireuta, M., J. Andrews, and P.Daoust (2017). The Relative Deviation from Predicted Values as a Tool to Prioritize Units for Failed Edit Follow-Up in IBSP. Proceedings of the Survey Methods Section, SSC Annual Meeting, June 2017.

Turmelle, C., S. Godbout, and K.Bosa (2012). Methodological Challenges in the Development of Statistic Canada's New Integrated business Statistics Program. Proceedings of the ICES IV Conference. Montreal.

Turmelle, C., S. Godbout, K. Bosa, and F. Mills, (2014). A Quality Driven Approach to Managing Collection and Analysis. *Proceedings of the European Conference on Quality in official statistics*. Vienna.