



Reducing collection effort while maintaining data quality in business surveys

Jessica Andrews

Matei Mireuta

Loic Muhirwa

Pierre Daoust

FCSM March 7th 2018



Outline

- The Integrated Business Statistics Program (IBSP)
- How the Quality Indicator/Measure of Impact (QIMI) tool works
- Impact on surveys
- Future steps

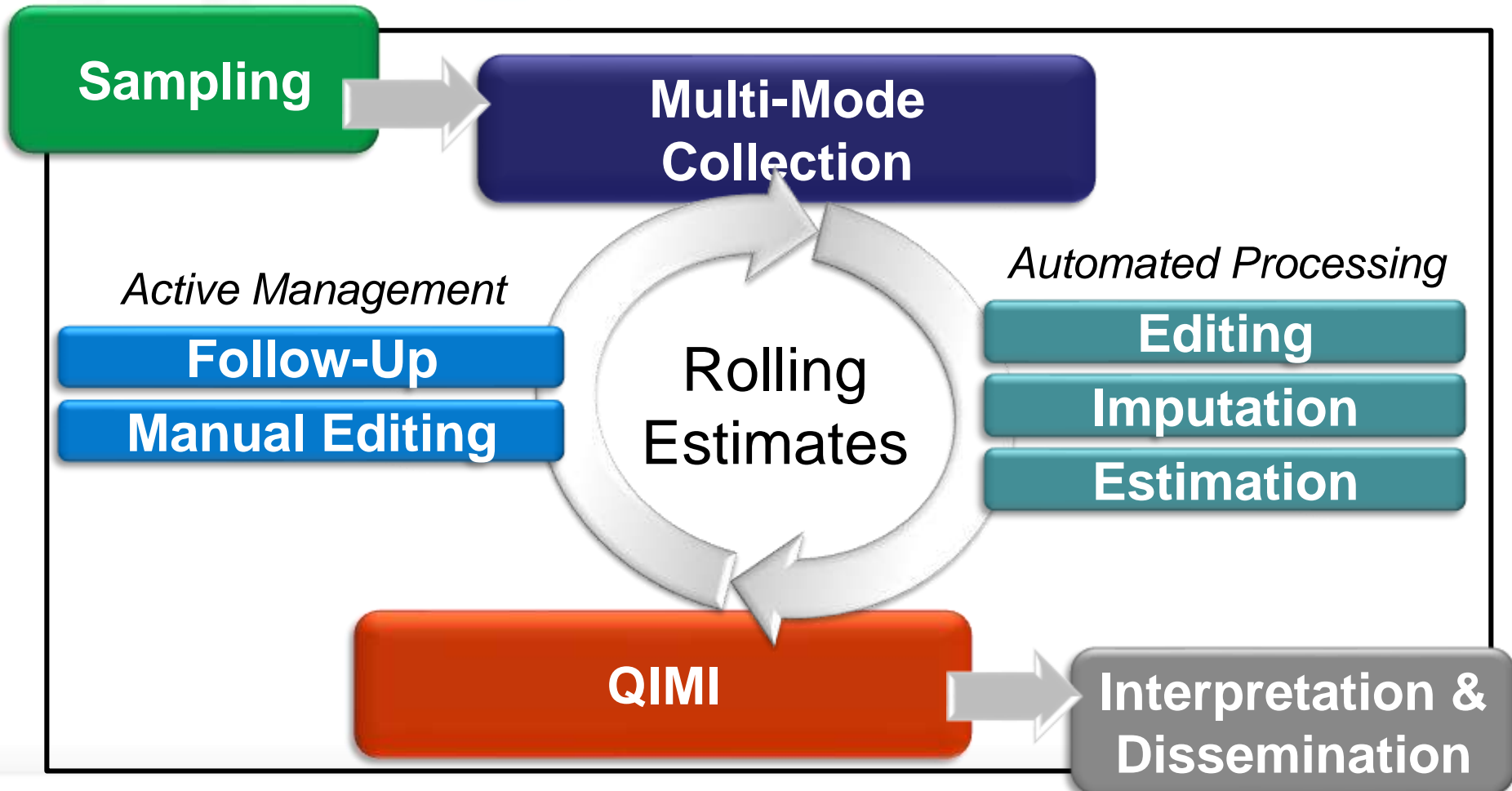


Integrated Business Statistics Program

- A recent platform intended to reduce costs and improve efficiency in processing business surveys at Statistics Canada
- Currently ~95 surveys are integrating into or produced using the IBSP
- QIMI (Quality Indicator/Measure of Impact) is a central part of the program and will be used for ~60 surveys in 2018 where the sample is large enough to warrant collection prioritization



IBSP Active Management Model





QIMI: Ideas

- Focus on a limited number of key estimates and appropriate quality measures
- Evaluate the quality of these estimates at regular intervals throughout collection with quality indicators
- Identify estimates where quality is low
- Identify units with significant impact on the quality of key estimates which have not reached quality targets
- Feed this information to collection and to analysts



QIMI: Implementation

- Every survey selects key estimates, a combination of key variables and key domains
- Each survey has sufficient key variables to capture the complexity of the survey, but not so many as to force all units to be high priority
- Key variables have mandatory collection edits so that units which require failed edit follow-up can be prioritized

QIMI: Implementation

- Quality can be measured in multiple ways and more than one quality indicator can be used at the same time
- For collection year 2017, two quality indicators were used
 - Key Variable Weighted Response Rate (Non-response follow-up)
 - Each variable's response rate is weighted by itself
 - Relative Deviation From Predicted Values (Failed edit follow-up)
 - Uses a predicted value that is compared to the reported/imputed value divided by the estimate total



QIMI: Implementation

- Before QIMI is run, factors of importance are set according to the relative importance of each key estimate
- Quality targets are set for each quality indicator for each key estimate
- Key estimates which are the most important to the survey will be given more stringent quality targets
- Once a quality target is met during collection the key estimate will no longer have impact on collection priorities
- This allows the QIMI prioritization lists to be very dynamic in prioritizing which domains to follow up and which variables in the survey



QIMI: Implementation

- Every survey unit has its measure of impact calculated for every quality indicator for each key estimate where the quality target has not been achieved
- The impact of a unit on key estimates is weighted with the factor of importance and distance from the respective quality targets for each key estimate to obtain a global measure of impact for the unit for each quality target



QIMI: Implementation

- The follow up list is generated using a combination of the ordered lists of global measure of impact for the quality indicators in use
- The priority list is broken down into 6 groups , with one group reserved for units no longer eligible for collection
- Collection workloads are built so that non response follow up concentrates on the top 30% of the eligible units. Failed edit follow up concentrates on the top 15% of the eligible units.

QIMI: In practice

- An agriculture survey had one key estimate with much lower quality than any other estimate
 - It contains the units with the largest difference from predicted and Key variable weighted scores
 - One unit reported in the wrong units causing a huge difference from predicted values
 - One unit has too large an imputation due to this respondent
 - Outlier removal of the respondents misreport contributes to the low quality score
 - Collecting either unit will fix the problem with the imputed unit if no other intervention is made
 - If not fixed through collection analysts are alerted to the problem both with the imputation and the respondent unit



QIMI: Impact

- In collection year 2015 3 surveys piloted QIMI, in 2016 ~36 surveys used QIMI, In 2017 ~50 surveys with more joining each year
- Collection quality is now assessed over many key estimates rather than as a single weighted response rate, thus overall data quality has improved
- Improvement in the percentage of quality targets met both in time and in effort over the three year period , even though many quality targets have been increased since the first year



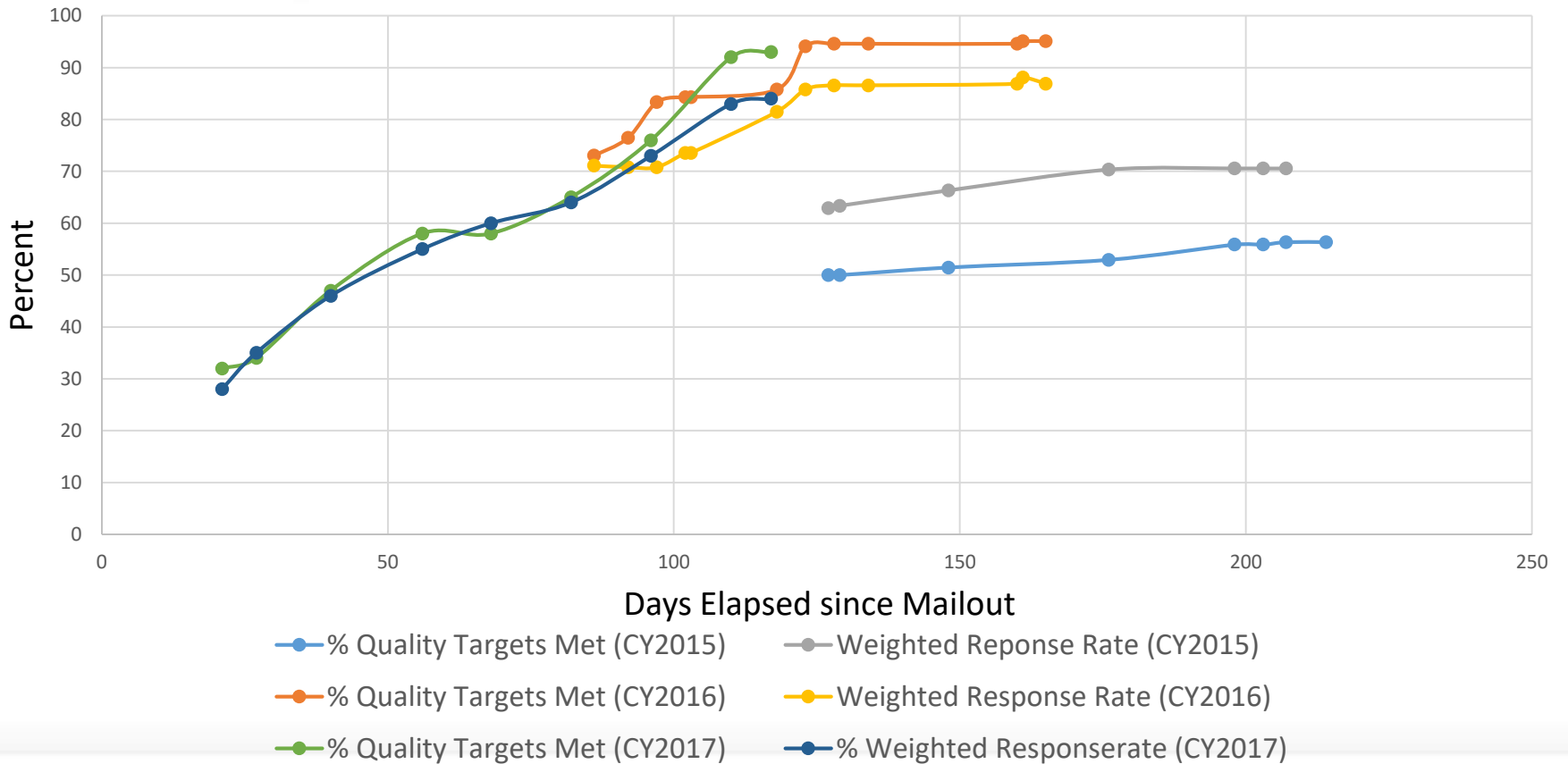
QIMI: Impact

- Many of the original group of surveys were able to shorten their collection period in 2017 without impact to the response rates
 - This is allowing more of our surveys to publish data within 12 months of the end of the reference year
 - A third of the first wave surveys published within 12 months for RY2016, as opposed to a single survey for RY2014
- In the future meeting QIMI collection targets will signal that collection will end early and bring cost reductions to surveys



QIMI: Impact

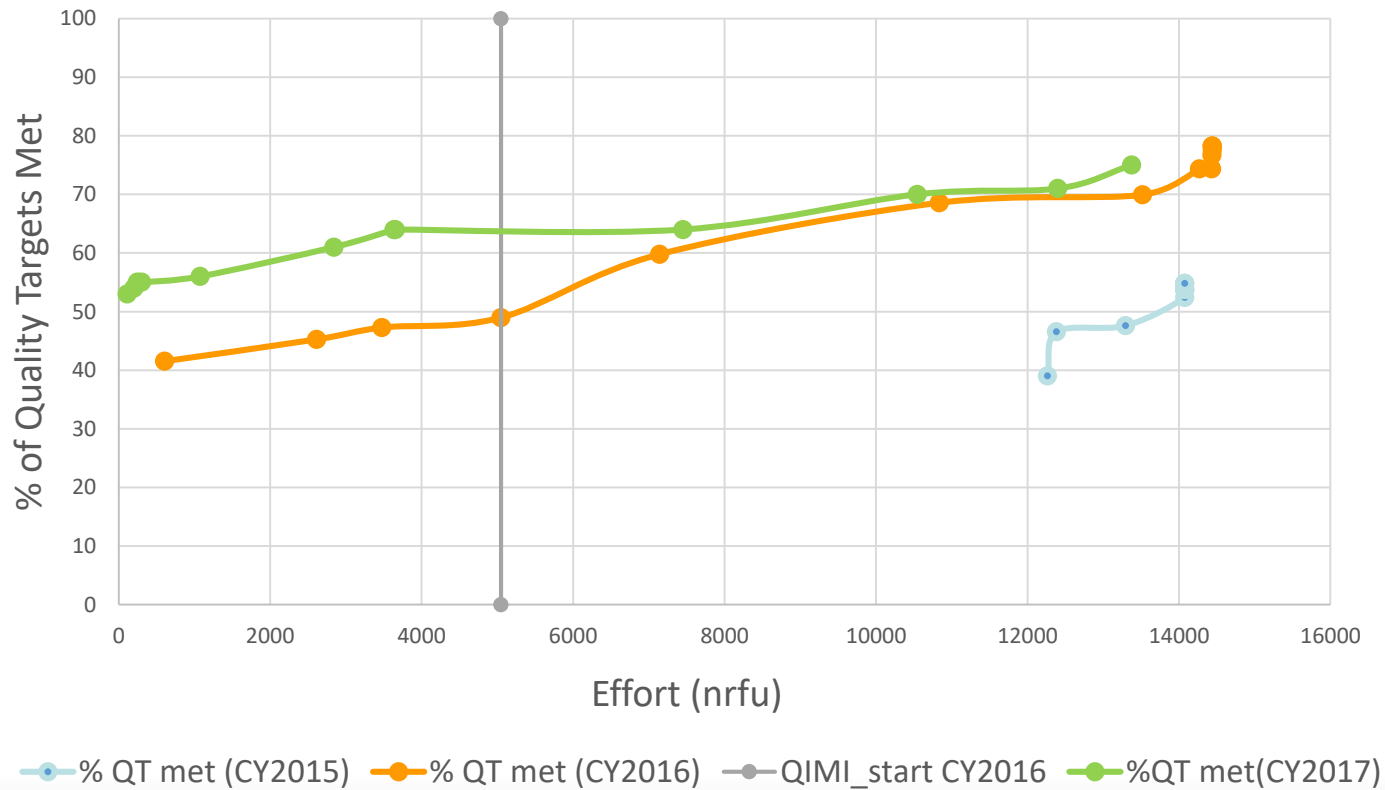
Surveying Mapping : Quality Targets Met and Reponse Rates vs. Days Elapsed since Mailout





QIMI: Impact

Specialized Design Survey: % of Quality Targets Met vs. Effort (nrfu)



QIMI: Future Work

- New quality indicators will be added over the next year
 - Weighted and Unweighted: For surveys where data is required for modelling purposes or there is interest in small units
 - CV : directly linked to whether estimate is publishable
 - In 2018 the IBSP system will introduce the calculation the variance due to imputation.
 - QIMI will then be able to prioritise units based on the impact of switching a unit from imputed to reported on the CV of key estimates
 - Priority of units will be lowered that are easy to impute while those which are difficult to impute will increase their priority
- These indicators can be combined with those already in use depending on survey needs and aims



QIMI: Future work

- The use of paradata can further improve QIMI
- Two sources- Transaction history files and Eqlogs
 - Transaction history files
 - Contain information on all messages and phone calls made to respondents and the result of these calls
 - EQ logs
 - Contain information on respondents behaviour in the EQ in terms of movement between pages and time per page
 - Do not contain information on computer use outside EQ and do not contain respondents keystrokes.



QIMI: Future work

- Priority of units stuck in the survey or who have needed multiple calls to respond in the past can be increased
- The priority of units which are likely to have completed the bulk of the survey can be decreased
 - If respondents have spent a reasonable amount of time on important pages of the survey we can assess the probability that they have supplied the necessary information.



QIMI: Conclusion

- QIMI is already helping surveys to obtain better data for the complete survey, improve quality and publish earlier
- As changes continue to be made to the QIMI system we expect further improvements both in data quality and in reductions in effort and cost.



**For more
information please
contact**

**Pour plus
d'information
veuillez contacter**

Jessica.Andrews@Canada.ca

Matei.Mireuta@Canada.ca

Loic.Muhirwa@Canada.ca

Pierre.Daoust@Canada.ca