Exploring Web Survey Paradata to Improve Survey Design

At Statistics Canada, electronic questionnaires (web questionnaires or e-questionnaires -EQ) have long been recognized as very important reporting options for survey collection. Various electronic data reporting tools have been explored over the last 20 years. Statistics Canada (STC) has continued its attempts to develop an effective and efficient e-questionnaire solution that will meet respondent expectations, comply with Statistics Canada requirements for confidentiality, security and data quality standards, and comply with Government of Canada requirements for accessibility and common look and feel.

The key drivers for moving to web-based questionnaires were:

- Requests from respondents to provide an electronic means for reporting their data;
- Provide a secure, convenient and simple means to respond to our surveys;
- Reduce respondent burden and help counteract anticipated declining response rates;
- Reduce costs by reducing mail, data capture and follow-up collection costs.

Establishment surveys began using the first iterations of web questionnaires in March of 2011. The number of business surveys offering web self-response as part of a multi-mode collection strategy has increased considerably since those early days. Today, nearly 75% of ongoing business surveys offer web self-response.

As multi-mode collection including self response became more and more prevalent for establishment surveys at Statistics Canada, significant effort has been put into improving the respondent's experience with the questionnaire. For example, it is a policy at Statistics Canada to conduct qualitative testing on new or revised questionnaires (including web questionnaires) before proceeding to collection. Effort was also made to improve collection strategies aimed at increasing self-response. Examples of strategies included using Secure Access Code letters for recruitment, and promoting the ease of self-response during CATI or NRFU activities.

This presentation will examine paradata generated by Microsoft Internet Information Services (IIS) logs for the electronic questionnaire to mine for lessons that can be learned by survey design teams. These logs record HTTP transactions. For Statistics Canada's web questionnaires, this means that the logs show actions taken in the web questionnaire by both the respondents and the interviewers for a particular survey. The logs do not contain the questionnaire response data, nor do they contain information regarding web pages accessed by the respondent outside of the web questionnaire. The raw logs contain massive amount of information, but Statistics Canada has done considerable work to clean them up and parse out the most important information. They can now be used by survey analysts to understand a range of web paradata about a survey. Among the many types of variables included in the log files, notable information includes:

- Time per page;
- Type of browser and device used;
- Help button usage;
- Questionnaire path taken;
- Language (English or French) used;
- Time and day;

-    Edits triggered.

Work on using the logs has evolved over time at Statistics Canada. First use dates to 2015 on business surveys. Logs were then used to examine the results of the 2016 Census. Since 2017, versions of the logs that are more user friendly have been made available. As the use of web questionnaire paradata grows, Statistics Canada can use this information to improve the respondent experience by modifying the design of questionnaires using information learned in the paradata.

This presentation will focus on web paradata from Statistics Canada's "mission-critical" monthly business surveys:

-    Monthly Retail Trade Survey (MRTS)
-    Monthly Manufacturing Survey (MSM)
-    Business Payrolls Survey (BPS)
-    Monthly Wholesale Trade Surve (MWTS)

Together, these surveys represent approximately 28,500 sampled units per month, and roughly 60% of those (17,000) self respond via the internet.

The **Monthly Survey of Manufacturing** collects information on sales of manufactured goods, inventories (including raw materials, goods in production and finished products), unfilled orders, and production capacity. This information is used by various stakeholders in the public and private sectors to monitor the economic health of this important segment of the Canadian economy. The objective is to measure month-to-month changes in the main production variables and use this information to show the trends at the industry and provincial level. The monthly sample is approximately 5,500, and the average monthly response rate is approximately 95%. The survey has been offering web self-response since October 2017.

The **Monthly Retail Trade Survey** collects sales, e-commerce sales, and the number of retail locations by province, territory, and selected Census Metropolitan Areas (CMA) from a sample of retailers. Retail sales estimates are a key monthly indicator of consumer purchasing patterns in Canada. Furthermore, retail sales are an important component of the Gross Domestic Product, which measures Canada's production, and are part of many economic models used by public and private agencies. The Bank of Canada relies partly on monthly retail sales estimates when making decisions that influence interest rates. Businesses use retail sales estimates to track their own performance against industry averages and to prepare investment strategies. The monthly sample is approximately 5,800 and the average monthly response rate is 95%. The survey has been offering web self-response since August, 2016.

The **Business Payrolls Survey** is a key input to the Survey of Employment, Payrolls and Hours (SEPH), which provides a monthly portrait of the amount of earnings, as well as the number of jobs (i.e., occupied positions) and hours worked by detailed industry at the national, provincial and territorial levels. SEPH data provide the principal input to labour income estimates: they also serve as a proxy output measure for about 15% of real gross domestic product and "nominal" gross domestic product. SEPH data are also used by the Canada Revenue Agency (CRA), to revise the maximum pensionable earnings and retirement savings plan contribution limits, and by the private sector, for contract escalations and wage rate determinations. The monthly sample is approximately 15,000, and the

average monthly response rate is 90%. The survey has been offering internet self response since late 2012.

The **Monthly Wholesale Trade Survey** provides information on the performance of the wholesale trade sector and is an important indicator of the health of the Canadian economy. In addition, the business community uses the data to analyse market performance. This survey presents estimates of monthly sales and inventory levels for wholesale merchants in Canada, each province and territory. A variety of organizations, sector associations, and levels of government make use of the information. Governments are able to understand the role of wholesalers in the economy (5-6% of the Gross Domestic Product, depending on the year), which aid in the development of policies and tax incentives. The monthly sample is approximately 3,600 and the average monthly response rate is 95%. The survey has been offering web self response since August of 2016.