Proceedings of Statistics Canada Symposium 2014
Beyond traditional survey taking: adapting to a changing world

Testing Collection Strategies for Online Self-Reporting Surveys

Margaret Wu, François Brisebois, and Lecily Hunter[1]

Abstract

In January and February 2014, Statistics Canada conducted a test aiming at measuring the effectiveness of different collection strategies using an online self-reporting survey. Sampled units were contacted using mailed introductory letters and asked to complete the online survey without any interviewer contact. The objectives of this test were to measure the take-up rates for completing an online survey, and to profile the respondents/non-respondents. Different samples and letters were tested to determine the relative effectiveness of the different approaches. The results of this project will be used to inform various social surveys that are preparing to include an internet response option in their surveys. The paper will present the general methodology of the test as well as results observed from collection and the analysis of profiles.

Key Words: Electronic questionnaire; Mail survey; Response rates; Response Profile.

1. Introduction

The use of online self-reporting surveys has become increasingly popular in recent years at Statistics Canada and other national statistical agencies around the world. In particular, Statistics Canada has begun to adopt the online collection mode in household surveys as a complement to other modes of collection, through the use of an electronic questionnaire (EQ).

In Canada, the Census of Population first offered an online response option in 2006 (Tallon and Ramsay, 2010) and continued to do so in 2011 (Hamel, 2012). For other household surveys, incorporating an internet response option is ongoing, and it is expected that internet collection will eventually become the main mode of collection. The Canadian Labour Force Survey conducted a pilot test using the EQ in 2013 (Gambino, Laflamme and Wu, 2014). The Canadian General Social Survey on Social Identity conducted a pilot test using the EQ in 2012 and began to offer an EQ option in regular collection starting in 2013 (Gambino, Laflamme and Wu, 2014).

The benefits to offering an online response mode are expected to be numerous, including potentially large cost savings relative to other modes of collection. However, the use of the EQ in household surveys at Statistics Canada is still in its infancy and questions remain on the best way to make use of this mode. A study was conducted in 2014 testing the effectiveness of two collection strategies for EQ, where sampled units were contacted only by mailed invitations. In comparison, the Labour Force Survey and General Social Survey made first contact by phone before offering an EQ option.

This paper will present the details and results of this study. Section 2 explains the objectives of the test, the collection strategies tested, and the methodology. Section 3 presents the take-up rates from these strategies. Section 4 presents the results of a respondent profile analysis and section 5 provides a summary of the results.

[1]Margaret Wu, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa ON, Canada, K1A 0T6, margaret.wu@statcan.gc.ca; François Brisebois, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa ON, Canada, K1A 0T6, francois.brisebois@statcan.gc.ca; Lecily Hunter, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa ON, Canada, K1A 0T6, lecily.hunter@statcan.gc.ca

2. Overview of the Study

2.1 Two Collection Strategies for EQ

In both strategies, paper invitations for an online survey were mailed to sampled units and this was the only method of contact. The objectives of the study were to measure take-up rates under these strategies and conduct a profile analysis of respondents. The two strategies were:

1. Household strategy: Letters were mailed to a sample of dwellings with an invitation to do the survey via EQ. The invitation was addressed to "The Occupant". Once online, the household respondent was asked to complete a household roster, at which point one person aged 15 or older in the household was randomly selected to complete the questionnaire. If the selected respondent was the same person who completed the roster, the EQ application continued to the content; otherwise, the EQ application ended and a new invitation to the named selected respondent (called the secondary respondent) was sent. If the household respondent provided an email address, the new invitation was emailed; otherwise it was mailed and the letter was addressed to the named selected respondent.

2. Target respondent strategy: Letters were mailed to a sample of selected people with an invitation to the EQ survey. Each invitation letter was addressed to the name of the selected ("targeted") person. With this strategy, there was no need for a roster and only the selected person could answer the EQ. Two different invitation letters were tested under this strategy; letter P1 was a standard letter, the same as the one used for the household strategy; letter P2 had one extra sentence stating that the respondent may be contacted by an interviewer if they do not respond.[2]

The main disadvantage of the household strategy was that the flow was more complicated. This strategy required a roster and sending invitations to a secondary respondent selected from that roster. The consequence was that there was a second level of non-response since there were potentially two respondents per household. The target respondent strategy had the advantage of a simpler flow but on the other hand required a person-level frame with names and up-to-date addresses, which is generally more difficult to obtain at Statistics Canada.

2.2 Frame and Sample

The target population for both strategies was composed of all households that completed Statistics Canada's 2011 National Household Survey (NHS) and that had mailable addresses available on the Dwelling Universe File (DUF) of Statistics Canada Household Survey Frame (HSF). The frame was therefore a combination of the NHS and the DUF, and was identical for both strategies tested. Canadian territories, Indian reserves, and collective dwellings were excluded. Note that mailing addresses were from the DUF, and the names used to address invitation letters for the target respondent strategy were from the NHS. Note also that the NHS took place two and a half years before this study, meaning that information on the respondent names could have been outdated.

The sample size for the household strategy was 30,000 households. The sample for the targeted respondent strategy was 6,500, which was split into two samples of 5,000 and 1,500, where the smaller sample received the non-standard invitation letter mentioning a possible follow-up. A simple random sample of households was taken within each region of Canada; for the target respondent sample, one individual aged 15+ was randomly selected from each selected household using the NHS roster.

2.3 Collection

Collection lasted approximately six weeks and took place in January and February 2014. Given the limited resources and time to conduct the test, an existing EQ module was reused: one module of the General Social Survey (GSS)

[2] There was actually no plan to follow-up with non-respondents. This sentence was added simply to test the impact on take-up rates under a "threat" of a possible follow-up.

related to social identity and internet use. The official survey title for the test was the Survey on the Use of Internet and Other Media (SUIOM).

3. Take-up Rates

The main objective of the test was to measure take-up rates from potential EQ strategies for household surveys.

Take-up rates for the household sample are presented in Tables 3-1, 3-2 and 3-3. Table 3-1 presents the counts and rates of post office returns, logins, and submits (i.e. cases that completed the survey content). Overall, after excluding post office returns, the login rate was 30% and the submit rate was 22% out of the 30,000 sampled. Table 3-2 breaks down the number of submitted cases by whether the same person was selected and therefore continued to the content immediately after the roster, or whether a different household member was selected and a new invitation was issued: cases where the same person was selected accounted for 75% of all content submits, and cases where someone else was selected account for the remaining 25%. Table 3-3 provides further information about cases where a different person was selected: there were 3465 cases where a different household member was selected from the roster; 2830 of these cases were sent invitations for the survey content. Note that not all cases were sent invitations if it was considered too late in the collection period; note also that there was a later cut-off date for email invitations than letter invitations because there was no need to account for mailing time. Out of the invitations sent, the overall login and submit rates were 58% and 57%. These rates were quite different depending on the method of contact: rates were about 20% higher for those contacted by email versus those contacted by mailed letter (64% login and 63% submit for email; 43% login and 42% submit for mail). Note however that it is the initial household respondent's choice whether to provide an email address for the selected household member which then determines the mode for the second invitation.

Take-up rates for the target respondent sample are presented in Table 3-4. The post office return rate was 12% overall. Excluding the post office returns, the login and submit rates were 31% and 30% overall. However these rates varied depending on the version of the invitation letter: login and submit rates were 30% and 29% for the sample sent the standard letter, and 34% and 34% for the sample sent the letter mentioning a possible follow-up.

Table 3-5 presents comparisons between the two samples and includes results from statistical tests of differences. The initial login rates for the household and target respondent samples are comparable at 30%. However, the submit rate for completed cases is significantly different (22% for household, 29% for target respondent with the standard letter), most likely from having to re-contact the household sample when someone else was selected. Comparing the two invitation letters in the target respondent strategy, the login rates are significantly different (30% standard letter, 34% follow-up letter), which indicates that the "threat" of a follow-up does have an effect on response. And, although not formally tested, the post office return rate is much higher for the target respondent sample (12% target respondent, 2% household). This is likely due to the contact information being out of date; recall that respondent names were taken from the NHS which took place two and a half years before the study, and it is almost certain that some selected persons would had moved during that time.

Table 3-1
Household Sample: Overall Results

	Count	Rate
Sample	30,000	
Post office returns	630	2%
Logins	8,673	30%
Submits (completed content)	6,324	22%

Note: login and submit rates exclude post office returns.

Table 3-2
Household Sample: Breakdown of Submits

Total submitted cases (completed content)	6,324	
└ Same person selected	4,720	75%
└ Different person selected	1,604	25%

Table 3-3
Household Sample: Results for the Secondary Respondent

	Roster submitted	Invitations sent	Login		Submit	
Total	3,465	2,830	1,632	58%	1,604	57%
Contacted by email	n/a	1,980	1,267	64%	1,244	63%
Contacted by letter	n/a	850	365	43%	360	42%

Notes: Roster submitted count excludes a few (12) dropped cases with errors in contact information. Count of invitations sent excludes 59 invitations by email where the invitation was bounced back.

Table 3-4
Target Respondent Sample: Results

	Letter P1 (standard)		Letter P2 (follow-up)		Total	
Sample	5,000		1,500		6,500	
Post office returns	597	12%	167	11%	764	12%
Logins	1,321	30%	454	34%	1,775	31%
Submits	1,295	29%	447	34%	1,742	30%

Note: login and submit rates exclude post office returns.

Table 3-5
Comparing the Strategies

Comparison	Group 1		Group 2		p-value
Login rate	Household	30%	Target respondent, standard letter	30%	0.60
Overall submit rate	Household	22%	Target respondent, standard letter	29%	<0.001
Login rate for letters (target respondent strategy)	Standard letter	30%	Letter mentioning follow-up	34%	0.02
Post office returns	Household	2%	Target respondent	12%	Not tested

4. Respondent Profiles

To determine which types of households and individuals would be more likely to respond to these strategies, logistic regression models were applied to find the characteristics significantly related to login. Characteristics were from the frame which included (proxy) demographics from the National Household Survey and dwelling and household type characteristics from the Household Survey Frame. The odds ratios of login for significant characteristics are presented in Figures 4-1 and 4-2.

For both the household and target respondent sample, the two most significant characteristics in explaining login were the response channel to the NHS (via internet or other method), and the household maximum education level. The odds of a household or individual logging in were 2 to 2.5 higher if the household responded to the NHS via internet than if the household responded via some other method. The odds of login increased as the education level

increases, more so for the household sample than for the target respondent sample. Overall, there were dwelling, household, and person-level characteristics that are significant in explaining login; some are significant in explaining login regardless of whether it is the situation of anyone in the household being allowed to login, or the situation where only a particular person selected could login.

Figure 4-1
Odds Ratios of Login for Household Sample

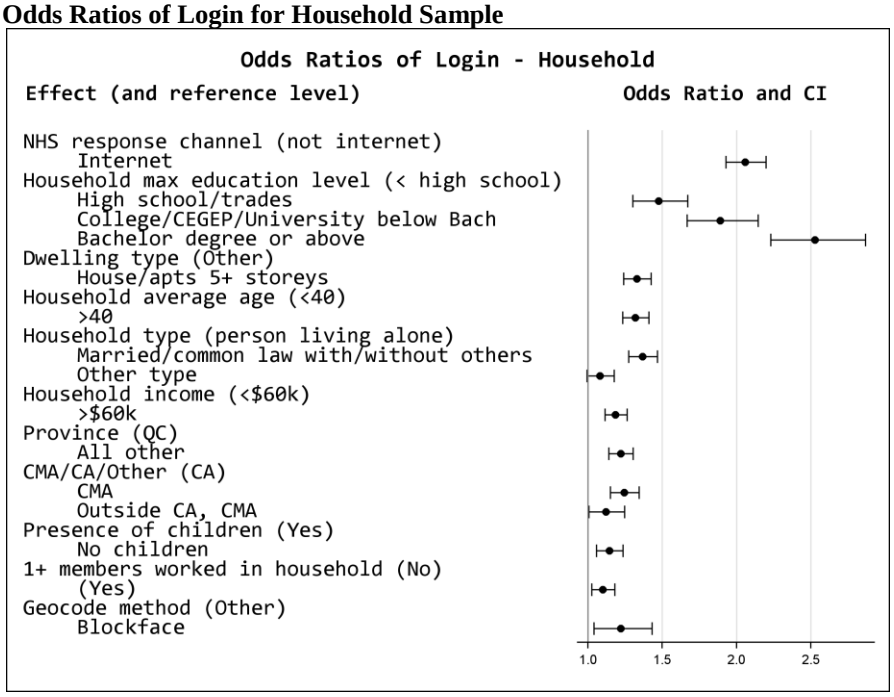
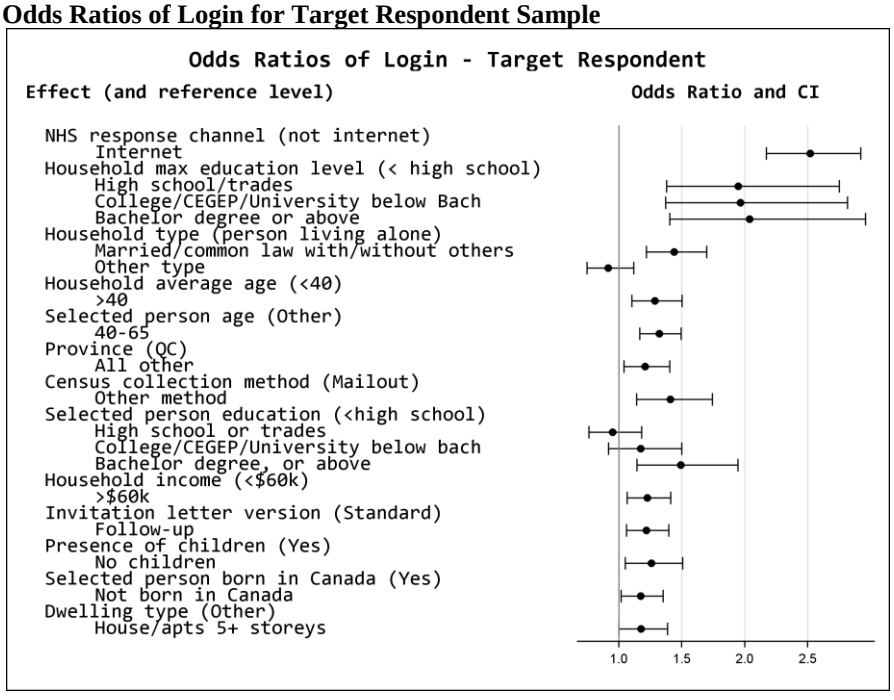


Figure 4-2
Odds Ratios of Login for Target Respondent Sample



5. Summary

This study explored the idea of contacting sampled units for EQ household surveys by mailed invitations. Two collection strategies were tested and compared. The initial log-in rates for the strategies were comparable at around 30% but the overall submit rate was significantly higher for the target respondent strategy, due to the loss in having to re-contact in the household strategy. However, it is not clear that one strategy was "better" than the other, as there were more post office returns for the target respondent sample, which indicated that the respondent names on the frame used for contact were somewhat out of date.

Additionally, certain segments of the population were more likely to respond to these strategies in this study. In particular, those who had responded to the NHS via the internet (compared to some other method) and those with higher education levels were more likely to respond.

References

Gambino, J., Laflamme, G., and M. Wu (2014), "Electronic questionnaires (EQ) for household surveys at Statistics Canada", unpublished report, Ottawa, Canada: Statistics Canada.

Hamel, M. (2012), "Internet data collection in the Canadian census of population", Economic Commission for Europe, Fourteenth Meeting on Internet Data Collection, May 2012. Available online at www.unece.org.

Karaganis, M., and G. Laflamme (2012), "Use of E-questionnaires in Household Surveys", unpublished report, Ottawa, Canada: Statistics Canada.

Tallon, P., and L. Ramsay (2010), "An Integrated approach to collection of Census data in Canada", *Proceedings of Statistics Canada Symposium 2010*, pp. 275-281.