Internet Response in the 2016 Canadian Census of Population and Analysis of Respondent Behaviour Using Paradata

Marie-Pier Lemieux[1]
[1]Social Survey Methods Division, Statistics Canada
100 Tunney's Pasture Driveway, Ottawa, ON, K1A 0T6, Canada

Abstract
Every five years, Statistics Canada conducts a mandatory census to provide population and dwelling counts for Canada as well as a wide range of information about Canada's demographic, social and economic characteristics. The methodology has considerably evolved over the years in order to reduce respondent burden, improve the efficiency of the Census Program and data quality. Every day during the 2016 Census, Statistics Canada has been able to collect paradata in order to analyze the respondents' behaviour. This paper provides results of ongoing investigations on paradata analysis to better understand the respondents' behaviour in preparation for the 2021 Canadian Census.

Key Words: Paradata, Canadian Census of Population, Online Questionnaire

1. Introduction

Every five years, Statistics Canada conducts a mandatory census to provide population and dwelling counts for Canada as well as a wide range of information about Canada's demographic, social and economic characteristics. The methodology has considerably evolved over the years in order to reduce respondent burden, improve the efficiency of the Census Program and ensure the quality of data. Building upon insight drawn from previous research and censuses, the 2016 Canadian Census of Population achieved one of its best collection rates with 98.4%, and moreover, the best internet response rate for a census since the Canadian Census's inception with 68.3%.

Statistics Canada is at the forefront of the use of internet-based data collection tools. A new corporate approach has been developed during recent years in order to take advantage of new technologies by leveraging the benefits of the online questionnaire. It has been a great opportunity to re-think the use of paradata and to ensure that all information needed to monitor collection is obtained. Daily during the Census, Statistics Canada has been able to analyze the respondents' behaviour by getting a variety of paradata such as the last page viewed on the online questionnaire by households, all the activities done by respondents such as logging in the questionnaire, saving the questionnaire and also which mobile device was used by the respondents. In addition, Statistics Canada derived more variables from paradata at the end of collection that have been useful to analyze the efficiency of validation

[1] Author to whom correspondence may be addressed: marie-pier.lemieux@canada.ca

messages and help pages, calculate the average completion time, and analyze the pattern of respondents' behaviour.

Section 2 of this paper begins by giving an overview of the 2016 Canadian Census of Population with an emphasis on the online questionnaire. Section 3 will present the results obtained so far from the analysis of paradata received from the census according to these sub-topics: devices and browsers used to complete the questionnaire, time to complete the census, characteristics of households who used a smartphone versus those who used a desktop and finally the behaviour of respondents in the online questionnaire. Section 4 will conclude this paper by resuming what has been done with paradata up to now.

2. Overview of the 2016 Canadian Census of Population

Canada's most recent census was held in May 2016. Approximately 15.5 million private dwellings and a population of more than 35 million people were enumerated. Two types of questionnaires were mainly used to collect data. The first type is the short-form (2A) that contains 10 questions and is distributed to 75% of dwellings. Questions concern the size and demographics of the household, the marital status and the language spoken at home. The second type is the long-form (2A-L) and it contains the same first 10 questions as the short-form, followed by 50 additional questions about various topics such as immigration, ethnic origins, mobility, education, labour and housing. This detailed form was distributed to the remaining 25% of dwellings.

For the 2016 Census, respondents had multiple options to fill out their census questionnaire: they could complete it by using the online questionnaire, they could complete it by using a paper questionnaire and then sending it back to Statistics Canada (via Canada Post) or they could call the census help line and have it completed via computer-assisted telephone interview. Dwellings that didn't fill out their form at the beginning of the non-response follow-up period were contacted by an interviewer to complete it, either by telephone or in person.

2.1 History of the online questionnaire

The use of the online questionnaire during the census is not new for Statistics Canada. It began in 2001 when a feasibility study was conducted in two sites chosen for their high internet connectivity rates back then. For the first time at Statistics Canada, a social survey offered the online questionnaire on the internet as a response mode (Léger 2015). Because of the success of this first test, the 2006 Census introduced the online questionnaire as a mode of response to all Canadians. Dwellings still received a paper questionnaire but the uniform resource locator (URL) of the census website as well as a unique online access code to allow them to login on the website and access the online questionnaire were printed on it. The online response rate was 18.3%.

For the 2011 Census, a new collection methodology, called the Wave Methodology, was tested and introduced. It is mainly based on the model proposed by Dillman (2007) which consists of a series of contacts with households to prompt self-response. In particular, it is reminding Canadians to fill out their questionnaire at specific times throughout the collection period. It is also encouraging respondents to use the online questionnaire to complete their census form, while mitigating the risk of a decline in overall response by offering other response modes (Mathieu 2017). A focus was made on mail-out areas, which

is where the vast majority of dwellings have addresses that are of the appropriate type and of sufficient quality for respondents to be contacted by mail for the Census. For 2011, around 75% of dwellings located in mail-out areas were selected to receive an invitation letter to complete their census online instead of a paper questionnaire. They were identified based on the probability of filling out their census and their form on the internet. The remaining dwellings received a paper questionnaire but still had the option to complete the census using the online questionnaire. The online response rate was 53.8%.

For the most recent census, the same approach was used but was expanded to include all dwellings in mail-out areas following results of a live test done during 2011 Census. Overall, approximately 82% of all dwellings received an invitation letter with the URL of the website and an online access code at the beginning of the collection period instead of a paper questionnaire. The online response rate was 68.3%.

2.2 Why use the online questionnaire?

Several factors have led Statistics Canada to use the online questionnaire for the Census. The main approach for the 2016 Census of Population was to leverage the process and technological advancements that were successfully implemented in 2011 while addressing the challenges that were encountered (Axalan and al. 2017).

Firstly, a user-friendly online questionnaire can improve the cost effectiveness by increasing self-response and reducing the cost of downstream processing activities such as paper imaging/scanning, editing and coding among others. It can also improve the timeliness by decreasing the time between when the respondent submits the questionnaire and when Statistics Canada receives it. As a result, it reduces the time of downstream processing activities, as mentioned above, and also shortens the time between the end of collection and the first results published.

Another interesting feature of the online questionnaire is the reduction of the respondent burden by showing only questions that are appropriate to the situation of the household. The online questionnaire allows us to maintain a high level of data quality by showing warning messages when a question is not answered or the data entry is invalid. Finally, by promoting the use of the online questionnaire, it reduces reliance on paper which has several benefits such as a decrease in the need to print paper questionnaires.

2.3 Collection results from 2016 Census of Population

The 2016 Census of Population has been a huge success by surpassing all planned goals. Overall, a census collection rate of 98.4% was achieved, which is an improvement from 2011 Census. As shown in Table 2, the census collection rate can be distributed according to the different methods of collection. One of the biggest achievements was about the online questionnaire rate of 68.3%, an increase of 15% from 2011 Census. It is also the highest ever achieved for a census in any country to date (Valente 2016). Another success has been the increase of the self-response rate, which is the combination of the online questionnaire and the paper questionnaire rates. The rate grows by 3.6% to get 88.8%. Those two rates combined allowed Statistics Canada to decrease significantly the number of cases to be treated at the beginning of the non-response follow-up period.

Table 1: Collection response rates by collection method and census years

Collection method	2011 Census	2016 Census
Census collection rate	98.1%	98.4%
Online questionnaire	53.8%	68.3%
Paper questionnaire	31.3%	20.5%
Other	13.0%	9.6%

Figure 1 shows online and mail daily return rates as well as the self-response collection rate between May 1st and June 15, 2016. By looking at the online returns only, a first peak occurred on May 3rd, which is the following day where dwellings received their invitation letter to complete their census online. On that day alone, Statistics Canada received 1,125,000 online questionnaires, which is close to 9.5% of the total response. A second peak occurred on May 10, the Census Day. Respondents were asked to complete their census questionnaire by May 10. During those first nine days, Statistics Canada received half of the total response.

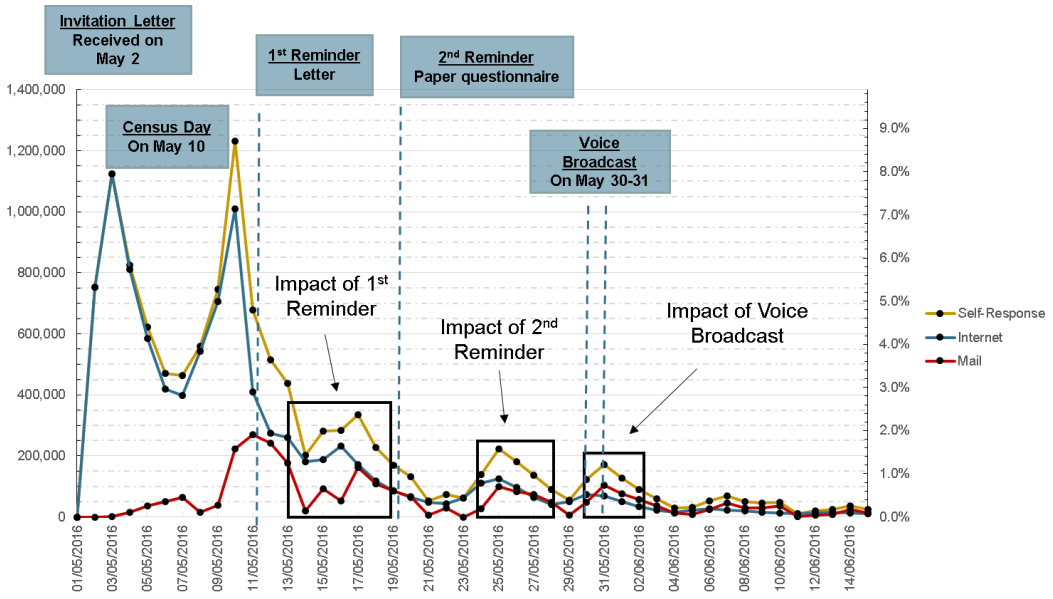

Figure 1: Online, mail and self-response daily rates between May 1st and June 15 during the 2016 Census

3. Paradata Analysis

Paradata have been used for many years at Statistics Canada and its use has been increasing greatly since the introduction of the online questionnaire. In the context of this paper, paradata are information gathered during data collection or statistical data production that are related to an identifiable person, business or organization.

The next subsections will show results that have been obtained so far by analyzing paradata. It will begin with the analysis on the electronic devices and browsers used, followed with the time to complete the online questionnaire, characteristics of households using smartphones and finally respondent behaviour in the online questionnaire.

3.1 Electronic devices and browsers used

For the 2016 Census, an emphasis was made to increase self-response, specifically by promoting the use of the online questionnaire and this enabled Statistics Canada to gather a lot of information about the electronic device and browser used to complete the census.

Based on over 9.5 million online responses, the clear majority of respondents used their desktop or laptop to complete their census. As shown in Figure 2, an additional 13% responded by using a tablet followed by 8% using a smartphone. A small proportion of respondents used non-typical devices like gaming devices (PlayStation, Xbox and the Wii) as well as smart TVs and e-readers. It is important to keep in mind, when testing the online questionnaire, any device that can connect to the internet can be used.

The browser used by respondents could also be identified. Almost half of the respondents used Chrome followed by Safari and Internet Explorer, both at 20%. Firefox was used by 11% of the respondents and a small 2% of all browsers couldn't be identified.

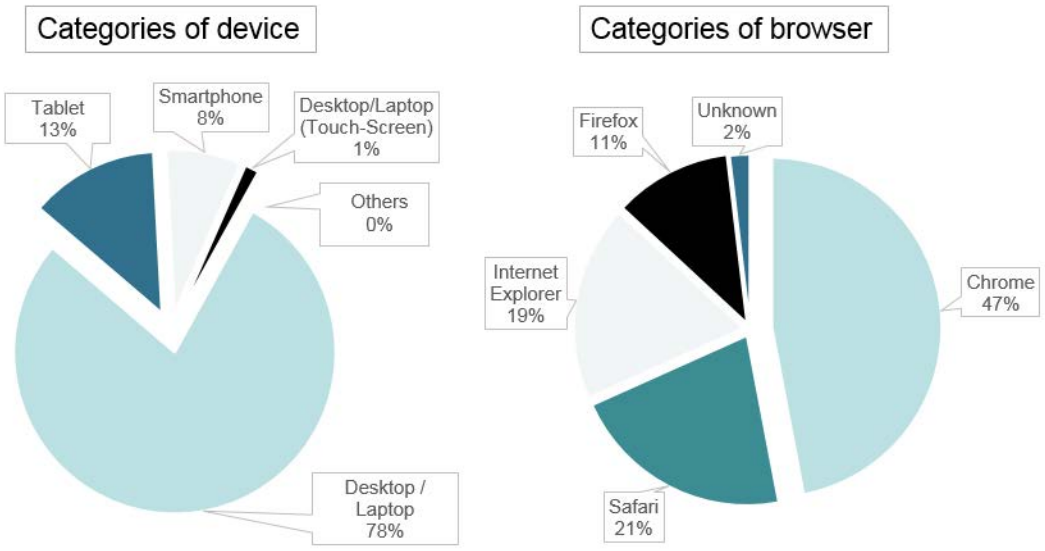

Figure 2: Devices and browsers used during the 2016 Census

3.2 Time to complete

Once the devices used to complete the census were identified, it was important to identify if there were differences in completing the census depending on the electronic device used, and if yes, understand the behaviours. At first, the time to complete the online questionnaire was analyzed by the type of the questionnaire: the short or long form. Table 2 shows the average time to complete the online questionnaire by type of device. It took an average of 8 minutes and 13 seconds for a respondent using a desktop or laptop to answer the short-form and a little over 36 minutes to answer the long-form. Only small differences are identified by comparing those results with other devices.

Table 2: Average time to complete the 2016 Census questionnaire by device category and by type of questionnaire

Form-type	Desktop/Laptop	Smartphone	Tablet
Short-Form	8 min 13 sec	+45 sec	+11 sec
Long-Form	36 min 05 sec	-41 sec	+1 min 0 sec

It is also interesting to look at these data according to the number of people in the household because it is expected that the more people in the household, the more time it would take to complete the questionnaire. As shown in Table 3, the average time to complete the online questionnaire increases with the number of people in the household. Also, respondents who used a tablet took almost the same amount of time to complete their census as those who used a desktop or laptop. Furthermore, there is a noticeable difference for the time to complete the questionnaire with respondents who received the census long-form and responded using a smartphone. The larger the household size, the larger the difference there is between smartphone users and desktop or laptop users.

Table 3: Average time to complete the 2016 Census questionnaire by device category, type of questionnaire and the number of people in the household

Form-type	# People in the household	Desktop/Laptop	Smartphone	Tablet
Short-Form	1	5 min 58 sec	+22 sec	+11 sec
	2	7 min 42 sec	+12 sec	+5 sec
	3+	10 min 03 sec	+39 sec	+1 sec
Long-Form	1	24 min 22 sec	-7 sec	+22 sec
	2	34 min 29 sec	-1 min 41 sec	+31 sec
	3+	45 min 19 sec	-2min 29 sec	+35 sec

The next step was to look at the time it took to answer each question to try to figure out if there is a different behaviour between smartphone users and desktop laptop users. For each page of the online questionnaire, the average time spent for smartphone and desktop/laptop users and the difference between both of them, was calculated. Table 4 shows the pages with the greatest difference where the respondents who used a smartphone took less time to answer the question. Those differences are concentrated in three subjects related to education, to labour and to housing questions. Pages with the highest differences, like pages 52 and 64, are questions where respondents must give a lot more information like the address of the employer, the type of diploma obtained, the name of the school and all expenses related to housing.

Table 4: Average time spent on education, labour and housing questions by device category

Category	Page	Average time for Smartphone (mm:ss)	Average time for Desktop/Laptop (mm:ss)	Difference (mm:ss)
Education	44	1:36	1:50	-0:14
	45	0:44	0:50	-0:06
	46	0:22	0:21	0:01
Labour	52	4:02	4:24	-0:22
	53	0:19	0:17	0:02
	54	2:36	2:42	-0:06
Housing	64	1:39	2:03	-0:24
	65	0:33	0:32	0:01
	66	0:35	0:36	-0:01
	67	0:13	0:14	-0:01
	68	0:31	0:44	-0:13

3.3 Characteristics of households

The distribution of households, by looking at the age of adults in the family and if they have children or not, was analyzed to see if there were any characteristics that could explain the difference between smartphone and desktop or laptop users. As shown on Table 5, younger households are more often using a smartphone to submit the online questionnaire and conversely, older households tend to use a desktop or a laptop. Likewise, having children seems to increase the use of smartphones versus the desktops or laptops for the age group of 30 to 50 years old.

By combining this analysis with the previous analysis that discusses the average time per page, one hypothesis that could be made is that smartphone users are younger so they may have fewer responsibilities and therefore have less things to mention about housing and labour questions. Another hypothesis would be that respondents who choose to use the smartphone are accustomed to using it for various tasks and may have tricks to help them finish the online questionnaire quicker. Moreover, the length of writes-in were analyzed and no significant differences between smartphone users versus other devices were identified. Finally, breakoff rates which is where respondents leave the questionnaire and never come back, were not impacted by the device used to complete the questionnaire (Ciceri and al., 2016).

Table 5: Proportion of smartphone and desktop users by age group and having children or not in the household

	Age of adults in the family	Smartphone (%)	Desktop/Laptop (%)
No children	All between [18-30[	12.88	6.98
	All between [30-50[	15.31	11.03
	All between [50-65[	7.08	15.62
	Different age range	13.79	22.20
With children	All between [18-30[	5.77	1.01
	All between [30-50[	26.82	16.16
	All between [50-65[	0.72	1.31
	Different age range	12.90	10.11
With or without children	All [65+	4.72	15.58
		100.00	100.00

3.4 Respondent's behaviour

Paradata helped Statistics Canada to analyze the behaviour of respondents in the online questionnaire. Each action done by the respondent, such as logging in, logging out, saving, abandoning, etc. had a code associated with it. A record on the paradata file gives a sequence of activities. Thereby, a pattern was identified for each user and a list with the most common pattern was created.

Table 6 shows the proportion of respondents who submitted or abandoned the form. Submitting or abandoning means that they never submitted? the online questionnaire, in one or multiple sessions. For the short-form, respondents mostly completed their questionnaire in one session with fewer than 2% abandoned. In the long-form, there's a higher proportion of respondents who completed their questionnaire in multiple sessions. It can be explained by the ability to save the questionnaire which is not an option for the short-form. Another thing that can also explain the higher percentage of respondents who abandoned is that they have more information to gather to answer the questions. They must collect all their bills to complete housing questions, get the address of their employers, etc. It should be mentioned that abandoning doesn't mean that Statistics Canada didn't receive a questionnaire from respondents. They could have chosen to complete it with a paper questionnaire or have been collected during the non-response follow-up period by an interviewer.

Table 6: Proportion of most common patterns by type of questionnaire and by one or multiple sessions.

	Pattern	Short-Form (%)	Long-Form (%)
One Session	Login – Submit	94.17	73.38
	Login – Abandon	1.11	2.48
Multiple Sessions	Login – Submit	4.15	20.68
	Login – Abandon	0.57	3.46
		100	100

4. Conclusion

The 2016 Canadian Census of Population was a huge success thanks to the collaboration of all those who worked on the project, and also because of the mixed methods of collection and the wave methodology. Statistics Canada has been able to surpass all collection goals, especially by being the world leader for the online collection method. Massively, Canadians completed their census using the online questionnaire and half of the total response was collected during the first nine days of collection.

All the paradata gathered from the online questionnaire allowed Statistics Canada to better analyze respondents by looking at their behaviour. Smartphone users compared to other device users complete the long-form questionnaire quicker and the difference increases with household size. Smartphone users are younger and answer writes-in questions quicker than non-smart phone users. Finally, respondents who received the long-form questionnaire completed it in multiple sessions more often than those who received the short-form.

The uses of paradata discussed in this document are only a small fraction of what can be done with paradata. The focus of this paper was on the behaviour of respondents for the online questionnaire, but paradata can also help to better design questionnaires. Other analysis could be by looking at the help pages that were consulted, how much time it took to create a password, how many days have passed before a respondent came back in the online questionnaires, etc. By encouraging respondents to complete their census using the online questionnaire, more paradata can be collected meaning more analysis can be done and more improvements can be brought to these collection methods.

Acknowledgements

I would like to thank my supervisors Jeannine Claveau and Danielle Léger for giving me the opportunity to work on a great project and also Keven Bosa, Patrice Mathieu and Jean-Pierre Morin for taken the time to review this paper.

References

Axalan, J., Franklin, S. and Robles, C. (2017), 2016 Census of Population Electronic and Paper Questionnaires Evaluation Report, Statistics Canada internal document

Ciceri, C. and Roy, L. (2016), 2016 Data Quality Report on EQ Self-Responding Household, Statistics Canada internal document

Dillman, D.A. (2007), Mail and Internet Surveys: The Tailored Design Method, second edition, Wiley

Léger, D. (2015), Use of the Internet in the Canadian Census, Statistics Canada internal document

Mathieu, P. (2017), Collection of the 2016 Census: Results and Post-mortem, paper presented to the Advisory Committee on Statistical Methods, May 2017

Mathieu, P., Morin, J.-P., Rodrigue, J.-F. and Taylor, J. (2012), Collection Methodology of Canada's 2011 Census and National Household Survey, Joint Statistical Meetings Proceedings, 2012

Valente, P. (2016), Overview of the Use of Technology for Data Collection in the UNECE Region – Practices Adopted in the 2010 Round and Plans for the 2020 Round, Technical Meeting on Use of Technology in Population and Housing Censuses, November 2016