



Future developments on the Canadian Census of Population

Marc Hamel*

Statistics Canada, Ottawa, Canada – marc.hamel@canada.ca

Julie Trépanier

Statistics Canada, Ottawa, Canada – Julie.trepanier@canada.ca

The Census of Population Program in Canada has seen the introduction of several innovative approaches over time. The most recent censuses were no exception with, for example, the introduction of internet data collection (2006) and a wave collection methodology (2011). As a result more than 68% of households completed their 2016 census form online, making Canada a world leader in internet collection. Overall, the collection response rate was 98.4%, most of it in self-response (88.8%), the best result ever on the Canadian Census. But Canada, like many countries around the world, faces a number of challenges in preparing for its next census in 2021. Statistics Canada will need to continue to innovate as it faces fiscal pressures and as concerns around response burden continue to rise. This paper briefly describes options that have been considered for the methodology of the 2021 Census and beyond. These include an increased use of information from administrative data sources to support or even replace the collection of some census information.

Keywords: first keyword; second keyword; third keyword; fourth keyword.

1. The current Canadian Census model

Canada conducts a census of population and dwellings every 5 years, using a traditional methodology. The current Canadian Census collection approaches are briefly described in Rodrigue and al. (2012). The most recent census took place in May 2016 and expanded on methodological changes introduced for the 2011 Census (Hamel and al., 2013). The main approach is a wave collection methodology which encourages response via internet while ensuring a high level of self-response. The basic approach, based on direct mail-out of invitation letters to respond online, was extended to a larger portion of dwellings (82%) in 2016. This led to high efficiency and quality as illustrated by record levels of self-response (88.8%) and Internet response (68.4%).

Two questionnaires are used, a short form collecting basic demographic information sent to 75% of households in 2016, and a long-form collecting more detailed socio-economic information for 25% of households. New for 2016 was the use Income Tax and Benefits files from the Canada Revenue Agency to replace the collection of income information traditionally collected on the long-form. Also new to 2016 was the addition of the immigrant admission category (McLeish, 2016). This information was obtained from administrative sources from the department of Immigration, Refugees and Citizenship Canada (IRCC), and integrated into the regular processing stream to benefit from the edit and imputation approach used for all census variables. This information will be released with the rest of the census outputs.

Since 2006, administrative information is increasingly used in the creation and maintenance of the Address Register (AR) serving as the dwelling frame for the 2016 Census. The last general block canvass activity was conducted for the 2006 Census, and a mix of various administrative sources and targeted listing activities has been used since to keep the register up-to-date (Dolson, 2011). The decision on where to conduct the targeted listing is based on an algorithm identifying areas with more 'noise' in the administrative sources, indicating an increased risk of under or over coverage of dwellings.

2. The drivers for change

The events around the 2011 Census and the decision from the government to conduct the long-form portion of the census as a voluntary survey, the 2011 National Household Survey (Hamel and al., 2013), prompted



Statistics Canada to conduct a new body of research into possible methodology options for future cycles. The 2016 Census Strategy Project was conducted from 2010 to 2012. It examined three main approaches to census-taking: the traditional census approach, the use of administrative registers, and the use of continuous measurement (Royce, 2011). The research project concluded that the traditional census approach was the only viable methodology for the 2016 Census Program given the current Canadian context (Statistics Canada, 2012).

This research also focused on approaches that could be considered beyond the 2016 cycle, notably those that rely on an increased use of administrative data. Cyr (2012) produced a report on the state of administrative data in Canada and their readiness to achieve a population register. The report examined three options to achieve a population register: a register built by the government, a register built upon an existing administrative program, and a statistical population register (sometimes referred to as a virtual population register). The report concluded that, given the current situation in Canada, “limitations exist in all three scenarios explored and none are attainable at this point” but it also stated that “...the third scenario of a virtual population register can be investigated by Statistics Canada, at least from a research standpoint, as conditions appear to be in place to move forward”.

3. Research since 2012

This work was at the genesis of further research into the possibility of creating a statistical population register. Statistics Canada has now created a few prototypes of what it calls the Canadian Statistical Demographic Database (CSDD), built from various administrative sources. The prototypes built to date attempted to replicate the 2011 or 2016 Census head counts.

The basic approach used to this point can be best described with the work done on the second prototype of the 2011 CSDD. The starting point for this prototype was the ‘Ident’ file from the Canada Revenue Agency (CRA) with a vintage of December 31, 2011. This file contains all historical federal tax filers since 1984, along with their latest contact information. Information on children was obtained from the Canada Child Tax Benefit file that includes the Universal Child Care Benefit recipients, and the Canadian Birth Database developed from provincial and territorial vital statistics. Using files from IRCC, the CSDD included permanent immigrants whose landing date was between 1980 and the 2011 Census Day (May 10) and non-permanent residents who had a valid temporary resident, work or student permit, or refugee claimants. Deaths were removed from the CSDD using an auxiliary file especially created as part of the research, and called the Improved Canadian Mortality Database, combining information from the Canadian Mortality Database developed from provincial and territorial vital statistics, with additional deaths (usually abroad) declared in taxation files. Approximately 100,000 additional deaths were identified from the CRA files since 1981.

Emigration is not recorded in an exhaustive way in Canada. Individuals who appeared “inactive” in Canada were removed from the CSDD using an auxiliary file derived by Statistics Canada and called the Fiscal Activity Indicator File (FAIF). This file included 20 signals (flags) of activity on 14 different CRA files from 2000 to 2011. The derivation of probabilities of being in the population based on these flags will be further explored in future research. In the meantime, approximately 3.5 million individuals were removed from the second prototype using deterministic rules, for example if an individual had not filed an income tax report or been issued tax slips since 2008. Approximately 200,000 individuals who would not have shown on the CSDD otherwise were added based on the FAIF.

The Address Register was used to code addresses to the dwelling level (AR_UID). The last step in the creation of the CSDD was to assign individuals their most likely address or AR_UID on Census day. Approximately 17% of addresses could not be coded to the AR_UID because they were not precise enough, were incomplete or incorrect. They were assigned to the next most precise geographical location in the hierarchy (i.e., the block face).



4. Assessment of the research to date

Population counts derived from the second 2011 CSDD were assessed against the published 2011 Census counts and the official Population Estimates (PEs) that Statistics Canada produces through the Demographic Estimates Program. PEs are a key input for the calculation of revenue transfers and contributions that are allocated to provinces and territories from the federal government under the *Fiscal Arrangement Act*. Every year, more than \$60 billion dollars, mainly on a per capita basis, are transferred through that process, so PEs are considered as benchmarks when comparing results at aggregated levels.

The PEs are currently rebased every five years based on results of census coverage studies carried out to estimate how many individuals were missed or counted more than once. The census net undercoverage estimates are derived by sex, age group and province or territory. Official PEs are then produced by sex, age and geographic areas down to Census subdivisions (municipalities) from the published Census counts adjusted for Census net undercoverage and incompletely enumerated Indian reserves. At more disaggregated levels below municipality, the CSDD was assessed against the 2011 Census Response Database (RDB) containing collected data before editing and imputation.

Table 1 shows that although the second 2011 CSDD prototype overestimates the PEs, it performs relatively better than the Census at the national level. The same is true at the provincial and territorial levels. The CSDD results were within 1% of the PEs for 11 of the 13 provinces and territories. At lower levels of geography, indicators favor the Census more than the CSDD. As the level of geography decreases, the precision of the information on the CSDD also decreases, mostly because of the lack of precise residential address information on administrative files.

Table 1 CSDD and published Census counts compared to Censal Population Estimates, Total population, Canada, 2011

CANADA		Number			Relative Difference with PEs			Sex ratio
		Total	Male	Female	Total	Male	Female	Number of males for 100 females
	PE	34,273,205	16,977,217	17,295,988				98.2
	Census	33,476,688	16,414,229	17,062,459	-2.3%	-3.3%	-1.4%	96.2
	CSDD	34,379,148	17,067,560	17,311,588	0.3%	0.5%	0.1%	98.6

Note: CSDD does include records with missing sex, provinces/territories and/or age

This prototype was also analyzed against the 2011 Census Response Database (RDB) from a household perspective. The RDB contains 13,083,485 dwellings, of which 10,355,338 are also on the CSDD. Remember that it was not possible to code the addresses of 17% of the individuals on the CSDD. The CSDD also includes 800,650 dwellings that are not found on the RDB. Analyses revealed that dwellings displaying the same household composition on the RDB and CSDD are 88.4% of the time in the mail-out area for the census, which tend to be urban areas and have better address information.

The prototype was finally matched to the 2011 RDB using a conservative deterministic hierarchical linkage approach, as a limited number of waves were performed. A total of 29,948,224 of the 35,178,827 CSDD records were successfully matched. The quality of the linkage was verified through the use of a sample that was manually checked. The number of valid links was estimated to be 97.7%. The analysis of this match reveal that 73% of the matched records (individuals) were placed in the dwelling where the 2011 Census found them. This percentage varies by province and territory.

5. 2016 prototypes



Two 2016 CSDD prototypes were also built. The first one was produced in June 2016 to assess its capacity to support a traditional census during its collection activities. The second one was produced in January 2017 to assess capacity to reproduce the census head count in approximately the same time frame as the census releases its own head count. The approach to build the two prototypes was very similar to the one used for the second 2011 CSDD prototype with a few notable differences.

Preliminary versions of administrative data files had to be used for the June version. For example, only preliminary birth and death files were available after 2012. Similarly, very little information from the 2015 taxation year was known at the time of the creation of the first 2016 CSDD prototype, most of the activity occurring around April 2016.

Statistics Canada also used the Social Insurance Register (SIR) which contains information about individuals who have been issued a Social Insurance Number (SIN) since the introduction of the program in 1964. The SIN is an individual nine-digit number required to work in Canada or to have access to government programs. The SIR has greatly improved Statistics Canada's ability to link different administrative data files containing only partial identification information. For example, information about immigrants on immigration files is limited to names, date of birth and sex. It does not include the SIN. The CRA Statements of Remuneration Paid issued by employers contain the SIN and partial information on the name, but not the date of birth and the sex. Before the acquisition of the SIR, it was challenging to properly link immigration and taxation records and to determine, for example, whether recent immigrants were displaying sufficient fiscal activity to be deemed still in Canada and kept on the CSDD. The 2015 Indian Registry was also used. Other improvements were made based on the experience with the 2011 CSDD, in particular in the use of the Fiscal Activity Indicator File.

The type of analyses performed on the 2011 CSDD prototypes are being reproduced for the 2016 versions. In particular, the second 2016 CSDD results are being compared to the 2016 Census counts published on February 8, 2017 and the preliminary PEs as of July 1, 2016. These PEs are still based on the 2011 Census counts adjusted for census net undercoverage to which are added the estimated demographic changes for the period from May 2011 to June 2016. In that regard, they differ from the PEs that were used in the 2011 CSDD analysis.

Table 2 presents the results at the national level. Nationally, the final estimate of the 2011 Census net undercoverage rate was 2.3%, compared with 2.8% for the 2006 Census. Should this rate prove to be of the same order of magnitude in 2016 (will be available in fall 2018), it would mean that the second 2016 CSDD has overestimated the population produced by the PEs

Table 2 CSDD, published Census counts and censal Population Estimates compared to preliminary Population Estimates, Total population, Canada, 2011 and 2016

	2016		2011	
	Total	% of Preliminary PE	Total	% of Preliminary PE
Preliminary PE	36,286,425	--	34,880,500	--
Censal PE	?	?	34,273,205	-1.7
Census	35,151,728	-3.1	33,476,688	-4.0
CSDD	36,663,801	1.0	34,379,148	-1.4

6. Live application of the CSDD in the 2016 Census

Special circumstances during the 2016 Census collection prompted Statistics Canada to use a different collection approach for the area of Wood Buffalo in the province of Alberta. Large forest fires forced the evacuation of over 100,000 people in early May 2016, at the same time as census collection started. The communities in the area remained evacuated until the end of July, 2016, the point at which it was planned to end census collection.



Given the limited number of direct responses received or collected for this area, Statistics Canada leveraged work on the first 2016 CSDD to create a 2016 short form census record for approximately 60% of persons from the evacuated area, in time to enter the census data processing phase. Given the limited maturity of the CSDD process at this stage, a lot of manual work needed to be done to make this use possible. A series of key decisions were made about how to determine dwellings occupancy, and how to treat inconsistencies between administrative sources used to validate the work. These included addresses in the administrative sources which were not on the AR used as the frame in this area, finding address discrepancies among the various sources for some households, etc. The combination of direct response and the information from administrative sources was used to produce and release census results for the area as part of the planned census releases.

7. Potential applications for the 2021 Census

Results achieved to this point have convinced Statistics Canada to put in place a fully dedicated team from different areas of the organisation to continue this research. In addition to research activities to build a Statistical Population Register (SPR) which will be replacing the CSDD, a more exhaustive Statistical Building Register (SBgR) will be developed to provide an accurate list of buildings in Canada, and more particularly occupied private dwellings. Key to this research will be the acquisition of new administrative data sources, mainly from provinces and territories, to provide more precise physical addresses. Research will also be conducted on alternative collection methodologies for areas where the SBgR and SPR are not of sufficient precision, and for the collection of relevant census variables not available from administrative sources.

Although the ultimate goal of the research is to develop a new approach for future censuses based on a combined methodology that maximizes the use of information from administrative sources, Statistics Canada is also assessing which parts of the research could be applied to support collection operations for the 2021 Census. A relatively small and diminishing proportion of dwellings and households require in-field follow-up in the Canadian Census, but the difficult nature of many of these cases makes it increasingly inefficient to conduct telephone or in-person non-response follow-up. We are researching approaches for reducing the expensive in-person follow-ups of dwellings selected for the short form by using available administrative data from the SPR or the SBgR, without having an impact on the quality of results? Should a direct replacement approach be used, i.e., directly using information from the SPR for non-response cases, or would it be preferable to use it to guide whole household imputation? Potential impacts on current coverage study methodologies will also be assessed.

Another challenge for the current approach in Canada is the reduction of the 'noise' in the AR. The current approach for the maintenance of the AR creates a certain level of overcoverage. This is in part created by some lack of precision in the administrative sources used to update the AR, and by the conservative approach on the use of the register to avoid causing undercoverage of dwellings. Any problem of overcoverage in the register is addressed at the time of collection by ultimately visiting every non-response dwelling to determine their status. In the current Canadian approach, with 88.8% self-response realized in 2016, the larger share of dwellings requiring in-person follow-up are now overcoverage cases (400,000 in 2016) or unoccupied dwellings (1 million in 2016). About 1.3 million responses were obtained during non-response follow-up in 2016, which represents only 9% of all census responses. Considering the important share of the program's costs necessary to conduct in-person work, any reduction in workload made possible by the new registers needs to be seriously considered.

Research and testing will be conducted to see how the various sources used to create the SPR and the SBgR can be leveraged to determine the validity of dwellings on the frame, and their likely occupancy status on Census Day. The objective is to produce rules that would reduce the number of cases requiring in-person follow-up without impacting coverage or overall quality. For example, signals from administrative sources that put into question the existence of a dwelling, combined with information from the initial mail-out of census invitation letters such as post office returns, could be used to determine the validity of a dwelling. The same could be done regarding the occupancy status of a dwelling. For example, timely information on utility consumption at



a given address combined with information from the SPR and information from the initial mail-out, could be used to derive a predicted occupation score. This information could be used to decide if in-person follow-ups should be done, or on the number of times a dwelling should be visited.

These approaches to potentially reduce in-person follow-ups of short form cases in the 2021 Census will be further expanded in 2026 and beyond, replacing the direct collection of census information by data from alternative sources.

8. Conclusion

The Canadian Census has a long history of introducing new and innovative approaches and methodologies to improve the quality of the outputs, the efficiency of the program, and to reduce burden on Canadians. These are still the drivers for the research into the use of alternative data sources and approaches to conduct future censuses. Moreover, the SPR and SBgR, along with Statistics Canada's Statistical Business Register (SBR), offer the opportunity to anchor the Canadian national statistics system on a set of interconnected base statistical registers in the future. Such a solid and common foundation will serve the coherence of the national statistics system and provides an efficient way to re-use existing data and to collect additional data via surveys to meet new data needs.

Biography

- 1) Rodrigue, J.-F., Mathieu, P., Morin, J.-P. and Taylor, J., (2012) « La méthodologie de collecte du Recensement et de l'Enquête nationale auprès des ménages de 2011 au Canada », Paper presented at the *Journées de méthodologie Statistiques* conference, Paris, France
- 2) Hamel, M. (2012), Internet data collection in the Canadian Census of Population, UNECE Expert Group Meeting on Censuses Using Registers, Geneva, Switzerland
- 3) Dolson, David (2011), Administrative Data Use in a Traditional Census, Paper presented at ISI
- 4) Mc Leish, S. (2016) Adding immigrant admission category to the Canadian census of population, UNECE Work Session on Migration Statistics, Geneva, Switzerland
- 5) Hamel, M. and Laniel N. (2013), Producing official statistics via voluntary surveys – the National Household Survey in Canada, Paper presented at ISI, Hong Kong
- 6) Royce, D. (2011), Preliminary Report on Methodology Options for the 2016 Census, Statistics Canada. <http://www12.statcan.gc.ca/strat/index-eng.cfm>
- 7) Statistics Canada (2012). Final Report on 2016 Census Options: Proposed Content Determination Framework and Methodology Options. http://www12.statcan.gc.ca/census-recensement/fc-rf/reports-rapports/r2_index-eng.cfm
- 8) Cyr, A. (2012). State of Administrative Data and their Readiness to Achieve a Population Register in Canada for the Census of Population, Statistics Canada internal report.