

Census of Population Reference Product

Guide to the Census of Population, 2016

Census of Population, 2016

Release date: November 29, 2017



Statistics
Canada

Statistique
Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at STATCAN.infostats-infostats.STATCAN@canada.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- | | |
|---|----------------|
| • Statistical Information Service | 1-800-263-1136 |
| • National telecommunications device for the hearing impaired | 1-800-363-7629 |
| • Fax line | 1-514-283-9350 |

Depository Services Program

- | | |
|------------------|----------------|
| • Inquiries line | 1-800-635-7943 |
| • Fax line | 1-800-565-7757 |

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under “Contact us” > “Standards of service to the public.”

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

- | | |
|----------------|--|
| . | not available for any reference period |
| .. | not available for a specific reference period |
| ... | not applicable |
| 0 | true zero or a value rounded to zero |
| 0 ^s | value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded |
| ^p | preliminary |
| ^r | revised |
| x | suppressed to meet the confidentiality requirements of the <i>Statistics Act</i> |
| ^E | use with caution |
| F | too unreliable to be published |
| * | significantly different from reference category ($p < 0.05$) |

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2017

All rights reserved. Use of this publication is governed by the Statistics Canada [Open Licence Agreement](#).

An HTML version is also available.

Cette publication est aussi disponible en français.

Table of Contents

Chapter 1 – Introduction	1
General information	1
Why does Statistics Canada conduct the Census of Population?	1
Why is the census in May?	1
Privacy and confidentiality	1
Consent to release personal information	2
Retention of census information	2
Official languages	2
Other languages / alternative formats	3
Chapter 2 – Census history	4
Chapter 3 – Communications	7
Key activities	7
Advertising	7
Outreach	7
Public relations	7
Media relations	8
Social media	8
Aboriginal strategy	8
Census website	8
Chapter 4 – Content determination	9
How are the census questions determined?	9
May 2014 Content Test	10
Changes to the 2016 Census questionnaire	11
Chapter 5 – Census of Population questionnaires	12
Forms used to enumerate usual residents of private dwellings	12
Forms used to enumerate usual residents of collective dwellings	13
Accessibility	13
Chapter 6 – Online questionnaire	14
Introduction	14
Reduced costs and increased quality	14
Mobile-friendly responsive design	14
Security	14

Chapter 7 – Field operations	15
Introduction	15
Census delivery methods	15
Census wave approach	16
Census collection methods	16
Census Help Line	18
Data collection verification and follow-up activities for the 2016 Census	18
Chapter 8 – Processing	19
Introduction	19
Receipt and registration	19
Imaging and data capture	19
Edits	20
Coding	20
Edit and imputation	20
Chapter 9 – Sampling and weighting for the long form	21
Chapter 10 – Data quality assessment	24
Introduction	24
Sources of error	24
Measuring data quality	24
Certification	25
Response rate for the 2016 Census of Population	26
Quality of long-form questionnaire estimates	26
Comparability of estimates from the 2016 Census long-form questionnaire with estimates from the 2011 National Household Survey	28
Comparability of estimates from the 2016 Census long-form questionnaire with estimates from the 2006 Census long-form questionnaire	31
Comparability of the variability of estimates from the 2016 Census long-form questionnaire with that of estimates from the 2011 NHS and the 2006 Census long-form questionnaire	31
Description of standard error	32
Derivation of the coefficient of variation (CV)	32
Distinction between standard error, response rate and global non-response rate	32
Conceptual and methodological differences between standard errors for the 2016 Census long-form questionnaire, the 2011 NHS and the 2006 Census long-form questionnaire	32
Chapter 11 – Dissemination	35
Introduction	35
How census data are used	35
Overview of 2016 Census of Population products and services	36

Connecting with Canadians.....	38
Availability of products and services.....	39
Data suppression	39
Chapter 12 – Census geography.....	44
Introduction.....	44
National Geographic Database	44
Hierarchical model of geographic areas for collection.....	45
Hierarchical model for the dissemination of geographic areas	45
The hierarchy of geographic areas: understanding the hierarchy, how geographies are related and data analysis	46
Geographic areas for dissemination.....	46
Non-standard or user-defined geographic areas for dissemination	48
Appendix 1.1 – Legislation.....	49
Introduction.....	49
Requirement to respond.....	50
Statutory references to the census	51
Constitutional law.....	51
Appendix 1.2 – Incompletely enumerated Indian reserves and Indian settlements.....	52
Appendix 1.3 – Difference between census counts and population estimates	54
Appendix 1.4 – Note describing the Wood Buffalo census subdivision data collection methodology and the use of administrative data sources	55
Introduction.....	55
Reference date	55
Data collection	55
Administrative data.....	55
Processing	56
Data quality for population and dwelling counts.....	56
Short-form questionnaire data quality.....	56
Appendix 1.5 – Information produced from the 2016 Census of Population.....	57
Who is included in the population of Canada?	57
Where are people counted?	57
Who is included in the data for each topic?	57
Appendix 1.6 – Results of the 92-year consent question.....	59
Release of personal census information after 92 years	59

Appendix 1.7 – Percentage change for population and dwelling counts at various levels of geography	60
Appendix 1.8 – Evaluation of the impact of updating 2016 Census language data	61
Potential impact on other variables	61
Evaluation of the impact on the weighting of the 2016 Census long-form questionnaire sample	67
Appendix 1.9 – Standard errors of estimates from 2016 Census long form sample	70
Highlights	70
Analysis of the 2006 Census long form, 2011 NHS and 2016 Census long form standard errors	70

Chapter 1 – Introduction

General information

Canada's most recent census was held in May 2016.

Census of Population data are important for all communities and are vital for planning services such as schools, day care, family services, housing, police services, fire protection, roads, public transportation and skills training for employment. Data are used by governments, businesses, associations, organizations and many others to make important decisions. The information provided by the 2016 Census of Population accurately reflects Canada's changing society.

Why does Statistics Canada conduct the Census of Population?

The *Statistics Act* requires that censuses of population and agriculture be conducted every five years, in years ending in '1' and '6.' The act also outlines the mandatory requirements for completing and returning census questionnaires. (For more information, refer to [Appendix 1.1, Legislation.](#))

The Census of Population is the primary source of sociodemographic data for specific population groups such as lone-parent families, Aboriginal peoples, immigrants, seniors and language groups.

Adjusted population counts from the census are used as the base for the Population Estimates Program. Population estimates in turn are used to determine representation in Parliament, to calculate transfer payments between levels of government and to support various government programs across the country.

Why is the census in May?

The timing of the Census of Population is driven by the need to maximize the number of Canadians who are home during enumeration. The May 10 date allows collection procedures to run smoothly, therefore costing less.

In addition, the mid-May date allows more time for final follow-up to be completed before the busy summer holiday period begins. Improved population coverage means better-quality data.

Privacy and confidentiality

Providing personal information to anyone—whether in a census, a survey, or any other manner—does involve some loss of privacy. But in virtually any country, it is recognized that the benefits to the public of accurate census data far outweigh this loss of privacy.

In Canada, great care is taken to ensure that information collected from the census is clearly in the public interest and cannot be obtained effectively from other sources.

Statistics Canada places the highest priority on maintaining the confidentiality of individual questionnaires. Stringent instructions and procedures have been implemented to ensure that confidentiality is maintained at all times. For instance, census data are processed and stored on a highly restricted internal network and cannot be accessed by anyone who has not taken the oath of secrecy.

What's more, data releases are screened so that anonymity is assured. Names, addresses and telephone numbers are not part of the census database used for dissemination, and private contractors do not have access to confidential data.

Only a limited number of Statistics Canada employees have access to personal and confidential information. Those employees are able to collect, handle and process completed questionnaires.

All Statistics Canada employees must take an oath of secrecy, which remains in effect for life, even after employment is completed. Employees are subject to fines and/or imprisonment should they reveal identifiable information derived from the census. Any possible breach of the confidentiality of census questionnaires is an exceedingly serious matter which would be investigated immediately and thoroughly, and subject to the full force of the *Statistics Act*.

Consent to release personal information

Respondents were asked for permission to make their information available to Library and Archives Canada (LAC), 92 years after the 2016 Census, so that it may be accessed by family members, genealogists, historical researchers, academics and journalists.

The person who completed the census questionnaire was instructed to consult with all members of the household who were included in the questionnaire before answering the consent question.

This information is important for preserving Canada's history for future generations. Only records for which consent was received will be released in 92 years.

Respondents who wish to change their response to the consent question on the 2006, 2011 or 2016 Census of Population questionnaire can do so by completing a [Request for change: Release of personal census data](http://www12.statcan.gc.ca/census-recensement/2016/ref/consent-consentement/change-modification-eng.cfm) (<http://www12.statcan.gc.ca/census-recensement/2016/ref/consent-consentement/change-modification-eng.cfm>).

Retention of census information

Statistics Canada, in consultation with Library and Archives Canada, determines the best means of preserving the records so that information can be released in 92 years (for respondents who consented to their information being made public at that time).

Census records up to and including the 1916 Census are available either online or as microfilm copies through Library and Archives Canada. The 1921 Census records have also been released to the public (through [ancestry.ca](http://www.ancestry.ca) (<http://www.ancestry.ca/>) by LAC).

A microfilm copy of the questionnaires from the 1921 to 2001 censuses is held by Statistics Canada. The original paper questionnaires have been shredded and destroyed.

The 2006 and 2011 censuses and the 2011 National Household Survey (NHS) (which replaced the long-form census in 2011) were not microfilmed. Paper questionnaires were converted to digital images and an archival database containing all responses (including those submitted via the Internet) was created. The original paper questionnaires were shredded and destroyed. The same procedure for retention of census information will be used for the 2016 Census.

Official languages

As early as 1871, census questionnaires were produced in English and French. This tradition became law in 1988 under the *Official Languages Act*. This act states that English and French are the official languages of Canada and that service to the public must be provided in both languages.

As in previous censuses, procedures were in place for the 2016 Census to ensure that members of the public received services in the official language of their choice.

Other languages / alternative formats

In addition to English and French, the 2016 Census questions and the reasons why the questions are asked were available in 11 Aboriginal languages and 11 Immigrant languages, as well as in Braille, in an audio version and as a sign language video.

It is important to note that the questions were translated for reference purposes only. The census questionnaire had to be completed online or on paper, in either English or French.

Aboriginal languages:

- Atikamekw
- Denesuline (Chipewyan)
- Dogrib
- Inuktitut (Nunavik)
- Inuktitut (Nunavut)
- Montagnais
- Northern Quebec Cree
- Oji-Cree
- Ojibway
- Plains Cree
- Swampy Cree

Immigrant languages:

- Arabic
- Chinese (simplified)
- Chinese (traditional)
- Korean
- Persian (Farsi)
- Portuguese
- Panjabi (Punjabi)
- Russian
- Spanish
- Tamil
- Vietnamese

Chapter 2 – Census history

Over the years, various changes have been introduced to the Census of Population in Canada. Below is a summary.

1666 – The first Canadian census was taken in New France by Intendant Jean Talon. The recorded population (excluding Aboriginal persons and royal troops) was 3,215. Information was obtained on age, sex, marital status and locality. In addition, the census identified professions and trades for 763 persons.

1666 to 1867 – Numerous censuses were taken at irregular intervals in the colonies of France and Britain that became parts of Canada.

1867 – *The Constitution Act, 1867* (formerly the *British North America Act*) included the requirement that a census be taken every 10 years (decennially) in order to determine representation by population in the new Parliament.

1871 – The first decennial census was taken in this year. The census enumerated the population of the four original provinces (Nova Scotia, New Brunswick, Quebec and Ontario). Manitoba and British Columbia, which had also joined Confederation, were surveyed separately.

The 1871 Census was the first to use the *de jure* method of enumeration rather than the *de facto* method used in Europe both then and now. The *de facto* method enumerates people where they are found on Census Day. The *de jure* method enumerates people according to their usual place of residence.

1881 – All census enumerators were required to take an oath of secrecy, a pledge still required today. The census was extended to include British Columbia, Manitoba and Prince Edward Island.

1891 – The population was prepared for the census enumerator's visit through announcements in newspapers and from pulpits.

1896 – A mid-decade census was held in Manitoba beginning in 1896 and then in Saskatchewan and Alberta beginning in 1906. These censuses were needed to measure the rapid growth taking place in the West.

1901 – The census content was expanded to include religion, birthplace, citizenship and period of immigration.

1905 – The census office became a permanent bureau of the federal government.

1906 – Beginning in 1906, the Prairie provinces of Manitoba, Alberta and Saskatchewan began to take a separate census of population and agriculture every five years to monitor the growth of the West.

1912 – Responsibility for conducting the census was transferred from the Department of Agriculture to the Department of Trade and Commerce.

1918 – The Dominion Bureau of Statistics was created.

1931 – Though compilation and tabulation for the 1931 Census were still carried out with mechanical equipment, a new sorter-tabulator developed by an employee of the Dominion Bureau of Statistics made production 50 times faster by allowing a whole data card to be read at once rather than one column at a time.

1941 – Sample information was collected for the first time. One in 10 households were asked additional content about their dwelling (type, number of rooms, cooking fuel used, etc.).

1951 – Canada's first census as a nation of 10 provinces and two territories used 'mark-sense.' This technology allowed for a generation of punch cards, greatly reducing processing time and costs.

1956 – A quinquennial (every five years) Census of Population and Agriculture was held in all provinces across the country, replacing the mid-decade censuses of the Prairie provinces.

The 1956 Census included questions on radios, colour televisions, home freezers and vacation homes. These kinds of questions were dropped after the 1966 Census as they were incorporated into various household surveys.

1971 – Under the *Statistics Act* of 1971, the Dominion Bureau of Statistics was renamed Statistics Canada. The act also confirmed that a Census of Population and a Census of Agriculture would be taken every five years (quinquennially).

Self-enumeration was first introduced in 1971. With the exception of Indian reserves and remote areas, census questionnaires and completion instructions were dropped off at private homes and respondents were asked to complete their own questionnaires. In population centres of 10,000 persons or more, respondents were asked to mail their completed questionnaires back in a pre-addressed envelope. In other areas, questionnaires were picked up by census enumerators.

1991 – All respondents in self-enumeration areas (over 98% of the population) were asked to return their completed census questionnaires by mail. The return rate was 85%, with more than 27 million people in over 10 million households counted.

Information on common-law partners was also collected for the first time.

2001 – For the first time, data were collected on same-sex couples.

2006 – The questionnaires were delivered by Canada Post to about 70% of households. The remaining 30% received the questionnaire from an enumerator as in previous censuses.

For the first time, all Canadians could answer the census questionnaire online.

This was also the first time people were asked if they agreed to have their personal information released in 92 years for the purpose of research and education.

Respondents were also asked whether they would give Statistics Canada permission to access their tax files. This permission was sought in an effort to reduce response burden.

The definition of spouse was expanded to include same-sex married couples.

2011 – The positive response to the online option (18.5%) in 2006 prompted a major change in methodology for the 2011 Census. In May, a letter was delivered to 60% of Canadian dwellings. This letter replaced the traditional paper questionnaire and explained how respondents could complete the questionnaire online.

About 20% of dwellings received a questionnaire package by mail. For the remaining 20%, questionnaires were dropped off by enumerators.

Information previously collected by the mandatory long-form census questionnaire was collected as part of the new voluntary National Household Survey (NHS).

The 2011 Census of Population questionnaire (short form) consisted of the same content as the 2006 Census short-form questionnaire, with the addition of two questions on language.

2016 – In November 2015, the government reinstated the long-form census questionnaire, replacing the National Household Survey. Most households (75%) received the short-form census questionnaire while 1 in 4 households (25%) received the long-form questionnaire.

To reduce the burden on Canadians, manage collection costs and get the most accurate information on income, Statistics Canada accessed income information retrieved from personal income tax and benefits files, replacing income-related questions on the 2011 National Household Survey questionnaire and on previous censuses.

Chapter 3 – Communications

The Census Communications Team provided pro-active communications programs that were creative and cost-effective, while maintaining a positive, non-partisan, corporate image of Statistics Canada.

For the 2016 Census, census communication materials focused on:

- increasing awareness of the 2016 Census
- promoting self-response through online collection
- increasing self-response rates by encouraging households to complete and return their census questionnaire in May 2016
- increasing participation among hard-to-enumerate groups
- promoting Statistics Canada's commitment to protecting the personal information of all Canadians.

Communications activities for the 2016 Census of Population took place throughout the entire collection process to inform Canadians about the benefits of completing the 2016 Census questionnaire and encourage them to complete their questionnaire online.

Key activities

The main activities of the 2016 Census Communications Program included advertising, outreach, public and media relations, social media, implementing an Aboriginal strategy and managing the census website.

Advertising

The advertising campaign was used to inform Canadians of the importance, relevance, and security of information collected by the Census of Population.

Television, radio, print and digital channels were used during the months of April and May 2016 to encourage households to complete their questionnaire, with a focus on online response.

Outreach

The 2016 outreach strategy extended the reach of traditional census promotional activities by engaging with key stakeholders at the community level to promote census job opportunities, as well as the benefits of the census.

Outreach activities were prioritized to target specific organizations whose constituencies were among the hard-to-enumerate and included contact with community leaders, data users and organizations, among others.

Public relations

Public relations activities supported census recruitment and collection. Primary activities included contacting federal departments and agencies to request that they support the census by disseminating census messaging to the public.

Public relations activities also reached out to specific groups, such as educators, foreign embassies in Canada and police services.

Media relations

Media relations maintained and monitored media coverage across the country during the census collection period (May, June and July 2016). Traditional and new media were used to report on key aspects and operational milestones of the census.

A media strategy was implemented to:

- maximize media coverage and generate public awareness of the census
- generate interest among traditional and new media
- generate media interviews with designated spokespersons
- respond to concerns in a timely manner.

Social media

Over the course of the past few years, Statistics Canada has been using social media to increase access to its relevant, accurate and timely statistical information and foster engagement, cooperation and information-sharing with the public. Statistics Canada's official social media accounts include Twitter, Facebook and YouTube.

The official social media accounts were used intensively to communicate and engage with the public during the 2016 Census of Population recruitment and collection phases.

Aboriginal strategy

To improve engagement levels among Aboriginal communities, Statistics Canada developed strategic plans, used internal and external partnerships and created customized support materials. The purpose of the Aboriginal strategy was to increase awareness among Aboriginal peoples about job opportunities within their communities, increase the number of Aboriginal applicants for census jobs, support the collection of data in Aboriginal communities and provide support materials that encouraged participation while respecting cultural diversity.

Census website

The 2016 Census website promoted the benefits of the census, provided information and instructions for completing the questionnaires and supported census recruitment.

Chapter 4 – Content determination

How are the census questions determined?

Determining the content of a census is an ongoing process involving user consultations, content testing and content approval. Content refers primarily to the content of the census short- and long-form questionnaires (also referred to as the 2A and 2A-L Forms).

Before each census, Statistics Canada undertakes a user consultation and testing program to determine the census questions, taking emerging social and economic issues into account.

While consultation is an ongoing process and data users are encouraged to communicate their data needs at any time, Statistics Canada holds a formal public consultation aimed at gaining feedback from census data users, on how they use the data and to identify data gaps. Results from this consultation feed into the content testing process, followed by the formulation of recommendations on final questionnaire content and the subsequent approval process.

For the 2016 Census of Population, Statistics Canada conducted comprehensive consultations with federal departments, provincial/territorial and municipal government organizations, Aboriginal organizations, organizations representing official language minority communities, organizations representing or providing services to Canadians with disabilities, immigrant communities, academia, non-profit organizations and the general public. These data users and interested parties across Canada were asked for their requirements for census information. The content was then tested qualitatively and quantitatively.

Final recommendations regarding content were based on the census' content determination framework which assesses users' information needs (e.g., strength of user need, size of population of interest, suitability of alternative sources); respondent burden and privacy concerns; and Statistics Canada considerations such as operational factors and costs.

The final questions were presented to and approved by Cabinet, and were published in the *Canada Gazette*.

Below is a description of how content consultation, content testing and content approval were done for the 2016 Census.

Step 1 – Content consultation

An ongoing consultation process with census data users allows Statistics Canada to identify if the content is relevant, how the data are used, and the importance of the census to the Canadian population. A formal content consultation process is typically set in place four years before a census year. During that time, Statistics Canada invites data users, stakeholders and the public to provide their feedback on what information is used, for what purpose, and what, if any, data gaps there may be that could be explored for inclusion in the next cycle of the census.

For the 2016 Census, the 2016 Census Strategy Project included consultations with senior representatives from various federal, provincial, territorial and municipal governments in the summer and fall of 2011, in order to develop a framework for determining census content. Furthermore, the Census Marketing Team (Census Operations Division) conducted the formal 2016 Census Program Content Consultation from September to November 2012: a two-tier consultation process allowed for feedback from all census data users.

The following two-tier approach was used:

1. Tier 1: A general call, broadly promoted, sought feedback from data users using a 'Discussion Points' questionnaire which asked users to rank the importance of census topics, the uses of essential topics, input on suggested modifications and questions about the level of geography used.

2. Tier 2: Selected data users were asked to report on their priority census data uses and requirements. This included more than 300 organizations (senior representatives of federal and provincial governments, municipalities, community groups, academia, business and special interest groups).

Further details can be found in the [Census Program Content Consultation Report, Census year 2016](http://www12.statcan.gc.ca/census-recensement/2016/consultation/contentReport-RapportContenu/index-eng.cfm) (<http://www12.statcan.gc.ca/census-recensement/2016/consultation/contentReport-RapportContenu/index-eng.cfm>), Catalogue no. 92-137-X, release date: September 10, 2014.

Step 2 – Content testing

Each cycle of the census involves extensive testing. Once feedback is obtained from data users during consultation, decisions are made by subject-matter experts about content that could be tested for the upcoming census.

The first part of this process involves qualitative testing. For the 2016 process, qualitative testing of content was conducted with the help of the Statistics Canada Questionnaire Design Resource Centre (QDRC) in three testing periods: June 2013, September 2013 and January 2014, which included focus groups to test the collection of the Social Insurance Number (SIN). Statistics Canada also conducted a focus group test within the deaf community to gain feedback as to whether the language questions on the census were well understood by this particular population. The deaf community focus group was held in the fall of 2014. The QDRC also conducted a test on mobile devices in February 2016. The purpose of this test was to gain feedback on how participants responded to the short- and long-form census questionnaires using their own mobile devices.

After qualitative testing is complete, the second part of the testing process involves quantitative testing. For the 2016 Census, the quantitative content test was held in May 2014.

May 2014 Content Test

In May and June 2014, a content test was conducted. This test was designed to meet two main objectives: to assess the impact of any proposed content changes and to measure the impact of adding a Social Insurance Number question. The Content Test required the development of different test and control forms, in English and in French, distributed among 11 analysis panels (independent samples).

Various modifications were tested, including changes to the length of the questionnaire, the wording of some questions, as well as examples and reference text to aid in responding to the questionnaire.

In total, 55,000 dwellings divided into 11 analysis panels were contacted during the Content Test. Both paper and online forms were tested, and questionnaires were available in both official languages. The mode of collection for this test was respondent self-enumeration followed by enumerator-assisted non-response follow-up for some panels.

Results from the Content Test shaped the final short-form and long-form questions that were presented to the Government of Canada for approval.

Details about this test can be found in the report [2016 Census Program Content Test: Design and Results](http://www12.statcan.gc.ca/census-recensement/2016/consultation/92-140/92-140-x2016001-eng.cfm) (<http://www12.statcan.gc.ca/census-recensement/2016/consultation/92-140/92-140-x2016001-eng.cfm>), Catalogue no. 92-140-X2016001, release date: April 1, 2016.

Changes to the 2016 Census questionnaire

The most significant change to the 2016 Census was the decision by the government to reinstate the mandatory long-form census questionnaire. In 2016, a sample of 25% of Canadian households received a long-form questionnaire, which also includes the short-form questions. The other households received a short-form questionnaire. The decision to re-instate the mandatory long form did not change the content of the long-form questionnaire.

The main content changes introduced for the 2016 Census of Population are the following:

On the short form and long form

- The questions on sex, date of birth, age and marital status were changed to an interrogative format.
- Statistics Canada informed respondents that their income information would be retrieved from personal income tax and benefits files, replacing income-related questions asked on the long form in previous censuses. Income data were obtained for all respondents and will be disseminated with the short- and long-form census variables.

On the long form

- The question on religion was not included in the 2016 Census since it is only asked every 10 years. The question will be considered for inclusion in the 2021 Census.
- The questions on activities of daily living were redesigned into one question, with sub-items providing more detail about the type of difficulties a person may have doing certain daily activities.
- The online long form had more detailed flows in Step C to help respondents determine who should be included on the census form. This took advantage of functionalities available to online questionnaires.
- Questions from the short form and long form were integrated into one form for 2016. The long-form paper questionnaire contained a transition message (introducing the long-form questions) located between the last question of the short form (Question 10) and the first question of the long form (Question 11). Question 10 and the transition text were eliminated from the online form to improve the flow. However, since the decision to reinstate the mandatory long form occurred in November 2015, it was too late to modify and re-print the paper long forms.

Step 3 – Content approval

Once Statistics Canada has determined the proposed census content, the content is sent to Cabinet for approval and an Order-in-Council (OIC) is signed by the Governor General. Once this approval has been obtained, there is a final approval process by Statistics Canada with an order signed by the Chief Statistician.

For the 2016 Census, the OIC was signed by the Governor General on January 29, 2016, and the final approved 2016 Census content and the OIC were published in Part 1 of the *Canada Gazette* (<http://canadagazette.gc.ca/rp-pr/p1/2016/2016-02-06/html/order-decret-eng.body.html>) on February 6, 2016.

The internal prescription order was signed by the Chief Statistician on February 25, 2016. All [2016 Census questionnaires](http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvInstrumentList&Id=152274) (<http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvInstrumentList&Id=152274>) are available on the Statistics Canada website.

Chapter 5 – Census of Population questionnaires

Most census data are collected using the [census short form](http://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=getInstrumentList&Item_Id=295241&UL=1V&) (http://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=getInstrumentList&Item_Id=295241&UL=1V&) or [census long form](http://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=getInstrumentList&Item_Id=295122&UL=1V&) (http://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=getInstrumentList&Item_Id=295122&UL=1V&). In 2016, a sample of 25% of Canadian households received a long-form questionnaire. Income data were obtained from personal income tax and benefits files (Also refer to: Section on [Changes to the 2016 Census Questionnaire](#)). Additional immigration data on admission category were obtained from administrative files from Immigration, Refugees and Citizenship Canada.

Forms used to enumerate usual residents of private dwellings

Most of the population in Canada reside in private dwellings. For residents of private dwellings, the census data are collected primarily by having one adult member of the household respond on behalf of the entire household through self-enumeration using an online form.

The short-form questionnaire (Forms 2A, 3A, 2C)

The census is the primary source of exhaustive demographic data in Canada and in 2016, collected the following information:

- date of birth and age
- sex
- relationships of household members (including marital or common-law status)
- knowledge of official languages
- language spoken most often
- other language spoken regularly
- first language learned
- access to personal 2016 Census information 92 years after the census.

Form 2A: The short form is used to enumerate all usual residents of all private dwellings in the 2016 Census.

Form 3A: The short form for individuals. This form is used to enumerate one person. It is delivered to usual residents in private dwellings who wish to be enumerated separately from other members of the household (e.g., roomers, lodgers, boarders).

Form 2C: The short form for persons living abroad. This form is very similar to the Form 2A and is used to enumerate residents who are overseas. In 2016, this included Canadian government employees (federal and provincial) and their families, and members of the Canadian Forces and their families.

The long-form questionnaire (Forms 2A-L, 2A-R)

Complementing the data collected by the short-form questionnaire, the long-form questionnaire is designed to provide information about people in Canada based on their demographic, social and economic characteristics. In addition to the short-form information, the 2016 long form collected the following information:

- activities of daily living
- sociocultural information
- mobility
- place of birth

- education
- labour market activities
- housing
- access to personal 2016 Census information 92 years after the census.

Form 2A-L: The long form. This form contains the census long-form questions.

Form 2A-R: This form is very similar to Form 2A-L and is used in early enumeration, canvasser and reserve areas only. It contains the census long-form questions with examples adapted for Indian reserves, and an additional question on Indian band housing.

Forms used to enumerate usual residents of collective dwellings

A collective dwelling refers to a dwelling of a communal, institutional or commercial nature. Most of the population in Canada reside in private dwellings, but a small portion lives in facilities such as nursing or senior homes, jails, lodging or rooming houses, or even hotels or motels.

People may live in a collective dwelling either temporarily or permanently. For instance, guests of hotels and motels typically stay for a short time and are considered temporary residents, or they may be considered foreign residents if their main home is outside Canada. Guests or the hotel's owners or managers could live there permanently as their main (or only) residence. Such persons are considered 'usual residents' at that address and need to be counted at that address for census purposes.

For the usual residents of institutional collective dwellings, census short-form information was collected in 2016 using Forms 2A and 3A.

Accessibility

Forms 2A and 2A-L were available in electronic format online and in paper format, including large print. All other forms were available only in paper format.

The online forms were designed to comply with Web Content Accessibility Guidelines, to ensure that they were accessible for respondents using assistive technologies.

All census forms were available in both official languages. In addition to English and French, the 2016 Census questions were available in 11 Aboriginal languages and 11 Immigrant languages (see [Chapter 1](#) for the list of languages). Respondents could obtain a copy of the census questions in any of these languages by visiting the census website or by calling the Census Help Line.

Chapter 6 – Online questionnaire

Introduction

The use of the online questionnaire in the census has continued to increase since its introduction in 2001. In 2016, the census form was submitted online for 68.3% of private dwellings, which is an increase from 53.9% in 2011 and 18.3% for the 2006 Census.

Reduced costs and increased quality

Online questionnaires provide cost savings and data quality gains over the use of paper forms. Online forms have better response rates for individual questions and better quality data since there are edits that prompt a respondent when invalid data are entered or data are missing. The online census questionnaires also prefilled fields based on information provided by the respondent (e.g., the names of members of the household) which reduced the likelihood of one member's answers being recorded under another person. Automated skip patterns alleviated respondent burden by skipping questions that were not applicable. The online forms also had help information available for respondents who wanted additional information about a census question. The census long form had a stop and finish later feature, so that respondents could save the form and return later. A password was used to retrieve the saved questionnaire.

Mobile-friendly responsive design

The Treasury Board of Canada Secretariat required that the 2016 Census online forms use the Web Experience Toolkit (WET), which is used to build and maintain innovative websites that are mobile friendly. WET has the ability to adjust the layout of the web page according to the screen size of the mobile device. The online questionnaires were developed using WET to prepare for the increase of mobile users in Canada. In 2016, 79.6% of households used a personal computer, 12.9% used a tablet and 7.5% used a smartphone to submit the census questionnaire.

Security

Statistics Canada takes the protection of confidential information provided online seriously. A secure login process and strong encryption are key elements in helping to prevent anyone from eavesdropping or tampering with census information.

To protect the security of personal information on the Internet, Statistics Canada incorporated the following safeguards:

- Strong encryption technologies ensure security of data passing between respondents' computers and the web server.
- Data submitted to the web servers are encrypted before being stored and remain encrypted until they are transferred to the high-security internal network.
- Census data are processed and stored on a high-security internal network.
- Powerful firewalls, intrusion detection and stringent access control procedures limit access to back-end systems and databases. Census employees who have proper authorization and who have affirmed an oath of secrecy can access census data but only from secure Statistics Canada locations.

Chapter 7 – Field operations

Introduction

During the field operations' data collection phase, the objective was to ensure that responses were obtained from all households in Canada. Field operations included: delivering questionnaires, following up with non-respondents, conducting interviews, determining the occupancy status of a dwelling.

Census delivery methods

For most private dwellings (99%), respondents were asked to complete the questionnaire for themselves and all members of the household.

Starting May 2, 2016, all private dwellings in the mail-out (MO) area (approximately 82% of private dwellings in Canada) received a bilingual invitation letter to complete the questionnaire online. As in 2011, this letter contained a secure access code (SAC), the URL of the 2016 Census website, as well as a telephone number, allowing the respondent to request a paper questionnaire if preferred.

For 17% of dwellings, enumerators dropped off paper questionnaire packages. List/leave (L/L) door-to-door delivery took place from May 2 to May 9, 2016. During the L/L operation, census enumerators listed all private and collective dwellings in specific areas in their Visitation Record (VR). The paper questionnaire had a SAC in the upper right-hand corner so that respondents could fill out the questionnaire online.

For the remaining 1% of dwellings, the questionnaires were completed by having enumerators conduct personal interviews (canvasser areas).

Census wave approach

Statistics Canada implemented a wave approach for the 2016 Census. The following table outlines the key dates for the different waves in mail-out (MO) and list/leave (L/L) areas.

Mail-out (MO) refers to the mail delivery of questionnaires. Each census questionnaire was directed to the dwelling rather than to a specific person.

List/leave (L/L) refers to non-mail out areas. Drop-off needed to be done in person because mail delivery to a civic address was not possible.

Collection phase	Main activity	Coverage	Key start date
Wave 1	MO areas received an invitation letter with a SAC. No questionnaire package mailed.	All MO dwellings (82% of all dwellings).	May 2, 2016
	L/L areas received a questionnaire package which included paper versions of short or long forms in both official languages.	All L/L dwellings (17% of all dwellings).	May 2-9, 2016
Wave 2	MO areas, excluding early non-response follow-up (NRFU) areas, received a reminder letter with a SAC.	All non-responding MO dwellings, excluding those in early NRFU areas.	May 10, 2016
	MO areas (early NRFU areas only) received a questionnaire package.	All non-responding MO dwellings, in early NRFU areas only.	May 10, 2016
	L/L areas received a reminder card.	All L/L dwellings.	May 10, 2016
Wave 3	MO areas (excluding early NRFU areas) received a questionnaire package.	All non-responding MO dwellings, excluding those in early NRFU areas.	May 18, 2016
	MO areas (early NRFU areas only): start of non-response follow-up (NRFU) by enumerator.	All non-responding MO dwellings, in early NRFU areas only.	May 20, 2016
	L/L areas: start of non-response follow-up (NRFU) by enumerator.	All non-responding L/L dwellings.	May 20, 2016
Wave 4	MO areas (excluding early NRFU areas): voice broadcast message and start of NRFU.	All non-responding MO dwellings, excluding those in early NRFU areas.	Voice broadcast: May 30-31, 2016 NRFU: June 1, 2016

Census collection methods

Enumeration of usual residents of private dwellings

Most residents of private dwellings completed a form by themselves. In some regions, interviews were conducted.

Self-enumeration

Self-enumeration was the most common mode of collection: in MO and L/L areas, an adult member of the household was asked to complete the census questionnaire for all members of the household. If the questionnaire was completed on paper, the respondent returned it by mail in a pre-addressed envelope to the Data Operations Centre (DOC).

One in four private dwellings received the long form (2A-L), while three in four dwellings received a short form (2A). Residents who wished to be enumerated separately could fill out their own form by requesting a short form for individuals (3A).

Canvasser and reserve enumeration

This method was used in remote and northern areas of the country, and on Indian reserves where other collection methodologies were deemed to be less efficient. Canvasser methodology was also used in what is called early enumeration (EE) areas where enumeration takes place before Census Day (May 10, 2016), the reference date used is the date on which the household is being enumerated.

Early enumeration took place from February to April 2016 for populations who tend to migrate out of their communities starting in May.

In 2016, approximately 1% of dwellings were enumerated using the canvasser enumeration method. In these cases, an enumerator visited the dwelling and completed a questionnaire (2A-R) for the household by interview.

Enumeration of collective dwellings

A collective dwelling refers to a dwelling of a communal, institutional or commercial nature. There are different types of collective dwellings, depending on the types of services provided. The type of collective dwelling determined how it was enumerated.

Collective dwelling enumeration was conducted from May 2 to June 3, 2016. The enumeration of all types of collective dwellings followed the same general procedures regardless of whether they were in an MO or L/L area, with field staff ensuring that each usual resident was enumerated.

For 2016, there were four main types of enumeration procedures covering the different types of collective dwellings.

(1) Administrative records and/or an interview with the collective dwelling census contact or an administrator, for institutional collective dwellings

This approach was used in hospitals, shelters, nursing homes and/or residences for senior citizens and correctional or custodial facilities. Enumerators obtained mandatory data by accessing administrative records. If administrative data were not available, then for nursing homes and/or residences for senior citizens, enumerators may have interviewed residents using Form 2A and Form 3A for live-in employees or employees' family members.

(2) Canvasser for lodging and rooming houses

For lodging and rooming houses, enumerators completed the questionnaire (Form 3A) with each resident with a face-to-face interview.

(3) Self-enumeration for religious establishments and Hutterite colonies

For religious establishments such as convents, monasteries or seminaries and Hutterite colonies, enumerators left the questionnaire (Form 2A) with a collective dwelling census contact for distribution to residents, then arranged pickup of the completed questionnaires.

(4) Usual resident (UR) count only

For some collective types, the census collected only a count of usual residents (UR). This was used in establishments with temporary accommodation services such as hotels, campgrounds, YMCA or hostels and other establishments such as school residences, military bases, work camps or vessels.

Enumeration of residents outside Canada

To enumerate Canadian residents posted outside Canada, Form 2C was used (see [Chapter 5](#)).

Census Help Line

The Census Help Line (CHL), a free, nationwide, multilingual service, was available to all respondents. The toll-free number was printed on the census questionnaire and guide, and advertised in all census communications materials.

Data collection verification and follow-up activities for the 2016 Census

Apartment occupancy verification (AOV) – The purpose of the AOV is to be able to verify the occupancy status of all units in an apartment building through one management contact. Pre-contact is done to collect contact information for large apartment buildings. This verification can only be done for units in buildings with contact information. It is an important exercise as it helps to produce a more accurate status of occupancy for these types of dwellings and to reduce the workload of the census non-response follow-up (NRFU) activity. AOV was conducted from Statistics Canada regional offices from March 1 to 31, 2016.

Dwelling occupancy verification (DOV) – The status of dwelling occupancy was verified immediately prior to non-response follow-up (NRFU) in MO areas. Dwelling occupancy verification (DOV) was conducted from May 20 to 27, 2016, to identify as many unoccupied dwellings as possible, close to Census Day, in order to remove these dwellings from the NRFU workload. The accuracy of the unoccupied status is higher if identified closer to Census Day, May 10, 2016. This operation is independent from AOV described above.

Non-response follow-up (NRFU) – The purpose of NRFU was to obtain a completed questionnaire from all households that did not return a questionnaire. This was the final collection activity in MO and L/L areas. Follow-up was first done by telephone when numbers were available. If a completed questionnaire could not be obtained by telephone, personal visits were conducted until a completed questionnaire was obtained. In L/L areas, follow-up was generally conducted from May 20 to July 15, 2016, and in the MO areas from June 1 to July 31, 2016. NRFU was not conducted in areas of early enumeration and canvasser areas, as completeness was verified during the enumeration process.

Failed edit follow-up (FEFU) was a process through which follow-ups were made on selected dwellings to resolve missing and incomplete responses on the census short-form questions. FEFU was conducted from Statistics Canada regional offices for those questionnaires received at the Data Operations Centre (DOC) that were identified as having failed edits because they were missing information. It was primarily focussed on cases with potential household composition issues. FEFU was conducted from May 10 to August 14, 2016.

Chapter 8 – Processing

Introduction

The step after collection, known as the processing phase, began May 2, 2016, with the process of editing and coding responses for approximately 15 million private and collective dwellings.

Receipt and registration

For the 2016 Census, electronic responses from online questionnaires were received from the Collection Management Portal (CMP) and registered in the Census Processing System (CPS) hourly before entering the edit and coding workflow. CPS also registered interviewer responses received through the Census Help Line (CHL), non-response follow-up (NRFU) and failed edit follow-up (FEFU) on a regular basis during collection and/or follow-up.

Paper questionnaires that were returned by mail were registered in Canada Post sorting plants by scanning the bar code on the front of the questionnaire before delivery to the Data Operations Centre (DOC). To confirm receipt by Statistics Canada, the questionnaires were removed from the envelopes and scanned again at the DOC. Whenever Canada Post was unable to read the barcodes (for instance, when forms were inserted into envelopes backwards), the questionnaires were removed from the envelopes and the barcode scanned when the envelope was delivered to Statistics Canada.

Registrations of all questionnaires from Canada Post were transmitted to the CMP daily. Enumerators were notified (via the CMP) of which questionnaires had been received so that they could stop contact for these respondents during NRFU procedures.

Paper questionnaires that were completed by enumerators during NRFU were shipped by their supervisors (crew leaders) directly to the DOC where they were registered. All such questionnaires were then data captured similar to other paper responses.

Imaging and data capture

Once paper questionnaires were registered, the next step was document preparation and scanning for data capture of responses.

Steps

1. **Document preparation** – Mailed-back questionnaires were removed from envelopes. In order to ensure that questionnaires were ready to be scanned, operators removed foreign objects such as clips and staples from the documents. Forms were also cut into single sheets using guillotines (large paper cutters).
2. **Scanning** – Scanning, using high-speed scanners, created digital images from the paper questionnaires.
3. **Automated image quality assurance** – An automated system verified the quality of the scanning for capture purposes. Images failing this process were flagged for rescanning.
4. **Automated data capture** – Optical mark recognition (OMR) and intelligent character recognition (ICR) were used to extract respondent data. When the system could not recognize the handwriting (known as the write-ins), keying was done by an operator from the scanner images. Paper forms that could not be scanned (e.g., too damaged) or were filled out with a pen or pencil that could not be read by the automated capture systems, were sent for transcription (i.e., the data were transcribed to a new form).
5. **Check-out** – This quality assurance process ensured that the questionnaire images and captured data were of sufficient quality and that the paper questionnaires were no longer required.

Edits

As the paper questionnaires were captured and the online questionnaires received, an interactive process of manual and automated edits was performed to ensure that problems and inconsistencies were identified and resolved.

1. **Blank and minimum content** – This automated edit identified questionnaires with no information or insufficient information to continue processing. These cases were returned to the field for non-response follow-up (NRFU) by census enumerators.
2. **Multiple responses** – A household may have multiple forms (e.g., large households require more than one paper form to complete the census). This automated edit identified households with one or more missing questionnaires. These cases were held in a queue until all questionnaires were received.
3. **Coverage edits** – These edits were conducted for both private and collective dwellings and ensured that the reported number of members of a household was consistent with the responses provided, including the number of names listed. Errors were resolved by an automated process or through interactive verification by DOC staff by manually examining the captured data and scanned images (where available) to help determine the appropriate solution.
4. **Failed edit follow-up (FEFU)** – Short-form questions that needed further coverage or content clarification were transmitted to the Statistics Canada regional offices for FEFU collection and transmitted back to the DOC for the CPS for subsequent processing.

Coding

Written responses were converted to numerical codes before they could be tabulated for release purposes. For the 2016 Census, all written responses on the questionnaires underwent automated and interactive coding to assign each one a numerical code using reference files, code sets and standard classifications. Reference files were built using actual responses from past censuses for the automated match process. Subject-matter experts are responsible for developing, testing and maintaining the reference files that are used for automated and manual coding. Subject-matter coded all the write-ins that were referred by the first- and/or second-level coders and certified the codes prior to delivery to edit and imputation.

Edit and imputation

The data collected in any survey or census contains omissions or inconsistencies. These errors can be the result of respondents missing a question, or they can be due to errors generated during processing.

After the initial editing and coding operations were completed, the data were processed through the final edit and imputation activity. The final editing process detected errors and the imputation process corrected them.

Chapter 9 – Sampling and weighting for the long form

For the 2016 Census, Canadian households are counted using two main types of questionnaires: the short-form questionnaire and the long-form questionnaire. The long-form questionnaire includes the same questions as the short form, as well as a series of questions aimed at providing a comprehensive portrait of the Canadian population and Canadian households. The long-form questionnaire is sent to a sample of the population.

The estimates produced from the responses to questions found on both questionnaires are obtained from a **census of population**. As such, all respondent households for both types of questionnaires contribute to a given number. That is the case for the population count for a specific age group, for example.

The estimates produced from the responses to at least one question from the long form are obtained from a **sample survey**. In this case, only the respondent households from the long-form sample contribute to the estimate, such as the unemployment rate estimate or the estimate of the population by highest level of education.

Selecting the sample for the census long-form questionnaire

The long-form questionnaire sample is selected from small geographic areas that, together, cover the country, called collection units (CUs). The CUs determine the strata for the sample plan. There are five types of CUs: list/leave, mail-out, collective dwellings, Indian reserves and canvasser enumeration. For the two last types of CUs, enumerators conduct personal interviews. In each CU (or stratum), a list of dwellings is drawn up and a systematic sample of private dwellings is chosen, with a sampling fraction of one in four. Collective dwellings are excluded from this draw. There are exceptions with respect to the sampling fraction: all private dwellings in CUs where enumerators collect data are selected for the long-form sample. Households in private dwellings selected for the sample are asked to complete the census long form. Other households—i.e., those in the private dwellings that are not part of the long-form sample, as well as those in collective dwellings, which are excluded from sampling—are asked to fill out the short form.

The sample for the long-form questionnaire is divided equally among geographic areas to ensure estimates are reliable for all regions across the country and to give the same relative importance to all geographic units of a given size. The sampling fraction was increased in 2016, compared with one in five for the previous census long-form questionnaire in 2006. In 2011, the response rate for the National Household Survey (NHS), a voluntary survey, was lower than for the 2006 Census long-form questionnaire. For the 2016 Census long-form questionnaire, a sample of one in four households was selected to reduce the risk of lower participation than in the past.

Weighting the sample for the census long-form questionnaire

The final responses to the long-form questionnaire are weighted so that they represent the Canadian population living in private dwellings. Weighting is the process of grouping the sample weight calculation and various adjustments to that weight. These include a weighting adjustment for the coverage of occupied dwellings based on the results of the Dwelling Classification Survey (DCS), an adjustment to correct the total non-response of sampled households, and a calibration of the weights of respondent households to totals derived from the census.

First, each household is given a sample weight equivalent to the inverse of its probability of selection in the sample. In CUs where enumerators conducted personal interviews, this weight is 1. In other CUs, this weight is generally 4. It is higher than 4 and no greater than 7 in list/leave CUs with a number of dwellings that is not a multiple of four because of how the systematic sample in this type of CU is drawn. In these CUs, the systematic sample is not random and the sampled dwellings are those listed 4th, 8th, 12th, etc. For example, if one such CU has seven dwellings, the sampled dwelling, i.e., the fourth one listed, will have a sample weight of 7 to represent all dwellings in its CU.

In the sample selected for weighting, several types of responses to the long form can be differentiated. First, there are households that answered at least one question from the long form that was not in the short form. These households are defined as “respondent households” for the long form. Then, there is a fraction of households that

answered only questions found on both the questionnaires or, similarly, only questions on the short form. Finally, there are certain households that did not respond to any questions. The last two types of households are referred to as “non-respondent households” for the long form.

In CUs where enumerators conducted personal interviews, i.e., Indian reserve CUs and canvasser enumeration CUs, non-response to the long-form questionnaire is accounted for by imputation. Data for households that did not respond to any questions are imputed using data from a respondent household in the same type of CU. Other non-responses are imputed for partial non-response. All private households in these CUs that are not part of incompletely enumerated Indian reserves and establishments keep their sample weight of 1 for estimation purposes. Other private households and collective households are attributed a final weight of nil and thus do not contribute to the estimates.

In the other types of CUs, several adjustments are made to the weight, and a different imputation method is used. The following describes the processing in these CUs. Only respondent households for the long-form questionnaire are assigned a non-null weight at the end of the weighting stages, meaning that they are the only ones to contribute to the long-form questionnaire estimates. Partial non-response for these households is compensated for by imputation.

Non-respondent households for the long-form questionnaire are nevertheless taken into consideration in the census figures. In fact, for all enumerated households that did not answer any questions, all responses are imputed for questions found on both questionnaires based on data from a household that answered at least one such question. The remaining non-responses to these questions for all enumerated households are imputed for partial non-response.

Before proceeding with the imputation for the census total non-response, the census undercoverage of occupied dwellings is estimated using the DCS, and this undercoverage is corrected by changing the occupancy status of certain dwellings. The incorrect classification of dwellings on Census Day is in fact one source of coverage error. This error can occur when an occupied dwelling is classified as unoccupied or when an unoccupied dwelling is classified as occupied. The purpose of the DCS is to estimate the number of these classification errors. To this end, a sample of private dwellings for which no census questionnaire was returned are contacted, and information is gathered on their occupancy status on Census Day and, if the dwelling was occupied, on the number of usual residents.

The weighting steps that follow the assignment of the sample weight are carried out after imputing for total non-response and for partial non-response to questions found on both questionnaires. All these weight adjustments are done by calibration. Calibration consists of applying the smallest adjustment possible to the weight so that the weighted estimates coincide with known counts. These known counts are referred to as control counts.

At each stage, the country is divided into geographic areas, and each area is calibrated independently. Four types of geographic units can be used, depending on the weighting stage: the dissemination area (DA), the aggregate dissemination area (ADA), the census subdivision (CSD) and the super aggregate dissemination area¹ (SADA). DAs are small areas, consisting of one or more neighbouring dissemination blocks and including 400 to 700 people. ADAs are groups of adjoining DAs, most often including 5,000 to 15,000 people. ADAs respect provincial and territorial borders, as well as the boundaries of census divisions (CDs), census metropolitan areas (CMAs) and census agglomerations (CAs) subdivided into census tracts (CTs) in effect for the 2016 Census. CSDs are also groups of DAs that respect the boundaries of CDs. They correspond to municipalities or areas treated as municipal equivalents for statistical purposes. SADAs are groups of adjoining ADAs, most often including 50,000 to 150,000 people. SADAs respect provincial and territorial borders and, most of the time, the boundaries of CDs.

The unit of measure for control counts can be the household or the person. Some control counts are derived from responses to questions found on both questionnaires. They are related to geography, age, sex, marital or

1. This geography is not part of the dissemination geographies. It was created especially for weighting.

common-law status, dwelling type, size of household, family structure and knowledge of official languages. Other control counts are derived from administrative data matched to census records. These are counts derived from individual income tax data, immigration data and data from the Indian Register. However, for a given region, several control counts are eliminated based on certain criteria to maximize the general quality of estimates.

The first sample weight adjustment makes the coverage of the selected sample correspond to that of the census. In fact, the imputation for total non-response and for census undercoverage based on the DCS does not allow the type of questionnaire to be taken into account. This means that the sample coverage after imputation can differ from the census coverage. To make them correspond, the sample weight is calibrated for all households targeted for the long-form questionnaire in the sample, whether or not they responded. This adjustment is made independently by SADA. All control counts are derived at that level, except some counts of households and individuals in the ADAs that make up the SADAs. The weight of households that are not targeted for the long-form questionnaire is set to 0. After the adjustment, the control counts correspond to the weighted counts for the sample.

The weight (adjusted for coverage) of respondent households is then adjusted for non-response using a logistic regression model that predicts the likelihood of response. This is done at the SADA level using a calibration of weights of respondent households based on the model. The control counts are the same as for the first adjustment, and the model's prediction variables are the variables that correspond to these counts. The weight of non-respondent households is set to 0. As a result, the control counts correspond to the weighted counts of respondent households.

The final adjustment consists of calibrating the weight (adjusted for non-response) of respondent households to more control counts. This ensures a certain consistency with census counts and attempts to reduce the variability of long-form questionnaire estimates. Calibration is again done independently by SADA. For this adjustment, a greater number of counts is chosen at the ADA level, and household and person counts are chosen by cross-tabulating ADAs and CSDs.

The weighted estimates from the long-form questionnaire may differ from census counts for characteristics found in both. In particular, this is the case when looking at a geography with boundaries that do not correspond to ADAs and SADAs. Furthermore, the smaller the geographic area, the greater the likelihood that estimates from the long-form questionnaire will differ from the census counts. When there are differences, the 2016 Census figures should be considered of higher quality and users should prioritize them, as they are not affected by the sampling variance or the slightly higher non-response error of the long-form questionnaire. Estimates from the long-form questionnaire for characteristics found in both forms should be used as contextual information when analyzing data specific to this questionnaire.

A detailed technical guide to sampling and weighting for the long-form questionnaire will be available in 2018. It will give further details on the weighting and estimation process.

Chapter 10 – Data quality assessment

Introduction

Data quality assessment provides an evaluation of the overall quality of census data. The results of this assessment are used to inform users of the reliability of the data, to make improvements for the next census, to adjust census data for non-response and, for two coverage studies (reverse record check and the Census Overcoverage Study), to produce official population estimates. Quality assessment activities take place throughout the census process, beginning prior to data collection and ending after dissemination.

Sources of error

However well a census is designed, the data collected will inevitably contain errors. Errors can occur at virtually every stage of the census process, from material preparation to creation of the list of dwellings, data collection and processing. Census data users should be aware of the types of errors that can occur, so they can assess the usefulness of the data for their own purposes.

Main types of errors:

Coverage errors occur when dwellings and/or persons are missed, incorrectly enumerated or counted more than once.

Non-response errors occur when some or all information about individuals, households or dwellings is not provided.

Response errors occur when a question is misunderstood or a characteristic is misreported by the respondent, the census enumerator or the Census Help Line operator.

Processing errors can occur at any stage of processing. Processing errors include errors that can be made at data capture during coding operations, when written responses are converted into numerical codes, and during imputation, when valid (but not necessarily accurate) values are inserted into a record to replace missing or invalid data.

Sampling errors apply only when answers to questions are obtained from a sample. This type of error applies only to the 2016 Census long-form questionnaire.

Measuring data quality

Many data quality studies have been conducted for recent censuses to allow data users to assess the impact of errors and improve their own understanding of how errors occur. For the 2016 Census, special studies examine errors in coverage and data quality, i.e., non-response, response and processing.

Three studies are conducted to measure coverage errors:

1. **Dwelling Classification Survey** – One of the sources of coverage error in the census is the misclassification of dwellings on Census Day. This error can occur when an occupied dwelling is classified as unoccupied, or when an unoccupied dwelling is classified as occupied. The purpose of the Dwelling Classification Survey is to study these types of classification errors and adjust counts, if necessary. A sample of dwellings for which no census questionnaire was returned is contacted, information is collected on the occupancy status and, if occupied, on the number of usual residents.

This information is used to adjust the census data for dwellings, households and persons. This is done by correcting the classification errors and adjusting household size distribution through imputation for dwellings that did not return the questionnaire. It is done in time for the initial population count release.

2. **Reverse Record Check** – This study provides estimates of persons missed by the census (after accounting for the adjustments described in the Dwelling Classification Survey above). Estimates are developed for each province and territory and for various population subgroups (e.g., age-sex groups and marital status).

For the provinces, this study comprises two steps:

- Step 1: Selecting a sample of persons who should have been enumerated in the census, using sources such as the previous census, birth registrations, immigration and non-permanent residents' records, and the sample of persons missed in the Reverse Record Check from the previous census.
- Step 2: Linking persons selected in Step 1 to the Census Response Database (CRD) to determine whether these persons were enumerated. The survey is then used to trace and interview persons who could not be linked with certainty to the CRD in order to collect additional information. Persons who have died or who emigrated prior to Census Day are identified using administrative records, such as the death register, or during tracing or the interviews.

For the territories, Step 1 consists in linking the persons on health insurance records to the Census Response Database to identify persons who were enumerated in the census. The Reverse Record Check sample is then selected among the unmatched persons.

The results of the Reverse Record Check are the most important source of information about persons missed in the census. However, unlike the Dwelling Classification Survey, the estimates are not used to adjust census data before the initial population count release.

3. **Census Overcoverage Study** – In the 2011 and 2016 censuses, double-counting of persons is determined by searching for linked records that have a high degree of matching on sex, date of birth and name. Linked records are sampled and checked manually, and results are used to estimate the census overcoverage (or the number of duplicate persons).

When combined with the results of the Reverse Record Check, the results of the Census Overcoverage Study provide estimates of net coverage error in census data. This net error is used to calculate the official population estimates.

Certification

Certification consists of several activities to rigorously assess the quality of census data at specific levels of geography in order to ensure that the quality standards for public release are met. This evaluation includes the certification of population and dwelling counts, and variables related to dwelling and population characteristics.

During certification, response rates, invalid responses, edit failure rates, and a comparison of data before and after imputation are among the data quality measures used. Tabulations for the 2016 Census are produced and compared with corresponding data from past censuses, other surveys and administrative sources. Detailed cross-tabulations are also checked for consistency and accuracy.

Depending on the certification results, census data can be released in one of three ways:

- First, the data may be released unconditionally, meaning that the data are of suitable quality.
- Second, the data may be released conditionally or with restrictions. In this case, the data will be released with a special note alerting users to possible limitations, or the data may be specially processed, for example, by combining reporting categories to address quality or confidentiality concerns.
- Finally, the data may be suppressed for quality reasons.

For more information on the quality indicators and certification results, see the [reference guides](http://www12.statcan.gc.ca/census-recensement/2016/ref/index-eng.cfm) (<http://www12.statcan.gc.ca/census-recensement/2016/ref/index-eng.cfm>) for the various domains of interest.

Response rate for the 2016 Census of Population

One of the key data quality measures used for the Census of Population is the response rate. Table 10.1 shows the response rates for the 2016 Census of Population both nationally and for each province and territory. The rates are provided for all occupied private dwellings for which a short form or long form was to be received and for the subset of occupied private dwellings for which a long form was to be received. For the long form, the unweighted response rate and the weighted response rate are provided.

The rates in Table 10.1 were calculated following data processing and data quality assessment. Response rates are calculated as follows: the number of private dwellings for which a questionnaire was filled out divided by the number of private dwellings classified as occupied according to the census database. The final classification of dwelling occupancy status is based on the data analysis collected by field staff, the data provided by respondents and the results of a quality study on the occupancy status of a sample of dwellings. The rates in Table 10.1 differ from the collection response rates previously disseminated because they take into account data processing and verification of the dwelling occupancy status and thus are considered final. With respect to weighted response rates, they are based on the long form's final sampling weights. The weighted response rates are therefore calculated as follows: the number of sampled weighted private dwellings for which a questionnaire was filled out divided by the number of weighted sampled private dwellings classified as occupied.

Table 10.1
2016 Census of Population: Response rates

Province/territory	Short and long form response rates	Unweighted response rates from the long form only	Weighted response rates from the long form only
	(%)		
Canada	97.4	96.7	96.9
Newfoundland and Labrador	97.4	96.6	96.8
Prince Edward Island	97.5	96.9	97.0
Nova Scotia	97.6	97.1	97.2
New Brunswick	97.6	97.1	97.2
Quebec	97.6	97.2	97.3
Ontario	97.6	97.0	97.2
Manitoba	97.4	96.3	96.9
Saskatchewan	96.7	96.2	96.3
Alberta	97.0	96.3	96.4
British Columbia	96.5	95.7	96.0
Yukon	95.8	93.5	95.2
Northwest Territories	93.9	92.8	93.1
Nunavut	92.7	92.7	92.7

Source: Statistics Canada, Census of Population, 2016.

Quality of long-form questionnaire estimates

Estimates produced from the long-form census questionnaire are obtained by means of a **sample survey**. Such surveys have two types of errors: sampling error and non-sampling error. The former occurs when a characteristic is estimated by measuring only part of the population instead of the entire population. The latter includes all errors that are not related to sampling. The latter type of error also occurs in **census counts**, i.e., counts produced from questions found in both the short-form questionnaire and the long-form questionnaire.

Sampling error

Sampling error is the difference that would be observed between an estimate from the long-form questionnaire and the actual value for the population if there were no non-sampling errors, i.e., if there were no coverage error, response error, processing error or non-response. It is inevitable when conducting a sample survey such as that conducted using the long-form census questionnaire.

Several factors influence sampling error. The sampling error is smaller when the sampling fraction is larger and when the sample size is larger. Ultimately, if the sampling fraction is 100%, as for the short-form census questionnaire, then the sampling error will be nil. It will also be small if the variability of the variable of interest in the population is low. This error also depends on the effectiveness of the sample plan. For example, it will be smaller if the populations in the strata of the sample plan are fairly homogenous, or, in the case of a characteristic measured at the level of the person, if the individuals in the households are fairly heterogenous. Finally, sampling error depends on the estimation methods used, such as the weighting method, as some are more effective than others. For example, when the weight of a survey is adjusted so that a weighted total is equal to the census total, the sampling error for that weighted total is nil. However, it must be noted that it is impossible to adopt a weighting method that would eliminate sampling error from all possible estimates drawn from the long-form questionnaire.

Sampling error cannot be measured directly. Indeed, to do so, the actual value of the variable of interest in the population would need to be known to subtract it from the estimate drawn from the long-form questionnaire, an estimate that should not include any non-sampling errors. However, its extent can be estimated using variability measures such as standard error and the coefficient of variation.

[Appendix 1.9](#) presents a measurement of sampling error (standard error).

Non-sampling error

Apart from sampling, several factors can lead to errors in survey results. These non-sampling errors can be of several types. Dwellings or individuals may have been incorrectly enumerated—this is coverage error. Respondents may not understand the questions and answer them incorrectly—this is response error. Responses may be entered incorrectly during data capture, or the coding of the responses may be incorrect. This is processing error.

Non-response error is also a non-sampling error. A distinction is made between partial non-response (no response to one or some questions) and total non-response (no response to the survey because the household could not be reached or refused to participate).

Non-sampling error is likely to bias estimates. Attempts were made to minimize it during each stage of collection and processing to reduce its impact. For example, as described in [Chapter 9](#), an attempt was made to correct non-response error and coverage error by means of imputation or by adjusting weights. However, a residual error remained following this process, except when the imputation of the missing value proved to be exact.

Moreover, non-sampling error is not as easy to measure as sampling error. Nonetheless, measures of variability produced from the long-form questionnaire estimate both sampling variability and variability caused by total non-response error, under the assumption that the model used in weighting to correct this non-response is accurate. In fact, total non-response variability is measured because it may be significant in a survey with a large sampling fraction.

Comparability of estimates from the 2016 Census long-form questionnaire with estimates from the 2011 National Household Survey

Users must be careful when comparing estimates from two surveys, as they can differ significantly in methodology, quality and target population.

The estimates from the 2016 Census long-form questionnaire were derived from a mandatory survey that had a high response rate, while the estimates from the 2011 National Household Survey (NHS) were derived from a voluntary survey. The response rate for the 2016 Census long-form questionnaire was 96.9%, while the 2011 NHS had a response rate of 68.6%. The definition of the target population of the 2016 Census long-form questionnaire and that of the 2011 NHS were exactly the same.

Non-response bias occurs when a survey's non-respondents are different from its respondents. The higher a survey's non-response, the greater the risk of non-response bias. The quality of estimates can thus be affected if such a bias is present. The risk of non-response bias was taken into account for the NHS. In fact, Statistics Canada conducted several studies on the 2011 NHS, as well as various simulations, before and after collection, to assess the potential risk of bias and its extent. A number of measures were taken to mitigate its effects. Non-response error may be greater for the estimates from the 2011 NHS than for the estimates derived from the 2016 Census long-form questionnaire, particularly for smaller domains of interest.

In addition, the response rate for the 2011 NHS varies significantly from one community to another, particularly small ones. The quality of the estimates and the risk of bias can thus vary significantly between different communities. For the 2016 Census, the response rates to the long-form questionnaire vary less between communities. There is therefore less variation in the quality of the estimates, and the risk of bias is very negligible. The estimates from the 2011 NHS can contain inaccuracies because of a lower response rate than the 2016 Census. Comparisons of estimates from the 2011 NHS and from the 2016 Census long-form questionnaire for a given geographic area must take into account the differences in observed response rates.

Unanswered questions are identified in returned questionnaires. Imputation replaces missing, invalid or inconsistent elements with plausible values. When carried out properly, imputation can improve data quality by replacing non-responses with plausible responses similar to those that the respondents would have given if they had answered the questions. It also has the benefit of producing a full dataset. Imputation for partial non-response (i.e., the unanswered questions in returned questionnaires) was greater for the 2011 NHS than for the 2016 Census long-form questionnaire. These differences are greater for questions in the second half of the NHS questionnaire (about work, the workplace, the mode of transportation, languages of work and housing). The 2011 and 2016 reference guides present imputation rates for each question at the provincial, territorial and national levels. Comparisons of estimates from the 2011 NHS and from the 2016 Census long-form questionnaire must take into account the differences in imputation rates.

Table 10.2 presents national-level imputation rates for variables from the 2011 NHS and the 2016 Census long-form questionnaire. The imputation rates for questions 2 to 9 were calculated differently in 2011 and 2016. In 2011, the imputation rate excluded imputation for household non-response, while this was included in 2016. For the other questions, the calculation method for the imputation rate can differ slightly between 2011 and 2016 for some variables.

Table 10.2

Imputation rates by question, for the 2011 Census, the 2011 NHS and the 2016 Census, Canada

Question	2011 Census and 2011 NHS	2016 Census
	(%)	
Q. 2 Sex	1.0	2.8
Q. 3 Date of birth	1.4	3.1
Q. 4 Marital status	2.0	4.3
Q. 5 Common-law status	5.0	5.1
Q. 6 Relationship to Person 1	2.4	3.2
Q. 7 Knowledge of languages	1.6	4.0
Q. 8 Language spoken most often	1.9	3.9
Q. 9 Mother tongue	2.3	4.3
Q. 10 92-year consent
Q. 11 Activities of daily living
Q. 12 Place of birth	2.0	1.0
Q. 13 Citizenship	2.3	1.3
Q. 14 Landed immigrant status	1.3	0.7
Q. 15 Year of immigration	12.5	9.4
Q. 16 Other language(s) spoken
Q. 17 Ethnic origin	5.8	4.5
Q. 18 Aboriginal group	3.7	1.1
Q. 19 Population group	3.9	2.0
Q. 20 Registered or Treaty Indian status	4.7	1.4
Q. 21 Membership in a First Nation or Indian band	3.8	1.8
Q. 22 Mobility one year ago	4.8	1.8
Q. 23 Mobility five years ago	6.9	2.4
Q. 24a Place of birth of father	6.0	1.6
Q. 24b Place of birth of mother	5.7	1.8
Q. 25 High school diploma or equivalent	4.6	1.2
Q. 26a Registered apprenticeship or other trades certificate or diploma	5.5	1.8
Q. 26b College, CEGEP or other non-university certificate or diploma	5.5	1.8
Q. 26c University certificate, diploma or degree	4.7	1.4
Q. 27 Major field of study	14.2	4.4
Q. 28 Location of study	12.1	3.1
Q. 29 Attendance at school	6.1	4.3
Q. 30 Hours worked	6.7	1.6
Q. 31 On lay-off or absent	10.5	4.5
Q. 32 Start of a new job	8.0	4.2
Q. 33 Job search	7.8	3.6
Q. 34 Reason for unavailability to work	10.3	3.1
Q. 35 Date last worked	8.7	6.2
Q. 36 and 37 Industry	13.6	6.2
Q. 38 and 39 Occupation	13.6	5.3
Q. 40 Class of worker	12.2	3.7
Q. 41 Incorporation status	8.1	5.1
Q. 42 Place of work status	11.3	3.7

Table 10.2
Imputation rates by question, for the 2011 Census, the 2011 NHS and the 2016 Census, Canada

Question	2011 Census and 2011 NHS	2016 Census
	($\%$)	
Q. 42 Workplace location	13.0	5.4
Q. 43a Mode of transportation	12.1	4.3
Q. 43b Vehicle occupancy	13.7	3.8
Q. 44a Time leaving for work	15.5	5.0
Q. 44b Commuting duration	14.8	5.3
Q. 45 Languages of work	12.9	3.1
Q. 46 Weeks worked last year	15.1	2.9
Q. 47 Full-time or part-time work	14.6	5.4
Q. 48 Amount paid for child care	3.6	31.1
Q. 49 Amount paid in support	12.6	4.3
F1 Household maintainer	11.8	2.0
F2 Housing tenure	10.7	1.8
F3 Condominium status	9.4	1.3
F4a Rooms	12.8	3.6
F4b Bedrooms	11.4	1.8
F5 Period of construction	13.5	2.9
F6 Condition of dwelling	10.7	1.7
F8a Payment—Electricity	19.5	6.8
F8b Payment—Fuel	19.0	7.0
F8c Payment—Water and other services	19.5	7.0
F9a Rent	14.3	5.4
F9b Subsidized housing	13.9	5.1
F10a Mortgage payments	18.0	5.1
F10b Property taxes included in mortgage payments	17.2	4.1
F10c Property taxes	20.8	7.4
F10d Value of dwelling	21.2	7.1
F10e Condominium fees	22.8	14.4

... not applicable

Sources: Statistics Canada, Censuses of population, 2016 and 2011, and 2011 National Household Survey.

The quality assessment of the NHS estimates deemed their quality to be acceptable overall for Canada, the provinces and territories, and census metropolitan areas. The NHS estimates were comparable with those from other data sources at the same geographic levels. At a more detailed geographic level, estimates from the NHS could not be compared with those from other data sources.

Users are encouraged to use the main quality indicator provided, the global non-response rate (GNR), to judge the quality of estimates from the 2011 NHS and the 2016 Census when assessing the reliability of comparisons. The GNR is an important measure of the quality of estimates from the NHS and the long-form questionnaire. It combines household non-response and partial non-response. For the NHS and the long-form questionnaire in particular, the GNR is weighted to account for the sampling. The GNR is a potential indicator of non-response bias. For each region for which data are published, the GNR is available for both the 2011 NHS and the 2016 Census long-form questionnaire. Nationally, the GNR was 26.1% for the 2011 NHS and 5.1% for the 2016 Census long-form questionnaire (see [Chapter 11](#) for more information). The GNR for the 2011 NHS and the GNR for the 2016

long-form questionnaire differ more for smaller geographic areas. In cases of greater discrepancy between the GNR for the 2011 NHS and the 2016 long-form questionnaire, users should take care in making comparisons. Users are also encouraged to read any quality notes that may be included with dissemination products.

Comparability of estimates from the 2016 Census long-form questionnaire with estimates from the 2006 Census long-form questionnaire

Estimates from the 2006 Census long-form questionnaire were derived from a mandatory survey. The response rate for the 2006 Census long-form questionnaire was 93.8%. The risk of non-response error was very low, and this risk is similar for estimates from the 2016 Census and 2006 Census long-form questionnaires.

The definition of the target population for the 2016 Census long-form questionnaire differs from the definition for the 2006 Census long-form questionnaire. The 2016 Census long form questionnaire targeted the total population usually living in Canada in private dwellings, in the provinces and territories. This target population includes persons who live on Indian reserves and in other Indian settlements; permanent residents; and non-permanent residents such as refugee claimants, holders of work or study permits, and members of their families living with them.

Foreign residents are not enumerated in the 2016 Census long-form questionnaire. These include representatives of a foreign government assigned to an embassy, high commission or other diplomatic mission in Canada; members of the armed forces of another country stationed in Canada; and residents of another country who are visiting Canada temporarily.

The 2016 Census long-form questionnaire also excludes persons living in institutional collective dwellings such as hospitals, nursing homes and penitentiaries; Canadian citizens living in other countries; and full-time members of the Canadian Armed Forces stationed abroad. Finally, the 2016 Census long-form questionnaire excludes persons living in non-institutional collective dwellings such as work camps, hotels and motels, and student residences.

The 2006 Census long-form questionnaire did not target exactly the same population. Compared with the 2016 Census long-form questionnaire, the 2006 questionnaire included persons living in non-institutional collective dwellings such as work camps, hotels and motels, and student residences. It also targeted foreign residents such as representatives of a foreign government assigned to an embassy, high commission or other diplomatic mission in Canada. These differences between the target populations of the 2016 and 2006 long-form questionnaires are minor and relate to only a very small percentage of the total population. Users must nonetheless take these differences into consideration when comparing estimates from 2016 and 2006.

Comparability of the variability of estimates from the 2016 Census long-form questionnaire with that of estimates from the 2011 NHS and the 2006 Census long-form questionnaire

As mentioned in the previous sections, estimates produced using data from a sample survey, such as those from the 2016 Census long-form questionnaire, include sampling error, i.e., an error stemming from the fact that only a sample of the population was observed. Sampling error is determined using variability measures such as standard error or the coefficient of variation (CV). In [Appendix 1.9](#), standard error is used to compare the variability of estimates from the 2016 Census long-form questionnaire with that of estimates from the 2011 NHS and the 2006 Census long-form questionnaire.

Moreover, the purpose of the 2016 and 2006 Census long-form questionnaires and of the 2011 NHS was to produce estimates for a series of questions for a variety of geographic areas, ranging from very large areas (such as provinces and census metropolitan areas) to very small areas (such as neighbourhoods and municipalities), and for various population groups, such as Aboriginal peoples and immigrants. These groups also vary in size, especially when cross-classified by geographic area. Such groupings are generally referred to as “domains of interest.” The purpose of this section and of [Appendix 1.9](#) is to compare the variability of estimates from 2016, 2011 and 2006, not to compare the estimates. However, sampling variability should be taken into account when

comparing estimates from these surveys, particularly for small “domains of interest,” as the observed differences can be the result of sampling variability rather than an actual difference in the population.

Description of standard error

The “standard error” of an estimate is a numerical measurement of the random component of its error. Standard error can be interpreted as follows. If the sampling, collection and processing for the long-form questionnaire could be repeated many times and if an estimate for a given characteristic were calculated every time, the estimates produced in about 68% of cases would be within one standard error of the census value (i.e., the value that would have been obtained if a census had been conducted instead of a sample survey). Furthermore, the estimates produced in about 95% of cases would be within two standard errors of the census value. This means that, in general, the lower the standard error, the more accurate the estimate. The standard error is a key element in deriving other measures of variability, such as the CV; in constructing confidence intervals; and in making statistical inferences (e.g., determining whether an estimate is significantly different from a given value or another estimate). Estimates of standard error for the 2016 Census long-form questionnaire will be published in early 2018 in a supplement to the profile of aggregate dissemination areas (ADAs). They will include estimates of standard error for ADAs, census divisions, provinces and territories, and Canada.

Derivation of the coefficient of variation (CV)

The CV of an estimate is the ratio of the estimate of standard error and the estimate itself, expressed as a percentage. Like standard error, the lower the CV, the more accurate the estimate. The CV is an interesting measure of variability, as it does not depend on the estimate's unit of measure. This makes it possible to compare the accuracy of estimates that have different units of measure. However, as the CV is a ratio, it tends to have a very large value when the quantity in the denominator (i.e., the estimate of interest) is very small. Thus, care is needed when interpreting the CV of a very small proportion.

Distinction between standard error, response rate and global non-response rate

Standard error does not measure bias, such as non-response bias. It is important not to confuse standard error, the non-response rate and the global non-response rate. In fact, the non-response rate indicates the risk associated with household non-response error, and the global non-response rate indicates the risk associated with household non-response error and partial non-response. However, standard errors calculated based on the 2016 Census long-form questionnaire, the 2011 NHS and the 2006 Census long-form questionnaire include total household non-response variability, to a certain degree and in addition to sampling variability.

Conceptual and methodological differences between standard errors for the 2016 Census long-form questionnaire, the 2011 NHS and the 2006 Census long-form questionnaire

Several factors influence the values of standard errors and can explain the differences between cycles. First, the target population, sampling methods and estimation methods differ from one cycle to another. Furthermore, the variability measured is not exactly the same for all cycles: sampling variability is estimated in all cases, but household non-response variability is not measured in the same way in all cycles. In fact, in 2006, household non-response variability was not measured in Indian reserve collection units and canvasser enumeration collection units, while it was measured in 2011 and 2016.

Factors that contribute to reducing sampling variability are a larger sample, a larger sampling fraction, lower variability of the characteristic in the population being studied, a more effective sample plan and more effective estimation methods. The extent of non-response, the differences and similarities in the characteristics of respondents and non-respondents, and the estimation methods are the main factors that influence non-response variability.

The following subsections describe these factors and some conceptual and methodological differences between the sample for the 2016 Census long-form questionnaire, the 2011 NHS and the 2006 Census long-form questionnaire.

Target population

The target population for each of the three surveys is different. The 2016 Census long-form questionnaire targeted the Canadian population as of May 10, 2016. The 2011 NHS targeted the Canadian population as of May 10, 2011, while the 2006 Census long-form questionnaire targeted the Canadian population as of May 16, 2006. Furthermore, the target population in 2011 and 2016 included only persons living in private dwellings, while the target population in 2006 also included persons living in non-institutional collective dwellings (approximately 1% of the population).

Sample plan

The sample plan for the census long-form questionnaire consists of only one sampling phase. The plan for the 2011 NHS is more complex and includes two sampling phases. The sampling fractions also differ from one cycle to the next. In 2016, one in four dwellings was sampled. In 2011, an average of one in three dwellings was sampled at first. After several weeks of collection, i.e., as of July 14, 2011, the initial sample was reduced. Only the respondents were kept (approximately two-thirds of the initial sample), plus a follow-up subsample of approximately one in three dwellings drawn from among the remaining non-respondents. In 2006, the sampling fraction was one in five.

Generally, the standard error should be lower for larger sampling fractions and for larger sample sizes. It should also be lower for the simplest and most “efficient” sample plans, i.e., the one-phase sample plans in 2006 and 2016.

Extent of household non-response

Household non-response reduces the number of responses observed, which increases the variability of estimates. The final non-response rate for the 2016 Census long-form questionnaire was 3.1%. In contrast, the unweighted non-response rate for the 2011 NHS was 31.4%, and the non-response rate for the 2006 Census long-form questionnaire was 6.1%. The estimates from the 2011 NHS are therefore more affected by the extent of household non-response than those from the 2016 and 2006 long-form questionnaires.

Differences in characteristics between respondents and non-respondents

In general, the characteristics of respondents and non-respondents in the sample should be as similar as possible. In fact, if they were perfectly comparable (e.g., if responding were independent of the characteristics of interest), there would be no non-response bias. Furthermore, since there would be no non-response bias, the measurement of non-response variability would measure the total non-response error.

On the other hand, if the characteristics of respondents and non-respondents were very different, then the non-response bias would be significant. This could present a problem, as it is not considered in the measurement of variability. Ultimately, it would be possible to have a non-response variability estimate of nil, but a significant bias. The higher the non-response rate, the greater the risk of such a situation occurring.

The standard errors for the 2011 NHS could be pushed downward more than those from the 2006 and 2016 long-form questionnaires, because of the differences in the characteristics of respondents and non-respondents. In fact, given the higher household non-response rate for the 2011 NHS, the characteristics of its respondents tend to be more homogenous.

Estimation methods

One way to minimize the impact of non-response is to use estimation methods, including weighting methods, that make good use of the available information. As the extent of non-response was greater in 2011, more information was used to reduce the non-response error than in 2006. In fact, in 2006, only geographic information and household size were used to adjust for non-response, while all census variables and some administrative data were used in 2011. In 2016, more information continued to be used to correct non-response, despite the higher response rate than in 2011. The main goal of using this information is to reduce the non-response bias. However, doing so can result in increased variability of estimates, for example if the variability of weights adjusted for non-response is greater. The methods for correcting non-response in 2011 and 2016 should therefore increase the variability of the final estimates more than those of 2006.

Calibration was carried out in the last stage of weighting for each cycle to produce the estimates (counts, proportions, averages, etc.). The calibration process ensured that certain estimates of survey counts corresponded to known counts. The calibration was done using counts from the census long-form questionnaire or administrative data matched to census records. In 2006, counts were based on demographic and geographic variables. In 2011, counts were added for family and language variables. In 2016, counts were also added based on administrative data matched to census records (i.e., data on income, on immigration and from the Indian Register). The effect of calibration, apart from allowing for concordance with census counts, is to reduce the variability of estimates produced from variables related to control counts. It would thus be possible to see less variability for variables related to the calibration subjects used for a given cycle.

Moreover, although the number of calibration subjects has increased over time, the overall number of constraints used in the calibration has been reduced from one cycle to the next. This is because simulations used for developing estimation methods for 2011 and 2016 revealed that tighter calibration can lead to higher variance estimates for the variables that were least related to the calibration subjects. This could therefore explain in part the differences between the standard errors of the three cycles. [Appendix 1.9](#) presents the measures of sampling error, as standard errors, for estimates from the 2006 Census long-form questionnaire, the 2016 Census long-form questionnaire and the 2011 NHS.

Chapter 11 – Dissemination

Introduction

A primary goal of the Census of Population dissemination process is to ensure that census and geography products and services meet the primary needs of the majority of data users. Similar to previous Census of Population releases, Statistics Canada strives to provide more data free of charge to the public, while at the same time seeking ways of publishing census results in a timely and user-friendly manner. The first 2016 Census of Population results were made available to data users on February 8, 2017. All major releases of standard products were scheduled to take place in 2017, 10 months earlier than in 2011.

How census data are used

Governments, businesses, associations, community organizations and many others use census data extensively. The following are some examples:

- The federal government uses population counts from the decennial census (held in years ending in the number 1, for example, 2001, 2011) to realign the boundaries of federal electoral districts. These data are required under the *Constitution Act, 1867*, and ensure equal representation of the population in the House of Commons.
- Data from the decennial and quinquennial censuses are used to produce population estimates.
- These estimates are used in the calculation of transfer payments from the federal government to the provinces and territories, and from the provincial and territorial governments to municipalities.
 - In 2016/2017, the Government of Canada projects allocating roughly 68 billion dollars to provincial and territorial governments through its major transfers (Canada Health Transfer [CHT], Canada Social Transfer [CST], Equalization and Territorial Formula Financing [TFF]) and direct targeted support. Even a small error in the estimates could lead to the misallocation of billions of dollars.
- Government departments need to know the age trends of the population to estimate future demands for child tax benefits and old age pensions.
- Communities use census information on population growth and movement for planning services such as schools, daycare, police services and fire protection.
- Town planners, social welfare workers and other government agencies use census information on families.
- Transportation planners for provincial, territorial, regional and municipal governments use census information to analyze traffic flows, assess existing transportation services and plan for changes to these services and to road networks.
- Life insurance companies base their premium tables on census age data.
- Businesses determine new factory, store and office locations based largely on the size and distribution of the population in different areas.
- Manufacturers of household and farm equipment are guided by census data in determining the best market locations for their products. They can also assess the benefits of developing specific products by knowing the characteristics of the population in particular areas.

Overview of 2016 Census of Population products and services

Products and services from the 2011 Census Program were evaluated for their continued usefulness and relevance in 2016.

The 2016 Census of Population products and services are meant to:

- satisfy policy and market analysis data needs
- be easy to use and understand
- contain information giving users both a historical and geographical perspective
- present data by topic
- offer levels of geography with the potential to better meet users' needs
- provide users with new product options.

The Census of Population web module was also updated to enhance the user experience and improve accessibility for data users.

The 2016 Census of Population products and services line consists of five main types of products and services:

(1) Data products

These products and services have been designed to present a wide range of census information, including population and dwelling counts, and data by variable and topic. These products are released for standard geographic areas and include:

- Data tables
- Census Profile
- Age pyramids
- Aboriginal Population Profile
- Portrait of Official Language Communities
- Public Use Microdata Files

Indicators are also available on the [census website](http://www12.statcan.gc.ca/census-recensement/index-eng.cfm) (<http://www12.statcan.gc.ca/census-recensement/index-eng.cfm>), highlighting some key facts and figures at a national and provincial/territorial level.

(2) Analytical products

These products, specifically designed for the electronic medium, provide data and interpretation for selected characteristics on key findings from the 2016 Census topics. Analysis products include:

- *The Daily*
- Census in Brief
- Infographics
- Videos
- Data visualization tool.

(3) Reference products

These products are designed to help users make the most of census data. They cover various aspects of the census and are intended to support the use of the data by giving users a better understanding of the methods and concepts used. The list of reference products includes:

- Guide to the Census of Population
- Dictionary
- Reference guides
- Technical reports
- Release and concepts overview

(4) Geography products

Geography products for the 2016 Census reflect both the changes to geography concepts, as well as the more precise geometry and more detailed base map visible features (such as water, roads and road names).

A new web-based *GeoSuite* is introduced for the first time. It presents the same basic functionality as the desktop version while providing links to illustrated definitions and maps, among the other new features.

The list of geography products includes:

Reference documents

- Illustrated Glossary
- Geography Catalogue
- Reference guides
- Working papers
- Interim List of Changes to Municipal Boundaries, Status and Names

Maps

- Reference maps
- Thematic maps
- GeoSearch

Spatial information products

- Cartographic boundary files
- Digital boundary files
- Road Network File (available annually)

Attribute information products

- Geographic Attribute File
- GeoSuite (desktop version and new Internet-based version)
- Correspondence Files

(5) Custom services

These services allow for products and services to be tailored to more specific and complex requests than can otherwise be accommodated by the standard products. User-defined tabulation services are made available upon the release of each variable.

Custom services include:

- Census custom tabulations
- Semi-custom tabulations
 - Semi-custom profiles
 - Target group profiles
- Semi-custom cross-tabulations
- Census data file at the Research Data Centres (RDCs)
- Geography custom services
- Custom area creation
- Custom product creation
- Custom map creation
- Print-on-demand

Connecting with Canadians

Statistics Canada continues to use new media to provide access to relevant, accurate and timely statistical information and to foster engagement, cooperation and information-sharing among people who use statistical information. For 2016, the Census of Population is again engaging with the public through Facebook, Twitter and YouTube, and during chat with an expert sessions. A list of the different initiatives can be found on the [Stay connected](http://www.statcan.gc.ca/eng/sc/index) (<http://www.statcan.gc.ca/eng/sc/index>) page.

Availability of products and services

For 2016, census data will continue to be disseminated primarily via the Internet through Statistics Canada's website. Each release of data is summarized and published in *The Daily*.

Six official 2016 Census of Population data releases were scheduled between February 2017 and November 2017.

Census of Population major release dates

Release topic	Release date
Population and dwelling counts	February 8, 2017
Age and sex Type of dwelling	May 3, 2017
Families, households and marital status Language	August 2, 2017
Income	September 13, 2017
Immigration and ethnocultural diversity Housing Aboriginal peoples	October 25, 2017
Education Labour Journey to work Language of work Mobility and migration	November 29, 2017

In addition, two 2016 Census geography product releases took place on November 16, 2016 and February 8, 2017. A Census of Agriculture release was also scheduled for May 10, 2017.

For a complete list of release dates, refer to the [2016 Census Program release schedule](http://www12.statcan.gc.ca/census-recensement/2016/ref/release-dates-diffusion-eng.cfm) (<http://www12.statcan.gc.ca/census-recensement/2016/ref/release-dates-diffusion-eng.cfm>).

Data suppression

Introduction

Published census data go through a variety of automated and manual processes to determine whether the data need to be suppressed. This is done primarily for two reasons: (1) to ensure that the identity and characteristics of respondents is not disclosed (which will subsequently be referred to as **confidentiality**) and (2) to limit the dissemination of data of unacceptable quality (which will subsequently be referred to as **data quality**).

Overview of suppression for confidentiality reasons

Confidentiality refers to the assurance that Statistics Canada will not disclose any information that could identify respondents. Various confidentiality rules are applied to all data that are released or published to prevent the publication or disclosure of any information deemed confidential. If necessary, data are suppressed to prevent direct or residual disclosure of identifiable data. Consequently, geographic areas with a population below a certain threshold are not published.

Random rounding

All counts in census tabulations undergo random rounding, a process that transforms all raw counts into randomly rounded counts. This reduces the possibility of identifying individuals in the tabulations.

Preventing disclosure

The risk of direct or residual disclosure must also be addressed when determining product content. A number of factors must be considered when assessing the risk of disclosure. The detail of individual variables, cross-classification of variables and the geographic level of the data will all contribute to the level of risk. For example, there may be no risk in producing tables with the number of persons in the dwelling and detailed groupings of age by various characteristics of the household members for large geographic areas. However, the risk of disclosure would increase for lower levels of geography.

Area suppression for standard and non-standard geographic areas

Area suppression is used to remove all characteristic data for geographic areas whose population size is below a certain threshold. The population size threshold for all standard areas or aggregations of standard areas is 40, except for blocks, blockfaces and postal code-defined areas. Consequently, no characteristics or tabulated data are released if the total population of the area is less than 40. However, for six-character postal code areas, areas built from the block or blockface, the population size threshold is 100. These population size thresholds are applied to 2016 Census data as well as to all previous census data.

Overview of suppression due to quality

By reviewing data quality, the dissemination of data whose quality is deemed unsatisfactory can be restricted, if necessary. Quality indicators are produced for all standard geographic areas for which data are released.

Global non-response rate

The global non-response rate (GNR) is an important measure of census data quality. It combines total non-response (households) and partial non-response (questions). This measure is used for the 2016 Census, as it was for the 2011 and 2006 censuses. The GNR is calculated for dissemination of the short-form questionnaire counts and long-form questionnaire estimates. For the long-form census questionnaire, the GNR is weighted to take sampling into account. A lower GNR indicates a lower risk of non-response bias and, as a result, a lower risk of inaccuracy.

The GNR is the main dissemination criterion associated with the quality of the 2016 Census short-form questionnaire counts and long-form questionnaire estimates. The counts and estimates for geographic areas with a GNR equal to or greater than 50% are not published in the standard products. The counts and estimates for these areas have a high risk of non-response bias, and in most cases, should not be released. The 50% threshold was set based on analyses produced following the 2011 Census and National Household Survey of the GNR in relation to non-response bias indicators. Those analyses showed that with a GNR of 50% or higher, the level of bias was sufficiently high to make the estimates below an acceptable quality.

Table 11.1 shows the GNR for the 2016 Census short-form and long-form questionnaires at the national level and for each province and territory. At the national level, the GNR for the 2016 Census short-form questionnaire is 4.1%.

Table 11.1
Global non-response rate for the 2016 Census short-form and long-form questionnaires, Canada, provinces and territories

Provinces and territories	Short-form questionnaire	Long-form questionnaire
	Global non-response rate (%)	
Canada	4.0	5.1
Newfoundland and Labrador	4.0	6.8
Prince Edward Island	4.1	5.6
Nova Scotia	3.9	5.5
New Brunswick	3.9	4.9
Quebec	3.7	4.3
Ontario	3.7	4.6
Manitoba	4.3	5.3
Saskatchewan	4.8	6.3
Alberta	4.7	6.1
British Columbia	4.9	6.1
Yukon	6.7	6.8
Northwest Territories	7.5	8.8
Nunavut	7.9	8.7

Source: Statistics Canada, Census of Population, 2016.

Coverage of data published from the 2016 Census short-form questionnaire

Canada has a total of 152 census metropolitan areas (CMAs) and census agglomerations (CAs). For all of these areas, the GNR of the short-form questionnaire is below 50% and 2016 Census data are available in the standard products. In addition, the 2016 Census short-form questionnaire standard products are available for all 293 census divisions (CDs) and 338 federal electoral districts (FEDs).

With the threshold for the global non-response rate for the short-form questionnaire counts set at 50%, 2016 Census data can be published for the vast majority of census subdivisions (CSDs) or municipalities. As a result, short-form questionnaire counts are published in the standard products for 4,585 CSDs, representing 99.6% of the 4,603 CSDs with a population of more than 40 inhabitants (those with a population of less than 40 are not published for confidentiality reasons). Table 11.2 shows the distribution of the total number of CSDs and the number of CSDs for which data have been published, by province and territory.

Table 11.2

Data published by census subdivision, 2016 Census short-form questionnaire counts, Canada, provinces and territories

Provinces and territories	Total number of census subdivisions	Census subdivisions for which data have been published		
		number	percentage	population covered (%)
Canada	4,603	4,585	99.6	99.99
Newfoundland and Labrador	354	354	100.0	100.00
Prince Edward Island	111	111	100.0	100.00
Nova Scotia	85	85	100.0	100.00
New Brunswick	268	268	100.0	100.00
Quebec	1,174	1,171	99.7	100.00
Ontario	543	537	98.9	99.99
Manitoba	212	211	99.5	99.96
Saskatchewan	837	835	99.8	99.97
Alberta	398	396	99.5	99.97
British Columbia	536	532	99.3	99.99
Yukon	25	25	100.0	100.00
Northwest Territories	34	34	100.0	100.00
Nunavut	26	26	100.0	100.00

Note: CSDs for which data have not been published for confidentiality reasons are excluded from this table. They have a population less than 40.

Source: Statistics Canada, Census of Population, 2016.

Coverage of data published for the 2016 Census long-form questionnaire

In Canada, the global non-response rate (GNR) for census metropolitan areas (CMAs) and census agglomerations (CAs) is below 50%, and 2016 Census data are available in the standard products. Standard products for the 2016 Census long-form questionnaire are available for all 293 census divisions (CDs) and all 338 federal electoral districts (FEDs).

As with the short-form questionnaire, the GNR threshold for the long-form questionnaire estimates is set at 50%—as a result, 2016 Census data can be published for the majority of census subdivisions (CSDs) and municipalities. Long-form questionnaire estimates are published in the standard products for 4,561 CSDs, representing 99.8% of the 4,567 CSDs with a population of more than 40 inhabitants (data on those with a population of fewer than 40 are not published for confidentiality reasons). Table 11.3 shows the distribution of the total number of CSDs and the number of CSDs for which data have been published, by province and territory.

The GNR for the long-form questionnaire is calculated from a sample—it is therefore an estimate of the global non-response. As with all estimates calculated from a sample, the GNR is subject to sampling error. There is variability associated with the GNR for the long-form questionnaire. For each geographic level, the GNR estimate is presented to users. For regions with a very high response rate for the long-form questionnaire (e.g., a response rate above 99%), the GNR estimate may be zero or very close to zero. A GNR of zero for a given region does not necessarily mean that the response rate for the long-form questionnaire in that region was 100% or that all returned questionnaires were completed in full.

Table 11.3

Data published by census subdivision, 2016 Census long-form questionnaire counts, Canada, provinces and territories

Provinces and territories	Total number of census subdivisions	Census subdivisions for which data have been published		
		number	percentage	population covered (%)
Canada	4,567	4,561	99.8	100.00
Newfoundland and Labrador	352	352	100.0	100.00
Prince Edward Island	111	111	100.0	100.00
Nova Scotia	85	85	100.0	100.00
New Brunswick	268	268	100.0	100.00
Quebec	1,169	1,169	100.0	100.00
Ontario	543	541	99.6	100.00
Manitoba	211	210	99.5	99.96
Saskatchewan	814	814	100.0	100.00
Alberta	395	395	100.0	100.00
British Columbia	534	531	99.4	99.99
Yukon	25	25	100.0	100.00
Northwest Territories	34	34	100.0	100.00
Nunavut	26	26	100.0	100.00

Note: CSDs for which data have not been published for confidentiality reasons are excluded from this table. These are CSDs with a population of fewer than 40.

Source: Statistics Canada, Census of Population, 2016.

Chapter 12 – Census geography

Introduction

There is a geographic component in every stage of the census cycle, from consultation through collection, processing and dissemination. Users are consulted about the geographic concepts used by Statistics Canada and about various options for disseminating standard geographic data. Small geographic areas are defined and mapped in detail so that every dwelling can be located during the data collection phase, while during the processing phase, the collected data are coded to the appropriate geographic areas within the geographic hierarchy. Finally, census data are disseminated by a variety of geographic areas which are designed specifically for disseminating data, along with supporting reference maps and other geographic data products.

National Geographic Database

The standard geographic areas that Statistics Canada uses for census and survey collection and dissemination activities are constructed, maintained and supported by detailed geographic data which are stored in a precise geographic database called the National Geographic Database (NGD).

The NGD is a joint Statistics Canada-Elections Canada initiative which develops and maintains a geospatial database which serves the needs of both organizations. The focus of the NGD is the continual improvement of quality and currency of geographic coverage using updated geospatial data provided by provinces, territories and local sources.

The NGD includes a digital representation of the boundaries of standard geographic areas, as well as their attributes such as names, types and codes, which are necessary for uniquely identifying each individual geographic area.

The NGD also contains additional geographic features, including a detailed road network, various hydrographic features such as lakes, rivers and coastal waters, and other selected visible features, for example, railroads. The road network also has associated attribute data, such as street names, types, directions and address ranges.

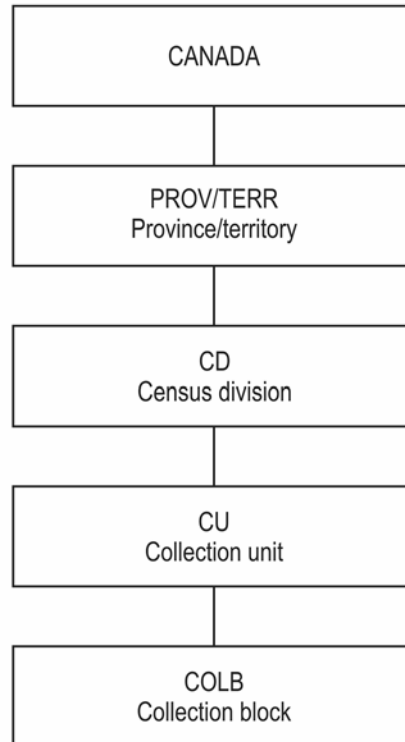
To take full advantage of census data, users are encouraged to develop a basic understanding of the geographic dimension of the data.

Hierarchical model of geographic areas for collection

The geographic areas used for census data collection are illustrated below in Figure 12.1, Hierarchy of geographic areas for collection, 2016 Census, and are different than geographic areas used for disseminating data.

The geographic areas used for census data collection include Canada, provinces and territories, census divisions, collection units and collection blocks (Figure 12.1).

Figure 12.1
Hierarchy of geographic areas for collection, 2016 Census



Data are not published using collection units or collection blocks and therefore are not represented in the [Figure 1.1](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/figures/f1_1-eng.cfm) (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/figures/f1_1-eng.cfm), Hierarchy of standard geographic areas for dissemination, 2016 Census (see the [Dictionary, Census of Population, 2016](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/index-eng.cfm) (<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/index-eng.cfm>), Catalogue no. 98-301-X).

Hierarchical model for the dissemination of geographic areas

Just as one can subdivide a population by sex, or into age and language groups, one can subdivide a population by different geographic areas. The geographic areas used for disseminating census data range in size from Canada, provinces and territories, all the way down to dissemination blocks, and are organized in a hierarchical model to illustrate the nature of their relationships to one another.

Standard geographic areas used for data dissemination and their relationships to one another are depicted in [Figure 1.1](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/figures/f1_1-eng.cfm) (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/figures/f1_1-eng.cfm), Hierarchy of standard geographic areas for dissemination, found in the *Dictionary, Census of Population, 2016*, Catalogue no. 98-301-X

- Each box in the hierarchy chart represents an individual geographic level, which is composed of one or more geographic areas.

- The relative position of each geographic level in the chart shows how it can be subdivided or aggregated to form other geographic levels. For example, the Canada level can be subdivided into six regions which comprise 13 provinces and territories which in turn are subdivided into 338 federal electoral districts (federal ridings), which are comprised of almost 500,000 dissemination blocks, the lowest level of standard geographic areas.
- The lines which join the boxes in the geographic hierarchy chart illustrate the relationship between the geographic areas which comprise each geographic level. In general, this is a 'one-to-many' relationship when moving from one geographic level down to a lower geographic level, for example, moving from 13 provinces and territories down to 338 federal electoral districts. From bottom to top, the relationship is 'many-to-one.'
- Each branch of the geographic hierarchy illustrates how different geographic levels relate to the geographic areas of lower geographic levels. For example, the geographic hierarchy chart shows that dissemination areas (DAs) group together to form census subdivisions (CSDs); they also group together to form census tracts (CTs); however, there is no exact fit relationship between census subdivisions and census tracts and therefore there is no line joining the CSD and CT boxes of these geographic levels within the geographic hierarchy chart.

The hierarchy of geographic areas: understanding the hierarchy, how geographies are related and data analysis

The geographic hierarchy illustrates how one can carry out geographic analysis starting with higher-level geographic areas and moving to the lower-level geographic areas (a top-down approach). For example, one can start with Canada and then look at each of the 13 provinces and territories, and continue by looking at individual or groupings of census divisions (CDs) and census subdivisions (CSDs). Conversely, using a bottom-up approach, one can start by examining specific individual lower-level geographic areas, census subdivisions (CSDs) for example, by comparing them with each CSD within a particular census division (CD), and then comparing CDs within the same province or territory, and eventually within or among regions and Canada as a whole.

Geographic areas for dissemination

In [Figure 1.1](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/figures/f1_1-eng.cfm) (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/figures/f1_1-eng.cfm), Hierarchy of standard geographic areas for dissemination, found in the *Dictionary, Census of Population, 2016*, geographic areas used for disseminating data are depicted as being either administrative areas or statistical areas. Administrative areas are defined, with a few exceptions, by federal, provincial or territorial statutes, and are adopted for the purposes of the census. Statistical areas, on the other hand, are defined by Statistics Canada, in cooperation with stakeholders, for the purpose of disseminating census data and complementing the structure of administrative areas.

This census cycle, a new statistical area called 'aggregated dissemination area (ADA)' is available. This level of geography will allow for more data to be released at smaller levels of geography. For more information on the ADAs, refer to the [aggregate dissemination area \(ADA\)](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo053-eng.cfm) (<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo053-eng.cfm>) definition in the *Dictionary, Census of Population, 2016*, Catalogue no. 98-301-X.

The number of geographic areas by province and territory for the 2016 Census is presented in [Table 1.1](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/tab/t1_1-eng.cfm) (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/tab/t1_1-eng.cfm), Geographic areas by province and territory, of the *Dictionary, Census of Population, 2016*, Catalogue no. 98-301-X.

Refer to the [Introduction to the geography universe](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geoint-eng.cfm) (<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geoint-eng.cfm>) section of the *Dictionary, Census of Population, 2016*, Catalogue no. 98-301-X for definitions and more detailed information on each of the following administrative and statistical areas.

Administrative areas

- Canada
- [Province or territory](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo038-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo038-eng.cfm>)
- [Federal electoral district \(FED\)](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo025-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo025-eng.cfm>)
- [Census division \(CD\)](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo008-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo008-eng.cfm>)
- [Census subdivision \(CSD\)](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo012-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo012-eng.cfm>)
- [Designated place \(DPL\)](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo018-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo018-eng.cfm>)
- [Forward sortation area \(FSA\)^{OM}](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo036-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo036-eng.cfm>)
- [Postal code^{OM}](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo035-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo035-eng.cfm>)

^{OM}: Postal code is an official mark of Canada Post Corporation.

Statistical areas

- Region
- [Census agricultural region \(CAR\)](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo006-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo006-eng.cfm>)
- [Economic region \(ER\)](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo022-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo022-eng.cfm>)
- [Census consolidated subdivision \(CCS\)](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo007-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo007-eng.cfm>)
- [Aggregated dissemination area \(ADA\)](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo053-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo053-eng.cfm>)
- [Dissemination area \(DA\)](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo021-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo021-eng.cfm>)
- [Dissemination block \(DB\)](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo014-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo014-eng.cfm>)
- [Statistical Area Classification \(SAC\)](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo045-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo045-eng.cfm>)
- [Census metropolitan area \(CMA\)](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo009-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo009-eng.cfm>)
- [Census agglomeration \(CA\)](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo005-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo005-eng.cfm>)
- [Census tract \(CT\)](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo013-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo013-eng.cfm>)
- [Census metropolitan influenced zone \(MIZ\)](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo010-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo010-eng.cfm>)
- [Population centre \(POPCTR\)](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo049a-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo049a-eng.cfm>)

Other

- [Place name \(PN\)](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo033-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo033-eng.cfm>)
- [Blockface](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo003-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo003-eng.cfm>)
- [Census subdivision - previous census](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo011-eng.cfm)
(<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo011-eng.cfm>)

Non-standard or user-defined geographic areas for dissemination

In most cases, the standard geographic areas for dissemination satisfy data user requirements for census data tabulations; however, there are also data users who require that data, which are not tabulated from the standard geographic hierarchy depicted in [Figure 1.1](http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/figures/f1_1-eng.cfm) (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/figures/f1_1-eng.cfm), Hierarchy of standard geographic areas for dissemination, *Dictionary, Census of Population, 2016*, be tabulated for specific geographic areas.

There are two basic types of non-standard or 'user-defined' geographic areas: areas that are customized aggregations of individual standard geographic areas, and areas that do not match the standard geographic areas at all. An example of the first type could be user-created sales regions within a metropolitan area, where the sales regions are created by combining one or more specific census subdivisions. An example of areas that do not match standard geographic areas could be user-defined market areas, school districts or transportation and utility corridors. When data users require that census data be tabulated for non-standard geographic areas, they may turn to the Custom Area Creation Service provided by Statistics Canada (see [Chapter 11 – Dissemination](#)).

Appendix 1.1 – Legislation

Introduction

Statistics Canada is required by law to conduct a Census of Population and a Census of Agriculture every five years, in the years ending in 1 and 6. The most recent Canadian census was held in May 2016.

The relevant provisions of the *Statistics Act* are as follows:

Subsection 19(1):

“A census of population of Canada shall be taken by Statistics Canada in the month of June in the year 1971, and every fifth year thereafter in a month to be fixed by the Governor in Council.”

Section 20:

“A census of agriculture of Canada shall be taken by Statistics Canada

- (a) in the year 1971 and in every tenth year thereafter; and
- (b) in the year 1976 and in every tenth year thereafter, unless the Governor in Council otherwise directs in respect of any such year.”

Subsection 21(1):

“The Governor in Council shall, by order, prescribe the questions to be asked in any census taken by Statistics Canada under section 19 or 20.”

Subsection 21(2):

“Every order made under subsection (1) shall be published in the *Canada Gazette* not later than thirty days after it is made.”

Requirement to respond

Just as Statistics Canada is required by law to conduct a census, respondents are required by law to complete their census questionnaires.

This requirement is set out in section 23 of the *Statistics Act*, which reads as follows:

Subsection 23(1):

“In lieu of or in addition to using agents or employees for the collection of statistics under this Act, the Minister may prescribe that a form be sent to a person from whom information authorized to be obtained under this Act is sought.”

Subsection 23(2):

“Subject to section 8, a person to whom a form is sent pursuant to subsection (1) shall answer the inquiries thereon and return the form and answers to Statistics Canada properly certified as accurate, not later than the time prescribed therefor by the Minister and indicated on the form or not later than such extended time as may be allowed in the discretion of the Minister.”

The requirement to respond is supported by the penalty provisions found in section 31 of the Act, which states:

Subsection 31:

“Every person who, without lawful excuse,

- (a) refuses or neglects to answer, or wilfully answers falsely, any question requisite for obtaining any information sought in respect of the objects of this Act or pertinent thereto that has been asked of him by any person employed or deemed to be employed under this Act, or
- (b) refuses or neglects to furnish any information or to fill in to the best of his knowledge and belief any schedule or form that the person has been required to fill in, and to return the same when and as required of him pursuant to this Act, or knowingly gives false or misleading information or practises any other deception thereunder is, for every refusal or neglect, or false answer or deception, guilty of an offence and liable on summary conviction to a fine not exceeding five hundred dollars or to imprisonment for a term not exceeding three months or to both.”

Statutory references to the census

Constitutional law

- a. A decennial census (i.e., a census every ten years) in the year 1871 and every tenth year thereafter is required under section 8 of the *Constitution Act, 1867* (formerly named the *British North America Act, 1867*).
- b. Conduct of the census is made the responsibility of the federal government under section 91, subsection 6 of the *Constitution Act, 1867*.
- c. Representation in the House of Commons is made dependent on decennial census data under section 51 of the *Constitution Act, 1867*, as amended by the *Representation Act, 1974*.
- d. The amending formula for the *Constitution Act* is made dependent on population data from the “latest general census” under section 38 of the *Canada Act, 1982*.
- e. A number of provisions relating provincial subsidies to population have been legislated and amended over the years. The following is a summary of this legislation:
 - The *Constitution Act, 1930*, Schedule, replaced the 1907 legislation with respect to the three Prairie provinces. A subsidy was made payable to these provinces (Alberta, Saskatchewan and Manitoba) based on quinquennial census population counts and a variable with these counts up to a maximum population of 1,200,000. These provisions are still in effect.

Legal opinions provided to Statistics Canada indicate a constitutional obligation to conduct a quinquennial census of the Prairie provinces exists until such time as their populations exceed 1,200,000. Since 1961, the population of Alberta has exceeded 1,220,000.
 - The *Newfoundland Act, 1949*, Schedule, part 26, made a federal subsidy to that province dependent on decennial census population counts. This provision is still in effect.
- f. Representation of Alberta and Saskatchewan in the House of Commons was made dependent on the mid-decade census of those provinces for the first mid-decade census subsequent to their creation only (i.e., 1906). Thereafter, representation was to be based on the decennial census of Canada (*Alberta Act, 1905*, section 6; *Saskatchewan Act, 1905*, section 6).

Appendix 1.2 – Incompletely enumerated Indian reserves and Indian settlements

In 2016, there were a total of 14 Indian reserves and Indian settlements that were incompletely enumerated. For these reserves and settlements, dwelling enumeration was either not permitted or was interrupted before it could be completed.

This represents a decrease compared to the 31 Indian reserves and Indian settlements that were incompletely enumerated in the 2011 Census. Note that in 2011, of the 31 incompletely enumerated Indian reserves and Indian settlements, 13 were not enumerated as a result of forest fires in Northern Ontario at the time of census collection. In 2016, there were no Indian reserves or Indian settlements that were not enumerated due to a natural disaster.

The 2016 Census population and dwelling counts are not available for the 14 incompletely enumerated Indian reserves and Indian settlements, and are not included in 2016 Census tabulations. Data for geographic areas containing one or more of these reserves and settlements are noted accordingly. Because of the missing data, users are cautioned that for the affected geographic areas, comparisons (e.g., percentage change) between 2011 and 2016 may not be precise. The impact of the missing data for higher-level geographic areas (Canada, provinces and territories, census metropolitan areas and census agglomerations) is very small. However, the impact can be significant for lower-level geographic areas (e.g., census divisions), where the incompletely enumerated Indian reserves and Indian settlements account for a higher proportion of the population. This is especially true for lower-level geographic areas where a particular Indian reserve or Indian settlement was incompletely enumerated for the 2016 Census and enumerated for the 2011 Census and vice versa.

Table 1 provides the list of incompletely enumerated Indian reserves and Indian settlements for the 2016 Census, along with population counts from the last two censuses (where available).

Table 2 shows, in alphabetical order, the list of incompletely enumerated Indian reserves and Indian settlements for the 2016 Census by province, census division and, where applicable, for the census metropolitan area or census agglomeration.

Table 1

2016 Incompletely enumerated Indian reserves and Indian settlements, showing enumeration status for the 2016 Census, 2011 and 2006 population counts (where available)

Province	Incompletely enumerated Indian reserves and Indian settlements, 2016	Enumeration status for the 2016 Census (reasons for absence of data)	Population, 2011	Population, 2006
Quebec	Kanesatake	Permission not given
	Doncaster	Permission not given
	Kahnawake	Permission not given
	Lac-Rapide	Permission not given
Ontario	Six Nations (Part) 40	Permission not given	946	..
	Six Nations (Part) 40	Permission not given	6,213	..
	Chippewas of the Thames First Nation 42	Permission not given	762	747
	Oneida 41	Permission not given	1,282	..
	Wahta Mohawk Territory	Permission not given
	Rankin Location 15D	Permission not given	..	566
	Goulais Bay 15A	Permission not given	..	82
	Pikangikum 14	Dwelling enumeration not completed – other	..	2,100
Alberta	Saddle Lake 125	Permission not given
British Columbia	Esquimalt	Permission not given

.. not available for a specific reference period. Incompletely enumerated Indian reserve or Indian settlement.

Permission not given: Band council did not give permission to enter their territory.

Dwelling enumeration not completed – other: Enumeration was not completed for reasons such as access restrictions, health and safety issues, etc.

Sources: Statistics Canada, Censuses of population, 2016, 2011 and 2006.

Table 2

Incompletely enumerated Indian reserves and Indian settlements in alphabetical order, 2016

Name	Type	Province	Census division	CMA/CA
Chippewas of the Thames First Nation 42	IRI	Ontario	Middlesex	...
Doncaster	IRI	Quebec	Les Laurentides	...
Esquimalt	IRI	British Columbia	Capital	Victoria
Goulais Bay 15A	IRI	Ontario	Algoma	...
Kahnawake	IRI	Quebec	Roussillon	Montréal
Kanesatake	S-É	Quebec	Deux-Montagnes	Montréal
Lac-Rapide	IRI	Quebec	La Vallée-de-la-Gatineau	...
Oneida 41	IRI	Ontario	Middlesex	...
Pikangikum 14	IRI	Ontario	Kenora	...
Rankin Location 15D	IRI	Ontario	Algoma	Sault Ste. Marie
Saddle Lake 125	IRI	Alberta	Division No. 12	...
Six Nations (Part) 40	IRI	Ontario	Brant	Brantford
Six Nations (Part) 40	IRI	Ontario	Haldimand-Norfolk	...
Wahta Mohawk Territory	IRI	Ontario	Muskoka	...

... not applicable

IRI = Indian reserve

S-É = Indian settlement

Source: Statistics Canada, Census of Population, 2016.

Appendix 1.3 – Difference between census counts and population estimates

The Census of Population is designed to conduct a complete count of the population. Inevitably, however, some individuals will not be enumerated (undercoverage), while others, usually less numerous, will be enumerated more than once (overcoverage).

To determine the number of people who were missed or counted more than once, Statistics Canada conducts postcensal studies of the coverage of the census population, using representative samples of the population. Results of these studies are usually available two years after Census Day. They are used, in combination with census figures and other sources, to develop the population estimates produced by Statistics Canada on a regular basis. Population estimates are used for equalization payments, to follow trends in the Canadian population on a quarterly basis and to understand the underlying components of population change (for example, births, deaths, immigrants, emigrants and non-permanent residents). Population estimates differ from census counts and are usually higher, because census counts are not adjusted for undercoverage or overcoverage.

Appendix 1.4 – Note describing the Wood Buffalo census subdivision data collection methodology and the use of administrative data sources

Introduction

On May 1, 2016, a wildfire began southwest of Fort McMurray, Alberta, and on May 3, swept through the community destroying many homes and buildings and forcing the largest wildfire evacuation in Alberta's history. Statistics Canada then decided to suspend census data collection (referred to as 'field data collection') in the evacuated areas.

Statistics Canada used a set of measures to ensure that residents of the Wood Buffalo census subdivision (CSD) (referred to as the Specialized municipality of Wood Buffalo or Wood Buffalo) were included in the 2016 Census of Population. Data for the evacuated area were derived from a combination of sources. First, many residents of the area responded online or by returning a paper questionnaire. Then field data collection was performed for a number of households using short- or long-form questionnaires. Lastly, short-form data were derived from a number of administrative data sources for the households residing in dwellings where field data collection was not possible. Data for all areas not evacuated due to the wildfire are from direct field data collection.

Reference date

For the 2016 Census, the reference date for data reporting is May 10, 2016. For residents of the evacuated areas during the wildfire, the reference date is May 1, 2016, to reflect the situation as it existed before the fire.

Data collection

Prior to the evacuation, and even in the following weeks when census data collection was suspended, some responses were received from the residents of the evacuated area. In August 2016, data collection was reinstated in Wood Buffalo and census representatives went door to door to complete census questionnaires. Efforts were focussed on collecting data for the one in four dwellings included in the long-form questionnaire sample. This was particularly important, as administrative data sources do not provide information for long-form questions. To further improve data quality, field data collection was also performed for dwellings in the areas for which no administrative data were available and for collective dwellings. In areas where enumerators prepare a list of dwellings and deliver census materials, field data collection was done for all dwellings.

Administrative data

Wherever possible and when no direct response had been received for a dwelling, data from various administrative data sources were used with a reference date as close as possible to May 2016, for variables such as name, date of birth, sex and marital status. As administrative data files did not contain information on language as collected on the census questionnaire, record linkages between the administrative sources and the 2011 Census database were performed. For successful linkages, the 2011 responses to the language questions were used as proxy for the 2016 language questions. Census questions for which no comparable information could be obtained from administrative data files, such as Relationship to Person 1 and common-law status, were derived during data processing.

Statistics Canada worked closely with both provincial and local authorities in Alberta to obtain access to administrative records to assist in the validation of the data derived from administrative data sources available in Statistics Canada.

Processing

If a census response was obtained for residents of a dwelling, this took precedence over any available administrative data. For the remaining cases, during data processing and for the calculation of response rates, data from administrative sources were considered as a response to the same extent as a direct response obtained through traditional collection methods.

Data quality for population and dwelling counts

For the population and dwelling counts, the Wood Buffalo CSD data went through the same quality assessments as the overall census data. A supplementary pre-validation activity was performed by Statistics Canada once data from field collection and administrative sources were combined. This additional step was done to certify that the alternative methods developed for this exceptional situation were providing satisfactory results.

Short-form questionnaire data quality

To obtain data on age, sex and families, Statistics Canada used administrative data for 54% of households, questionnaire data for 40% of households and imputation for the other data (5% of the remaining households).

For households for which administrative data were used, the age and sex data were taken directly from administrative data files. The distribution of the age and sex data from the administrative files and of the data taken directly from the completed questionnaires is comparable for both enumeration methods (administrative data and completed questionnaires). However, with respect to households enumerated using administrative data, the biggest determinant for attributing family characteristics was the use of marital status and parent-child relationship established during linkage with the tax data. A larger number of lone-parent families following processing of the administrative data than the completed questionnaire data was observed. The corresponding proportions were 19.2% (administrative data) and 9.8% (completed questionnaires).

This discrepancy had an impact on the proportion of families consisting of couples without children, which was 30.0% and 40.9%, respectively, depending on the enumeration method. There also seems to be a difference in the data on households taken from the administrative data files with respect to the number of people living common-law; the proportion from the administrative data is smaller than from data from traditional collection. There is also a significant difference in terms of the size of household; proportionally, there are far more one-person households and six or more person households in the administrative data than in the questionnaire data.

For households for which administrative data were used, the 2016 Census data on language were obtained from responses to the 2011 questions on language when linkage was possible. A comparison of the distribution of language variables does not show as many differences for households for which administrative data were used as it does for households for which the data came from completed questionnaires. Comparing the 2011 and 2016 figures for the family and language variables for the Wood Buffalo CSD must be done with caution.

Appendix 1.5 – Information produced from the 2016 Census of Population

Who is included in the population of Canada?

The Census of Population aims to produce counts for the total population of Canada. This ‘target population’ consists of: Canadian citizens (by birth or by naturalization); landed immigrants (permanent residents); and (since 1991) non-permanent residents and their family members living with them in Canada. (Non-permanent residents are persons who hold a work or study permit, or who are claiming refugee status.) All such persons are included in the population provided they have a usual place of residence in Canada (see [Where are people counted?](#)).

The total population also includes certain Canadian citizens and landed immigrants (permanent residents) living outside the country: government employees working outside Canada; embassy staff posted to other countries; members of the Canadian Armed Forces stationed outside Canada; and Canadian crew members of merchant vessels and their families. Together, they are referred to as persons living outside Canada.

Foreign residents are excluded from census data: for example, residents of another country visiting Canada temporarily, government representatives of another country posted in Canada and members of the armed forces of another country stationed in Canada.

Where are people counted?

The census counts people according to their usual place of residence. The 2016 Census questionnaire included questions and instructions to determine either the person’s sole residence or their main residence as of May 10, 2016. This location is then used in all data products by geographic area. It is also used to determine which people reside in the same dwelling together—an important aspect of census data. For more information about where people are counted, refer to the [Dictionary, Census of Population, 2016](#) (<http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/index-eng.cfm>), Catalogue no. 98-301-X, usual place of residence.

Who is included in the data for each topic?

Between the different topics released from the census, there are slight differences in what is included in the data. Refer to Table 1 for an illustration of the different statistical units and universes by the topics covered in the Census of Population.

Table 1
Statistical units and universes for the dissemination of the 2016 Census of Population, by topic

Topic	Statistical unit	Maximum universe available
Population counts	Persons	Total population
Dwelling counts	Dwellings	Private dwellings ¹
Type of dwelling (collective or private)	Dwellings, Persons	Occupied dwellings ¹ (or population in)
Age and sex	Persons	Total population
Marital status	Persons	Total population
Family and household characteristics	Persons, families and households	Private households (or population in, or families in)
Language	Persons	Population excluding institutional residents ²
Income	Persons, families and households	Private households (or population in, or families in)
Immigration and ethnocultural diversity	Persons	Population in private households
Housing	Households, persons	Private households (or population in)
Aboriginal peoples	Persons	Population in private households
Education	Persons	Population in private households
Labour	Persons	Population in private households
Journey to work	Persons	Population in private households
Language of work	Persons	Population in private households
Mobility and migration	Persons	Population in private households

1. Dwellings that are either unoccupied or occupied solely by foreign residents and/or by temporarily present persons are excluded from most data on dwellings. Most data on dwellings are presented for occupied private dwellings (occupied by at least one person for whom this is their usual place of residence). This universe corresponds to that of private households.

2. Comprises the population in private households and the population in non-institutional collective dwellings. Also referred to as the non-institutional population.

Source: Statistics Canada, Census of Population, 2016.

Generally, topics which are included in the short-form questionnaire directed to the 'target' population are available for the total population. Additional questions are asked on the long-form questionnaire to the population in private households, which is the total population, excluding persons living outside Canada (as described above) and excluding persons living in collective dwellings. Collective dwellings are classified as either institutional, such as hospitals, nursing homes and penitentiaries, or non-institutional, such as work camps, hotels and motels, and student residences.

The applicable population for each topic is also referred to as the universe for that topic.

Depending on the characteristics being portrayed, the data may represent these statistical units:

- Persons (or population)
- Dwellings
- Households
- Families

Appendix 1.6 – Results of the 92-year consent question

The consent question used in the 2016 Census gave Canadians the opportunity to make an informed decision about the release of their census records in 92 years.

The *Statistics Act* allows for the release of name-identified census information to be transferred to Library and Archives Canada and made publicly available 92 years after the 2016 Census of Population. Consenting to the release of census information will help future generations better understand the Canada of today, and will benefit historical, academic and genealogical research. If consent is not given, then personal information will remain confidential in perpetuity.

In 2016, 81.5% of respondents to the Census of Population answered yes, and the remainder either responded no or left the option blank.

Release of personal census information after 92 years

The following table shows how Canadians responded to the 92-year consent question on the 2016 Census questionnaire, both in Canada as a whole, and in individual provinces and territories.

2016 Census

Geographic name	% who responded yes	% who responded no/left blank
Canada	81.5	18.5
Newfoundland and Labrador	85.2	14.8
Prince Edward Island	85.1	14.9
Nova Scotia	86.6	13.4
New Brunswick	84.0	16.0
Quebec	83.4	16.6
Ontario	80.8	19.2
Manitoba	80.5	19.5
Saskatchewan	81.0	19.0
Alberta	79.8	20.2
British Columbia	80.2	19.8
Yukon	82.1	17.9
Northwest Territories	83.1	16.9
Nunavut	78.9	21.1

Source: Statistics Canada, Census of Population, 2016.

Appendix 1.7 – Percentage change for population and dwelling counts at various levels of geography

The percentage change for the population and dwelling counts is based on a 2011 count that may have been revised since the publication of the 2011 population and dwelling counts. If the 2011 count has been revised, it will be indicated by an 'r' beside the count. The 2011 population and dwelling counts may be revised due to the reasons identified below.

(1) The boundary of the geographic area has changed since the 2011 Census

When a boundary of a geographic area changed, Statistics Canada identified the impact of the change on the population and dwellings enumerated during the 2011 Census, and adjusted these counts to reflect the 2016 boundaries of the geographic areas. Most of the revisions to 2011 population and dwelling counts are due to this type of adjustment.

(2) A formal review of the 2011 population and dwelling counts identified an error

When Statistics Canada releases population and dwelling counts from the census, data users sometimes question the validity of the counts for a specific geographic area such as a municipality (census subdivision) or sub-municipal area (designated place). When requested by local authorities, Statistics Canada undertakes a formal review of the population and dwelling counts.

In 2011, 158 census subdivisions (CSDs) and 40 designated places (DPLs) underwent a formal population and dwelling count review and their counts were revised. For the 2016 Census products showing 2011 counts, these revised 2011 counts are presented in tables for CSDs and DPLs and used to calculate percent change between 2011 and 2016 at that level of geography.

The revised counts are presented in tables **only** for CSDs and DPLs. The 2011 counts for other levels of geography (e.g. province, territory, census division) are not revised. As a result, aggregating 2011 CSD counts to higher-level geographic areas (e.g. census divisions, provinces, territories, and Canada) may not sum to the counts presented on the higher-level geographic areas. As well, the percent change for all levels of geography, other than CSDs and DPLs, do not account for these revisions.

Appendix 1.8 – Evaluation of the impact of updating 2016 Census language data

After Statistics Canada was notified of anomalies, for various regions in Quebec, in the 2016 Census language data released on August 2, it conducted an in-depth investigation to identify their source.

During the census, Statistics Canada follows up with households that partially completed the census questionnaire. For the 2016 Census, Statistics Canada developed a computer program to perform certain steps of this operation. An error was identified in this computer program that affected French-language questionnaires.

The census language questions are the only questions for which response options differ between the French and English versions. The French version of the census questionnaire has given precedence to French in the wording of the questions and the response options. This only affects census questions on mother tongue, language spoken at home and knowledge of official languages. The example below illustrates this difference with respect to language spoken most often at home. This distinction was not taken into account by the new computer program used for partial non-response follow-up in 2016.

8 a) What language does this person speak most often at home?	<input type="radio"/> English <input type="radio"/> French Other language – specify: <input type="text"/>	8 a) Quelle langue cette personne parle-t-elle le plus souvent à la maison?	<input type="radio"/> Français <input type="radio"/> Anglais Autre langue – précisez : <input type="text"/>
--	--	--	--

The resulting error led to incorrect allocation of responses to the census language questions for roughly 61,000 individuals, mainly in Quebec. It resulted in an overestimation of the growth of English in Quebec between 2011 and 2016, both as a mother tongue and as a language spoken at home.

After correcting these allocation errors, Statistics Canada conducted an in-depth review to ensure that no other census questions were affected by an error and that the computer programs did not affect other variables. Moreover, Statistics Canada extensively reviewed the many data editing and control stages.

The results of this analysis and the corrective steps taken are described below.

Potential impact on other variables

Following the detection of the language allocation error, Statistics Canada conducted a review to determine if the error negatively impacted the quality of estimates for any other variables. After a thorough review of systems, programs and estimation procedures, Statistics Canada concluded that the impact on the estimates for all other variables was negligible.

For the 2016 Census variables, the vast majority of responses were received directly from respondents either on the census questionnaire or through a record linkage to administrative files. For example, 99% of place of birth responses came directly from respondents and 95% of income values came from high-quality administrative data (including tax data from the Canada Revenue Agency).^{2,3} The incorrect allocation of responses to the language variables could only potentially affect estimates for the remaining records which required imputation through the donor imputation process. In that process, a respondent's missing or invalid values (the recipient) are replaced

2. Estimates in this section are rounded

3. For detailed imputation rates by variable, please refer to the individual [reference guides](http://www12.statcan.gc.ca/census-recensement/2016/ref/index-eng.cfm) (<http://www12.statcan.gc.ca/census-recensement/2016/ref/index-eng.cfm>).

from other respondents with similar characteristics (the donor). The incorrect allocation of responses to the language variables could affect the estimates for that record in one of two ways:

1. A recipient record with a miscoded language is imputed with a donor. It could be that a donor with a different value would have been chosen had the recipient's language been coded correctly.
2. A recipient record with a correctly coded language is imputed with a donor. It could be that a donor was incorrectly chosen, based partly upon the donor having a miscoded language.

In this donor imputation process, donors are chosen by a score which reflects their similarity to the recipient. Usually, a donor is randomly chosen from among several similarly qualified donors. It is important to underscore that this random selection causes variability in the donor imputation process, and any re-running of the donor imputation system could result in slightly different estimates due to this variability.

Statistics Canada conducted an analysis to determine whether the language error affected the results of the donor imputation process. To begin with, the potential effect of the language error would be considerably mitigated by the following factors:

1. Not all variables used language variables to select donors for imputation. Income, immigration, ethnocultural, Aboriginal, mobility and migration, and language of work were the only topics which used language variables in their imputation processes. Age and sex, dwelling type, families, marital status, and living arrangements were not affected since language variables (or any of the variables listed above) were not used to select donors for imputation.
2. Very few imputed cases were affected by the language allocation error. As shown in [Table 1](#), most of the variables which used language to select donors were imputed fewer than 1,000 times among the affected cases.
3. The affected records were concentrated in Quebec but not in any particular municipality (census subdivision or CSD); rather, they were distributed among many municipalities.
4. Language is only one of several variables used in the donor imputation process. Depending on which variables are being imputed, the donor selection variables can include age, sex, geography, variables from linked administrative data (e.g., place of birth from linked immigration administrative records), other variables (e.g., immigrant status), and language. A recipient's characteristics are matched as closely as possible to a donor's characteristics across all of these dimensions. Because many dimensions are used, the importance of an error in any one dimension is significantly reduced.
5. For most variables, language is primarily a useful donor selection variable when it differentiates between immigrant languages, Aboriginal languages and official languages. Because the language allocation error was specific to miscoding between the two official languages, the importance of the error is smaller than if it had affected immigrant or Aboriginal languages.
6. Donors that are a close match on other characteristics will also tend to have similar or equal information for the variables being imputed, reducing the impact of using a different donor.

Table 1

Number of affected records imputed and their share of the total population, selected variables for which language variables were used to select donors during imputation, Canada and Quebec

Variable	Question	Number of affected records imputed	Percentage of total population (%)
Canada			
Income	Linked administrative data	1,940	0.01
Place of birth	Q12	340	0.00
Citizenship	Q13	300	0.00
Immigrant status	Q14	250	0.00
Year of immigration	Q15	410	0.00
Admission category	Linked administrative data	160	0.00
Ethnic origin	Q17	2,220	0.03
Aboriginal group	Q18	610	0.01
Visible minority	Q19	1,000	0.01
Registered or Treaty Indian status	Q20	620	0.01
Membership in a First Nation or Indian band	Q21	790	0.01
Mobility status, one year	Q22	910	0.01
Mobility status, five years	Q23	980	0.01
Place of birth of father	Q24a	630	0.01
Place of birth of mother	Q24b	690	0.01
Language of work	Q45	710	0.01
Quebec			
Income	Linked administrative data	1,760	0.02
Place of birth	Q12	320	0.02
Citizenship	Q13	280	0.01
Immigrant status	Q14	230	0.01
Year of immigration	Q15	370	0.02
Admission category	Linked administrative data	150	0.01
Ethnic origin	Q17	2,070	0.11
Aboriginal group	Q18	580	0.03
Visible minority	Q19	930	0.05
Registered or Treaty Indian status	Q20	560	0.03
Membership in a First Nation or Indian band	Q21	740	0.04
Mobility status, one year	Q22	840	0.04
Mobility status, five years	Q23	910	0.05
Place of birth of father	Q24a	580	0.03
Place of birth of mother	Q24b	640	0.03
Language of work	Q45	650	0.03

Note: The number of records imputed is unweighted and rounded. The percentages of the total population were calculated based on weighted estimates for all variables except income.

Source: Statistics Canada, Census of Population, 2016.

The analysis involved re-running certain steps of the imputation process, first, to assess the amount of variability in estimates arising from the imputation process itself, and second, to assess the size of the change arising from re-imputing the cases affected by the language allocation errors. If the size of the change arising from re-imputing the cases affected by the language allocation error was small relative to the variability in estimates arising from the imputation process itself, then it could be concluded that the impact of the language allocation error on the estimates was negligible.

Table 2 illustrates the variability introduced to the estimates by the imputation system overall.⁴ The table focusses on results from Quebec, where most of the affected cases were found. It focusses only on variables for which language variables were used to select donors during imputation. It shows results averaged across CSDs according to size, for selected income and ethnocultural indicators (median total income, median wages, proportion of the population with Aboriginal identity, proportion of visible minorities, proportion of immigrants, and proportion with British Isles ethnic origins⁵).⁶ When the data are re-imputed for all CSDs, these estimates may rise or fall. For example, for a CSD in the population range of 20,000 to 99,999, when the total income value increased, it increased by an average of \$16, and when it decreased it decreased by an average of \$15. This is an illustration of the small variability in the estimate that derives from donor imputation.

Table 3 shows what would be the effect of correcting only those records whose language allocation changed.⁷ For CSDs in this size class, in cases where the income estimate rose, it rose by \$3 on average; and in cases where it fell, it fell by \$4 on average. Thus, the change in estimates resulting from correcting the language error and re-imputing the results is small and falls within the variability inherent in the imputation process, and therefore has a negligible impact on the results. This was also true for different CSD size classes and for other variables which used language variables to select donors during imputation.⁸

The analysis of the potential impact of the language allocation error on estimates for the other variables from the 2016 Census concluded that the error had a very negligible impact. This was to be expected, given that the vast majority of records received responses directly from respondents or linked administrative records, and very few respondents had estimates that were affected by the language allocation error.

However, in some cases, the mother tongue, home language, or knowledge of official language variables were used to deterministically modify inconsistent responses related to the language of work and knowledge of non-official languages variables (e.g., if an individual reports that they speak Spanish most often at home, Spanish must also be one of the responses provided for the knowledge of non-official languages). Because of this direct dependence, the language of work and knowledge of non-official languages variables were re-processed for the 2016 Census using the corrected the mother tongue, home language, and knowledge of official language variables.

Based upon these results, there was no need to re-process any non-language variables for the 2016 Census.

4. The principle is to compare the data following two runs of the imputation process, with the difference in estimates being the result of imputation. Values from two imputations are shown for context.

5. Includes all respondents who reported as a specified ethnic origin, either as their only ethnic origin or in addition to one or more other ethnic origins, at least one of: Channel Islander, Cornish, English, Manx, Scottish, Welsh or British Isles origins, n.i.e.

6. For CSDs with population 250 or greater. For income variables, the analysis is restricted to the population living in private dwellings aged 15 years and older. For the ethnocultural variables, the analysis is restricted to the population living in private dwellings and to persons not in census families, lone parents, and persons in couples (with or without children). Note that not all imputation steps were re-done – only those necessary to conduct this experiment.

7. Multiple iterations of the imputation process were run to estimate the expected value for the amount being imputed. This makes it possible to examine the direct effect of correcting the error, net of the variability in imputation.

8. A separate experiment was run for the low-income rate. In this case it was only possible to use the value from one donor. Results were similar and are not reported.

Table 2

Comparing estimates generated through two imputation runs, selected variables for which language variables were used to select donors during imputation, census subdivisions (CSDs), Quebec

CSD population size	Values from one imputation	Values from another imputation	Average imputation effect	Average of positive effects	Average of negative effects
Median total income (\$)					
250 to 9,999	30,128	30,129	-3	69	-67
10,000 to 19,999	36,958	36,952	-10	20	-22
20,000 to 99,999	36,478	36,477	-5	16	-15
100,000+	34,469	34,468	2	13	-8
Median wages and salaries (\$)					
250 to 9,999	28,534	28,530	2	97	-96
10,000 to 19,999	35,725	35,722	-1	29	-32
20,000 to 99,999	35,606	35,607	-2	12	-16
100,000+	33,957	33,955	1	6	-5
Proportion with Aboriginal identity (%)					
250 to 9,999	7.0	7.0	0.00	0.20	-0.30
10,000 to 19,999	2.1	2.1	0.00	0.04	-0.06
20,000 to 99,999	1.3	1.3	0.00	0.02	-0.01
100,000+	1.6	1.6	0.00	0.01	0.00
Proportion of visible minorities (%)					
250 to 9,999	0.8	0.8	0.00	0.15	-0.18
10,000 to 19,999	4.5	4.5	0.00	0.01	-0.01
20,000 to 99,999	6.2	6.2	0.00	0.01	-0.01
100,000+	10.3	10.3	0.00	0.00	0.00
Proportion of immigrants (%)					
250 to 9,999	2.6	2.6	0.00	0.15	-0.23
10,000 to 19,999	9.0	9.0	0.00	0.00	-0.01
20,000 to 99,999	10.5	10.5	0.00	0.01	-0.01
100,000+	14.5	14.5	0.00	0.00	0.00
Proportion with British Isles ethnic origins (%)					
250 to 9,999	8.4	8.4	0.00	0.39	-0.40
10,000 to 19,999	11.7	11.7	0.02	0.09	-0.09
20,000 to 99,999	8.7	8.7	0.01	0.07	-0.05
100,000+	8.0	8.0	0.00	0.02	-0.01

Note: The analysis is on CSDs with a population of 250 or more persons. Cells represent the average value among CSDs in each size range. Population in private households only. For income variables, the analysis is restricted to the population living in private dwellings aged 15 years and older. For the ethnocultural variables, the analysis is restricted to the population living in private dwellings and to persons not in census families, lone parents, and persons in couples (with or without children).

Source: Statistics Canada, Census of Population, 2016, experimental analysis.

Table 3

Comparing expected estimates before and after re-imputing records with a language allocation error, selected variables for which language variables were used to select donors during imputation, census subdivisions (CSDs), Quebec

CSD population size	Expected value (before correcting language error)	Expected value (after correcting language error)	Average change	Average of positive changes	Average of negative changes
Median total income (\$)					
250 to 9,999	30,133	30,129	-3	24	-31
10,000 to 19,999	36,957	36,955	-2	3	-6
20,000 to 99,999	36,477	36,477	0	3	-4
100,000+	34,470	34,469	0	2	-1
Median wages and salaries (\$)					
250 to 9,999	28,535	28,536	1	44	-41
10,000 to 19,999	35,726	35,726	0	8	-8
20,000 to 99,999	35,609	35,607	-2	5	-6
100,000+	33,955	33,956	0	3	-2
Proportion with Aboriginal identity (%)					
250 to 9,999	7.0	7.0	0.00	0.17	-0.07
10,000 to 19,999	2.1	2.1	0.00	0.01	0.00
20,000 to 99,999	1.3	1.3	0.00	0.01	-0.01
100,000+	1.6	1.6	0.00	0.00	0.00
Proportion of visible minorities (%)					
250 to 9,999	0.8	0.8	0.00	0.08	0.00
10,000 to 19,999	4.5	4.5	0.00	0.00	-0.01
20,000 to 99,999	6.2	6.2	0.00	0.00	0.00
100,000+	10.3	10.3	0.00	0.00	0.00
Proportion of immigrants (%)					
250 to 9,999	2.6	2.6	0.00	0.13	-0.06
10,000 to 19,999	9.0	9.0	0.00	0.00	-0.01
20,000 to 99,999	10.5	10.5	0.00	0.00	0.00
100,000+	14.5	14.5	0.00	0.00	0.00
Proportion with British Isles ethnic origins (%)					
250 to 9,999	8.4	8.4	-0.02	0.25	-0.30
10,000 to 19,999	11.7	11.7	-0.02	0.02	-0.06
20,000 to 99,999	8.7	8.7	-0.01	0.02	-0.02
100,000+	8.0	8.0	-0.01	0.01	-0.01

Note: The analysis is on CSDs with a population of 250 or more persons. Cells represent the average value among CSDs in each size range. Population in private households only. For income variables, the analysis is restricted to the population living in private dwellings aged 15 years and older. For the ethnocultural variables, the analysis is restricted to the population living in private dwellings and to persons not in census families, lone parents, and persons in couples (with or without children).

Source: Statistics Canada, Census of Population, 2016, experimental analysis.

Evaluation of the impact on the weighting of the 2016 Census long-form questionnaire sample

The purpose of the weighting for the census long-form questionnaire sample is to be able to extrapolate the results of the sample to the entire population. The weighting is based on several characteristics that were collected for the entire population. More specifically, the weight of households in the long-form sample is calculated so that the sum of these household weights corresponded to the population counts for several characteristics. This technique is called weight calibration.

Weight calibration establishes consistency and decreases the variance among the estimates for the long-form questionnaire characteristics correlated with the calibration characteristics. Weight calibration is done independently for geographic areas called weighting areas.⁹ Up to 200 population characteristics can be used to perform weight calibration within a weighting area. On average, approximately 100 are used.

Among the characteristics affected by the language data error, only the “knowledge of official languages” characteristic (i.e., English only, French only, English and French, Neither English nor French) was used for calibration. Corrections to the language data were made for approximately 61,000 people. However, the “knowledge of official languages” data did not require corrections for all of these individuals. Corrections to the “knowledge of official languages” data were made for a little over 36,000 people, the vast majority (95%) of whom went from the category “English and French” to “French only.” Since these corrections concerned mostly people with a knowledge of French, they were more significant in certain geographic areas of Quebec.

Estimates from the weighting areas most affected by these corrections were evaluated. The original weights were calculated based on the “knowledge of official languages” data before the corrections. The evaluation determined whether these weights could be used to produce accurate long-form characteristic estimates in terms of bias and sampling variance. The impact on the estimates is more easily detected for variables correlated with “knowledge of official languages,” such as sociocultural variables or other variables involving languages that were not used in calibration.

“Mother tongue” data (i.e., English, French or another language), which were available for the entire population, were used as a test variable to study the impact on the accuracy of estimates. This variable is correlated with “knowledge of official languages,” and was a variable that required correction. Estimates for the four weighting areas most affected by the corrections were produced for the two “mother tongue” categories most affected (i.e., French and English), after correcting this characteristic. Estimates were produced first using the original weights, then using the weights calibrated to the corrected “knowledge of official languages” data, as well as to the other characteristics used to calculate the original weights.¹⁰

To study the impact on the bias of estimates, t-statistics (i.e., the difference between the total estimate and the population count, divided by the standard error of the total estimate) were calculated using the two weight sets. Then, a t-statistic was calculated for the difference between the two estimates (i.e., the estimate with the original weight minus the estimate with the corrected weight, divided by the standard error of this difference). Table 4 shows all the t-statistics.

9. These weighting areas are defined by the aggregate dissemination areas (ADA) and super aggregate dissemination areas (SADA). See the section “Weighting the census long-form questionnaire sample” in [Chapter 9](#).

10. Producing corrected weights requires much time, so corrected weights were produced for selected weighting areas for evaluation purposes only.

Table 4

Impact on the bias: T-statistics for certain mother tongue characteristics using the original and corrected weights for the four weighting areas most affected by the corrections

Mother tongue	Predominant city in the weighting area	T-statistic		
		Difference between the estimate with original weights and population count	Difference between the estimate with corrected weights and population count	Difference between the two estimates
French	Coaticook	1.92	1.75	1.62
English	Coaticook	-0.51	-0.31	-1.49
French	Rivière-du-Loup	1.15	1.02	0.93
English	Rivière-du-Loup	-0.87	-0.72	-0.99
French	Trois-Rivières	-0.79	-0.80	0.24
English	Trois-Rivières	-0.41	-0.37	-0.63
French	Drummondville	1.55	1.64	-1.19
English	Drummondville	-0.60	-0.70	1.30

Source: Statistics Canada, Census of Population, 2016, experimental analysis.

The absolute values of the t-statistics on the estimates based on the original weights and of those based on the corrected weights are less than 2. This is an indication that the estimates are not biased, regardless of the weight set used. The absolute values of the t-statistics based on the differences between the two estimates are also less than 2. This indicates that the two estimates are not significantly different.

To evaluate the impact on the sampling variance of estimates, the standard errors were calculated based on the two weighting scenarios. The results are shown in Table 5.

Table 5

Impact on sampling variance: Standard errors for certain mother tongue characteristics using the original and corrected weights for the four weighting areas most affected by the corrections

Mother tongue	Predominant city in the weighting area	Standard errors	
		Estimate with original weights	Estimate with corrected weights
French	Coaticook	166	163
English	Coaticook	140	137
French	Rivière-du-Loup	87	85
English	Rivière-du-Loup	60	59
French	Trois-Rivières	137	137
English	Trois-Rivières	92	91
French	Drummondville	101	101
English	Drummondville	72	72

Source: Statistics Canada, Census of Population, 2016, experimental analysis.

The standard errors calculated with the corrected weights are very similar to the standard errors calculated with the original weights. This indicates that using the original weights would not negatively affect the sampling variance of the estimates.

Mother tongue estimates were calculated after the data were corrected for the weighting areas most affected. The evaluation of these calculations shows that using the original weights does not change the accuracy of the estimates in terms of bias and sampling variance. In the geographic areas with fewer corrections, the impact of the corrections should be even less noticeable. Like the other characteristics used in calibration, “knowledge of official languages” was not systematically chosen in all the weighting areas. Based on the evaluation results above, correcting the weights is not necessary to ensure the accuracy of the estimates produced with the long-form questionnaire sample.

Appendix 1.9 – Standard errors of estimates from 2016 Census long form sample

This note analyzes the standard errors for different levels of geography for estimates of totals and averages for some variables common to the 2016, 2011 and 2006 censuses. It provides context for the measure of variability released for the 2016 Census long form, which is the standard error (SE). The note starts with some highlights of the SEs for the 2016 Census long form, 2011 National Household Survey (NHS) and the 2006 Census long form for Canada, the provinces and territories and four Census subdivisions (CSDs). Following these highlights, an analysis of the SEs for 2006, 2011 and 2016 is provided. Finally, in [the table](http://www12.statcan.gc.ca/census-recensement/2016/ref/se-et-eng.cfm) (<http://www12.statcan.gc.ca/census-recensement/2016/ref/se-et-eng.cfm>), some 2016 SEs are compared with the corresponding 2011 and 2006 SEs for nine characteristics.

Highlights

The following observations apply only to the variables and geographic areas presented in [the table](http://www12.statcan.gc.ca/census-recensement/2016/ref/se-et-eng.cfm) (<http://www12.statcan.gc.ca/census-recensement/2016/ref/se-et-eng.cfm>), i.e., the results do not necessarily extend to other variables and geographic areas.

- The 2016 SEs are considerably lower than the corresponding 2006 and 2011 SEs for estimates of counts in most of the geographic areas presented in [the table](http://www12.statcan.gc.ca/census-recensement/2016/ref/se-et-eng.cfm) (<http://www12.statcan.gc.ca/census-recensement/2016/ref/se-et-eng.cfm>).
- About half of the geographic areas in [the table](http://www12.statcan.gc.ca/census-recensement/2016/ref/se-et-eng.cfm) (<http://www12.statcan.gc.ca/census-recensement/2016/ref/se-et-eng.cfm>) have lower SEs in 2016 compared to 2006 and 2011 for all population characteristics. The other geographical areas have some population characteristics with higher SEs for 2016 compared to 2006 and 2011.
- SEs are much improved in 2016 compared to 2006 and 2011 largely due to improved weighting methodology.

Analysis of the 2006 Census long form, 2011 NHS and 2016 Census long form standard errors

Standard errors are provided for nine different characteristics of the 2006 Census long form, 2011 NHS and the 2016 Census long form. Levels of geography provided are national, provincial and territorial and census subdivision (CSD). Results for only four CSDs are presented. These CSDs are the same as those shown in a [previous note](http://www12.statcan.gc.ca/nhs-enm/2011/ref/cv/index.cfm?Lang=E) (<http://www12.statcan.gc.ca/nhs-enm/2011/ref/cv/index.cfm?Lang=E>) disseminated in 2013 about the precision of the 2011 NHS estimates. The same four CSDs were chosen for consistency purposes. These four geographic areas were initially selected to feature CSDs of varying regions and population sizes.

It should be noted that the SEs for Nunavut were nil in 2006, but greater than nil in 2011 and 2016. Since all households were included in the sample, there was no sampling variance in this territory. However, there was a small non-response variance, which was not measured in 2006 but was measured in 2011 and 2016. The following analysis will therefore not include Nunavut for 2006.

For the characteristics and geographic areas presented in [the table](http://www12.statcan.gc.ca/census-recensement/2016/ref/se-et-eng.cfm) (<http://www12.statcan.gc.ca/census-recensement/2016/ref/se-et-eng.cfm>), the 2016 SEs are almost always lower than the corresponding 2006 and 2011 SEs. It is also observed that the 2016 SEs are lower for more characteristics when compared to 2011 than when compared to 2006.

There are two types of estimates in [the table](http://www12.statcan.gc.ca/census-recensement/2016/ref/se-et-eng.cfm) (<http://www12.statcan.gc.ca/census-recensement/2016/ref/se-et-eng.cfm>): estimates of counts (i.e. people) and estimates of averages (i.e. income). One would expect to see, as the population size increases, the SE to increase for the estimates of counts and to decrease for estimates of averages. As expected, it can be observed that, at the Canada level, the SEs are higher for estimates of counts and lower for estimates of averages, than for estimates at the CSD level, due to the larger population size. Despite the fact that the estimates of counts at the Canada level have a higher SE, it does not mean that the quality of these estimates is lower than that of estimates at the CSD level.

For the six estimates of counts, the SEs in 2016 are roughly half of those for 2011 and 2006 for almost all of the geographic areas, which is a considerable reduction. Two exceptions to this reduction are the territory of Nunavut and the CSD of Joliette (Ville). In both geographic areas, when accounting for population size increases over time, the SEs of the six estimates of counts in 2016 are generally smaller or at least comparable to 2011 and 2006. For the three estimates of averages, the 2016 SEs are similar to 2011 and 2006 in large geographic areas (e.g. Canada, Ontario and Quebec) and either smaller or larger than those of 2011 and 2006 in smaller geographic areas.

About half of the geographic areas in [the table](http://www12.statcan.gc.ca/census-recensement/2016/ref/se-et-eng.cfm) (<http://www12.statcan.gc.ca/census-recensement/2016/ref/se-et-eng.cfm>) have lower SEs in 2016 compared to 2006 and 2011 for all characteristics. The other geographical areas have some characteristics with higher SEs for 2016 compared to 2006 and 2011. At the Canada level, in Prince Edward Island, New Brunswick, Ontario, Saskatchewan, British Columbia, Niagara Falls (City) and Port Coquitlam (City) the SEs are lower in 2016 than in either 2006 or 2011 for all characteristics. For estimates of counts, only four characteristics (non-immigrants, immigrants, visible minority and unemployed) have SEs lower in either 2006 or 2011 than in 2016. This only occurs in Alberta, Northwest Territories, Nunavut, Charlottetown (City) and Joliette (Ville). For estimates of averages, the 2006 and 2011 SEs are at times lower than the 2016 SEs estimates in Newfoundland and Labrador, Nova Scotia, Quebec, Manitoba, Alberta, Yukon, Northwest Territories, Charlottetown (City) and Joliette (Ville).

It was expected that the SEs would be, in general, lower for 2016 than for 2011 due to the simpler sample design, higher response rate and improved calibration procedures. However, they were also much lower than for 2006, especially for estimates of counts. Factors that may explain why the 2016 SEs are considerably smaller than those for 2006 include a larger number of responses due to a higher sampling fraction (i.e. 25% in 2016 versus 20% in 2006) and changes in weight calibration.

Calibration may either reduce or increase the variability of the estimates. On the one hand, variability may be reduced by calibration for long form sample variables that are correlated to calibration variables. On the other hand, when a very large number of calibration totals are used, the variability of the estimates may increase. The fact that, overall, considerably more calibration was performed in 2006 than in 2016 could, to some extent, explain why SEs were usually larger in 2006.