Catalogue no. 98-306-X2016001
ISBN 978-0-660-26769-2

Census of Population Technical Report

Sampling and Weighting Technical Report

Census of Population, 2016

Release date: September 11, 2018

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 *Statistics Act*
[E] use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

An HTML version is also available.

Cette publication est aussi disponible en français.

Table of contents

Introduction

The 2016 Census Program enumerates Canadian households using two main types of questionnaires: the short-form questionnaire and the long-form questionnaire. In 2016, a sample of 25% of Canadian households received a long-form questionnaire, which included the questions from the short-form questionnaire. The other households received the short-form questionnaire. In 2011, the government decided to eliminate the mandatory character of the long-form census questionnaire. However, it was reintroduced for the 2016 Census.

In addition to the short-form questions, the long-form questionnaire includes a series of questions to paint a full portrait of the Canadian population and households, according to their demographic, social and economic characteristics.

The estimates produced from responses to the questions on both questionnaires are obtained from the entire population via a census. All households that respond to both types of questionnaires contribute to a specific number, such as the population figure for a specific age group.

The estimates produced from responses to at least one question found only on the long-form questionnaire are obtained from the sample. In those cases, only respondent households in the long-form sample contribute to the estimate. The unemployment rate and the highest level of educational attainment are examples of this type of estimate.

The long-form sample is evenly distributed geographically to ensure a high degree of reliability of the estimates for all areas of the country and to grant the same degree of importance to all geographic units of a given size.

This technical report presents the methodology used to produce the estimates based on the 2016 Census of Population long-form sample. Chapter 1 details the collection methods used for the census and for the long-form sample. Chapter 2 describes how the sampling was applied for the long-form questionnaire. Chapter 3 explains the data processing procedures. Chapter 4 gives an overview of the procedures used to assign weights to the respondent units in the long-form sample to obtain estimates for the population. Chapter 5 covers different evaluations of the weighting procedures, while Chapter 6 provides an overview of the variance estimation methodology used for the 2016 Census long-form sample. A conclusion is presented in Chapter 7.

1. Census data collection

The purpose of data collection for the 2016 Census was to ensure that each of the 15.4 million dwellings in Canada was enumerated and that, for each occupied dwelling, the corresponding household completed a census questionnaire. The census enumerated the entire population of Canada, which consists of Canadian citizens (by birth and by naturalization), landed immigrants and non-permanent residents, as well as the family members living with them. Non-permanent residents are persons living in Canada who have a work or study permit or are claiming refugee status, as well as the family members living with them.

The census also enumerated Canadian citizens and landed immigrants who were temporarily out of the country on Census Day. This included federal, provincial and territorial government employees posted outside Canada; the staff of Canadian embassies abroad; members of the Canadian Forces stationed abroad; and all Canadians aboard merchant vessels.

1.1 Collection methodology

Since the 2011 Census, a new collection methodology has been used to collect census data. Referred to as the wave methodology, it involves contacting non-respondent households at key times to remind them to take part in the census and to encourage them to complete the questionnaire. In each wave, households are given the information they need to respond. Since every Canadian household is required by law to answer the census questions, this method is designed to encourage people to respond promptly online, while mitigating the risk of a decline in overall response and reducing the need for costly field follow-up.

This methodology varies with the collection method used to distribute the census materials for a given region.[1] These collection methods are described in the next section. In 2016, Canadian households had the option of responding online, completing a paper questionnaire (mail-back) or contacting the Census Help Line.

1.1.1 Collection methods: Delivery of census questionnaires

The three collection methods used for the Canadian census are mail-out, list/leave and canvasser. To make census collection as efficient as possible, Canada's territory is divided into small geographic units known as collection units (CUs). For the 2016 Census, Canada was divided into approximately 46,000 CUs. There are five types of CUs, and one of three collection methods is assigned to each one. The same collection method is used for all dwellings in a CU.

Table 1.1.1.1
Collection method by type of collection unit

Type of CU	Collection method
Mail-out	Mail-out
List/leave	List/leave
Canvasser	Canvasser
Collective dwellings	List/leave
Reserves	Canvasser

CU = Collection unit.

1. Data collection for the Wood Buffalo census subdivision was exempt from this methodology. On May 1, 2016, a forest fire broke out southwest of Fort McMurray, Alberta. Within two days (by May 3), it had ravaged the area, destroying many homes and buildings. Statistics Canada introduced an alternative strategy for the Wood Buffalo census subdivision, which encompasses Fort McMurray, because of the evacuation of residents during the collection period. For more information on the collection method used for the Wood Buffalo census subdivision, see Appendix D.

The collection methods are described below.

1.1.1.1 Mail-out

In CUs where the mail-out collection method is used, the postal system is used to deliver census materials. This delivery method ensures effective, coordinated distribution, without needing to recruit and train a large contingent of enumerators. Mail-out CUs are typically in urban areas. While mail-out CUs now include about 82% of Canadian dwellings, they cover only a tiny fraction of the country's territory. One-quarter of the private dwellings in these CUs are selected to receive the long-form questionnaire; the other dwellings receive the short-form questionnaire.

1.1.1.2 List/leave

List/leave CUs are typically in rural areas. In those areas, enumerators prepare a list of dwellings and deliver the census materials. About 17% of the dwellings in Canada are in list/leave CUs, which cover a large portion of the country's territory. One-quarter of the private dwellings in these CUs are selected to receive the long-form questionnaire. This collection method is also used for collective CUs.

1.1.1.3 Canvasser

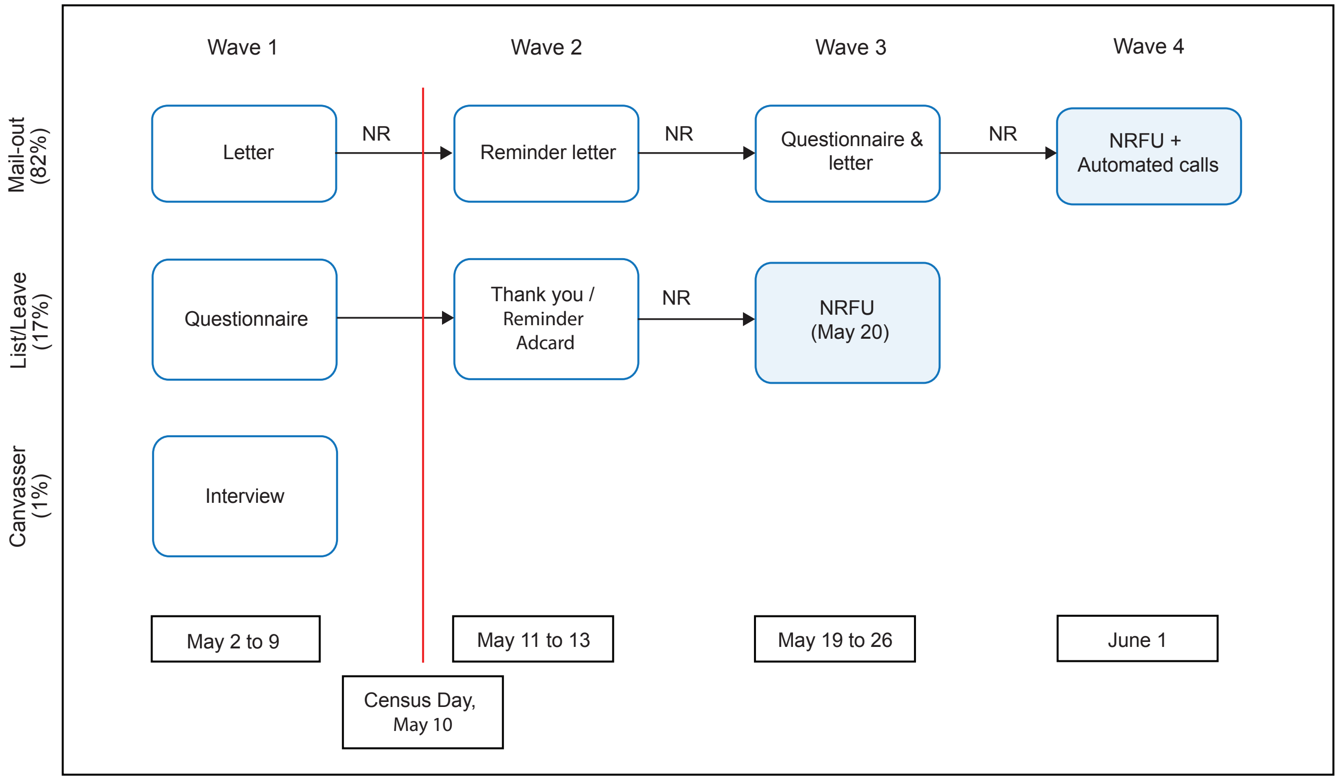
Canvasser CUs are either Indian reserves or in remote or difficult-to-access places. To limit the number of often costly and logistically complicated trips that enumerators have to make to these places for follow-up, the enumerators do more than prepare dwelling lists. They also complete a questionnaire with each household on the spot. Canvasser CUs cover just over half of Canada's territory, but only about 1% of dwellings. All private dwellings in these CUs receive the long-form questionnaire.

1.1.2 Census wave methodology

The wave methodology was designed to encourage online response and to offer an alternative for households that do not wish to complete their questionnaire online. This wave approach has many advantages for minimizing non-response and increasing the number of questionnaires completed online, which accelerates questionnaire registration, improves question flow and data capture, and ultimately enhances data quality.

The wave methodology is applied differently depending on the collection method. CUs were therefore consolidated by collection method, and a different wave methodology was developed for each of the three groups. Given the nature of canvasser CUs, a single-wave methodology was adopted for these units. The sections that follow and Figure 1.1.2.1 provide an overview of the wave methodology used for the 2016 Census.

Figure 1.1.2.1
Overview of the wave methodology in the 2016 Census



NR = non-response
NRFU = non-response follow-up
Source: Statistics Canada, Census of Population, 2016.

1.1.2.1 Mail-out collection units

In mail-out CUs, Wave 1 involved simply sending out a letter asking households to complete the questionnaire online using the secure access code (SAC) provided or to call an automated system on a toll-free line to have a paper questionnaire mailed to them. The Wave 1 letters were delivered by mail one week before Census Day (i.e., on May 2, 2016).

Wave 2 consisted of a reminder letter sent to all Wave 1 households who had not responded by a certain date, by region. The letter reminded the households that they were required by law to complete the census questionnaire. As the Wave 1 letter did, it also provided the SAC and the toll-free telephone number. It was delivered to households from May 11 to 13.

In Wave 3, from May 19 to 26, a paper questionnaire was sent to non-respondent households. The households could still respond online using the SAC printed on the front cover of the questionnaire. The questionnaire was accompanied by a letter indicating that if the questionnaire was not completed by May 31, 2016, an enumerator would contact the household by telephone or in person to complete the questionnaire. The letter also mentioned that if the household refused to answer the census questions, the case could be referred to the Public Prosecution Service of Canada, which would take appropriate action under the *Statistics Act*.

Wave 4, which began on June 1, 2016, consisted of field non-response follow-up (NRFU) and an automated reminder call at the beginning of the follow-up period. NRFU is described in section 1.1.3 of this document.

1.1.2.2 List/leave collection units

A different wave methodology was used for list/leave CUs. In Wave 1, enumerators delivered a paper questionnaire to all the dwellings in these CUs from May 2 to 9. The questionnaire also provided a SAC, so that the household had the option of responding online.

In Wave 2, from May 11 to 13, all the dwellings in these CUs received a thank you and reminder card in the mail. These were delivered whether they had responded or not, because sending mail to specific civic addresses without the name of the occupant was generally impossible in these areas.

The last wave involved going directly to field non-response follow-up as of May 20.

1.1.3 Non-response follow-up

As mentioned in the previous section, the last wave in the wave methodology was NRFU. In that wave, enumerators telephoned and visited households that had not responded to the questionnaire. Each non-respondent household for which a telephone number was available received an automated reminder call at the beginning of the NRFU period. This message reminded non-respondents of their legal obligation to respond to the census questionnaire.

NRFU was conducted in person by enumerators or over the telephone by call centre staff. Census employees had information at their disposal to manage their work in the Collection Management Portal. Being a computerized system accessible over the Internet, this portal also facilitated the gathering of information on collection progress and costs.

1.1.4 Dwelling occupancy verification

Before NRFU, field operations were also carried out for dwelling occupancy verification (DOV) of non-respondent dwellings. The purpose of DOV, which began shortly after Census Day, May 10, 2016, was to identify the greatest possible number of dwellings that were unoccupied on Census Day or cancelled (addresses that were not private or collective dwellings) before NRFU started. Identifying such dwellings close to Census Day should make occupancy classification more accurate and easier to perform. DOV also reduces the NRFU workload, since the unoccupied or cancelled dwellings it identifies do not require follow-up.

Nevertheless, errors in classifying a dwelling as occupied or unoccupied do occur during DOV and NRFU. Some dwellings classified as unoccupied are in fact occupied, and some non-respondent dwellings are unoccupied. As a result, another process, the Dwelling Classification Survey (DCS), is carried out after NRFU. It determines the occupancy status of a sample of dwellings for which no completed questionnaire was received (unoccupied dwellings, non-respondent dwellings or unresolved cases), and it estimates occupancy rates for all dwellings in the same situation across the country. The survey results are used to adjust the Census of Population counts during processing. More information on the DCS is provided in section 3.6.

1.2 Questionnaires and forms

Various questionnaires and forms are used to collect data for the Canadian census, and respondents have the choice of completing the census questionnaire online or using a paper questionnaire.

Visitation Record (VR) is used to list every occupied and unoccupied private or collective dwelling, agricultural operation and agricultural operator in the areas of the country where a list of dwellings is created before questionnaires are delivered (i.e., in list/leave collection units; see section 1.1.1.2). The VR serves as an address list for field operation and control purposes during data collection.

Form 2A (http://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=getInstrumentList&Item_Id=295241&UL=1V&) is the basic short-form questionnaire used to enumerate private dwellings. Each household is asked to list all household members who belong to the census population and to answer questions for them.

Forms 2A-L (http://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=getInstrumentList&Item_Id=295122&UL=1V&) and 2A-R (http://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=getInstrumentList&Item_Id=295299&UL=1V&) are the long-form questionnaire. The questions on form 2A are also on these forms, followed by additional questions on various subjects such as education, employment and dwellings. Form 2A-L is distributed to 25% of Canadian households in mail-out areas and list/leave areas, as described in sections 1.1.1.1 and 1.1.1.2; these are the long-form sample households. Form 2A-R has nearly all the same questions as form 2A-L and is used in areas where canvassing is the collection method. All households in these areas are included in the long-form sample.

Form 2C (http://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=getInstrumentList&Item_Id=295305&UL=1V&) is used to enumerate Canadians posted abroad. Form 2C has the same questions as form 2A.

Form 3A (http://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=getInstrumentList&Item_Id=295255&UL=1V&) is an individual census questionnaire used to enumerate persons in collective dwellings. It can also be used to enumerate usual residents in a private household who prefer to complete their own census questionnaire rather than be included in a form 2A.

1.3 Collection response rate

The overall collection response rate for the 2016 Census of Population was 98.4%. This rate was calculated directly from the collection results (as of August 4, 2016), i.e., before data processing and quality verification were completed. It represents the number of private dwellings for which a questionnaire was returned, divided by the number of private dwellings that enumerators coded as being occupied. The collection response rate for the long-form sample was 97.8% (for more information, see the 2016 Census of Population collection response rates (http://www12.statcan.gc.ca/census-recensement/2016/ref/response-rates-eng.cfm)).

2. Sampling[2]

When a sample survey is conducted, the sample selection must be planned properly. In sampling, a subset of the survey's target population is selected to receive the questionnaire. The responses of the subset are used to draw inferences for the entire population. Two types of sampling exist: probability sampling and non-probability sampling. Probability sampling is preferable when producing statistical inferences for the entire population is important, since the probability of unit selection can be calculated and the sampling error can be estimated. This chapter discusses the selection of the sample that received the 2016 Census long-form questionnaire.

2.1 Long-form sample universe

The census household universe was broken down into three parts: private households, collective households and households outside Canada. The long-form sample universe consists only of private households, including those living in private dwellings attached to collective dwellings in Canada. This universe excludes incompletely enumerated Indian reserves and settlements. Unless otherwise specified, the term "in scope" indicates that a household is part of the long-form sample universe (i.e., private households that are not living in incompletely enumerated Indian reserves and settlements). "Out of scope" refers to households not in the universe (i.e., households living in collective dwellings, outside Canada, or in incompletely enumerated Indian reserves and settlements).

2.2 Long-form sampling design

In most cases, the long-form questionnaire was distributed to one-quarter of the households in the long-form universe to gather demographic and socioeconomic data on the Canadian population. The sample was selected from the list of dwellings for the 2016 Census of Population. At the time the sample was selected, the addresses of out-of-scope dwellings were unknown. This meant that some dwellings erroneously received a long-form questionnaire. Once a dwelling was determined to be out of scope, no further collection or processing activities were carried out.

Dwellings were selected to receive the long-form questionnaire according to a stratified systematic sampling design. The sampling design strata were the CUs. For mail-out CUs, the sampling was systematic, with a one-quarter sampling fraction. The selection starting point was random. For list/leave CUs, sampling was systematic, and the sampled dwellings were every fourth one on the list, i.e., 4th, 8th, 12th, etc. For example, if a list/leave CU had seven dwellings, only one dwelling was selected. Finally, in canvasser CUs, all households were selected. These CUs were take-all strata.

The sampling design had one exception. Private dwellings attached to collective dwellings were added to the sample with certainty. However, they completed only the short-form questionnaire. Long-form questionnaire responses were later imputed for these households.

Except private households attached to collective dwellings, all households selected for the sample were asked to complete the long-form census questionnaire. Households in private dwellings that were not part of the long-form sample were asked to fill out the short-form questionnaire.

2. For more information on the history of sampling in Canadian censuses, see Appendix B.

3. Census data processing

3.1 Introduction

This chapter discusses the processing of all the completed questionnaires (all questionnaire types), which encompasses everything from the receipt of the questionnaires through to the creation of an accurate and complete census database. It describes the steps of questionnaire registration, questionnaire imaging and data capture, editing, error correction, failed edit follow-up, coding, dwelling classification and non-response adjustments, linkage of income data, imputation, weighting, and final response rates.

Automated processes, implemented for the 2016 Census, had to be monitored to ensure that all Canadian residences were enumerated once and only once. The Master Control System (MCS) was built to control and monitor the process flow, from collection to data processing. The MCS held a master list of all the dwellings in Canada, where each dwelling was identified with a unique identifier. This system was updated on an ongoing basis with information about each dwelling's status in the census process flow (e.g., delivered, received or processed). Reports were generated daily by the system and made accessible online to managers to ensure that census operations were efficient and effective.

3.2 Receipt and registration

Responses received through the Internet or help-line telephone interviews were received directly at the Data Operations Centre (DOC), where the receipt of the responses was registered automatically.

Respondents completing paper questionnaires mailed them back to the DOC. Canada Post registered their receipt automatically in multiple locations in Canada (as part of the normal mail flow process) by scanning the barcode on the front of the questionnaire through the transparent portion of the return envelope. The envelopes were then delivered to the DOC throughout each business day. Canada Post would also send files daily listing all census questionnaires received at each regional processing plant, by date of receipt.

The registration of each returned questionnaire was flagged on the MCS at Statistics Canada. A list of all the dwellings for which a questionnaire had not been received was generated daily by the MCS and transmitted to field operations to prevent follow-up on households that had already completed their questionnaire during NRFU.

3.3 Scanning and keying from images

In 2016, all paper census forms (2A, 2C, 2A-L, 2A-R, 3A) were imaged. The following steps were part of the imaging process:

- Document preparation: Mailed-back questionnaires were removed from envelopes and foreign objects (i.e., clips and staples) were detached in preparation for scanning. The questionnaires were batched by form type. Their spine was cut off to separate them into single sheets.
- Scanning: The questionnaires were converted to digital images.
- Automated image quality assessment: An automated system analyzed the images for errors or anomalies. Images failing this process were sent to be reviewed by a document analysis operator.
- Document analysis: At this step, images containing anomalies were presented to an operator for review. The operator could accept the image as is and send it directly to key entry (bypassing automated recognition), or the operator could send the entire questionnaire to be pulled at the check-out step. See below for more details on the key entry and check-out steps.
- Automated recognition: This step attempted to automatically recognize all handwritten responses and marks on the questionnaires.
- Key entry: Operators entered responses that automated recognition could not determine with sufficient accuracy. About 13% of all responses were sent to keying.

- Check-out: Once the questionnaires were processed through all of the above steps, the paper questionnaires were checked out of the system. Check-out is a quality assurance process that ensures that the images and captured data are of sufficient quality that the paper questionnaires require no subsequent processing. Questionnaires that had been flagged as containing errors were pulled at check-out and reprocessed.

3.4 Coverage edits, completion edits and failed edit follow-up

At this stage, a number of automated edits were performed on respondent data. These edits were designed to detect cases where the number of persons counted in the household was incorrect because of an error in collection, a respondent error or a data capture error. Most of these errors occurred on paper questionnaires, including:

- data erroneously entered in the wrong person column
- crossed off data that are captured in error
- data not being provided for every household member listed in the roster at the beginning of the questionnaire.

Errors that can occur both on paper and online include:

- data provided for the same person on more than one questionnaire (e.g., a person completes his or her own 3A questionnaire and is also included on the household 2A questionnaire)
- the receipt of duplicate questionnaires (e.g., a person completes the Internet version and his or her spouse completes the paper version and mails it back).

For about 58% of edit failures, the system resolved the case automatically. This was done when the error was such that the solution was obvious. The solutions included deleting false person data that were created because of respondent or capture error and deleting duplicate responses. The remainder of the edit failure cases were forwarded to processing clerks for resolution. An interactive system enabled the clerks to compare data across questionnaires and examine the images of paper questionnaires to detect data capture or respondent errors. Edit failures were resolved by deleting invalid or duplicate persons or by adding missing persons (i.e., creating blank person records), as necessary and appropriate.

Following the coverage edits, another set of automated edits was run. These edits detected cases where too many questions had missing responses or where data had not been provided for all the usual residents in the household, including cases where missing persons were added by coverage edit clerks. Households that failed these edits were followed up with. An interviewer called the respondent to resolve coverage issues and obtain missing responses, using a computer-assisted telephone interviewing application. For households that responded to the long-form questionnaire, only data missing for the short-form questions were followed up on. The data obtained through this follow-up activity were introduced into the system for subsequent processing steps. If the follow-up was unsuccessful, the data were imputed in the edit and imputation step (see section 3.9).

3.5 Coding

The census questionnaires contained questions for which answers could be checked off a list, as well as questions requiring a written response. Each written response was automatically assigned a numerical code according to Statistics Canada reference files, code sets and standard classifications. Reference files for the automated match process were built using actual responses from past censuses, as well as administrative files. Specially trained coders and subject-matter specialists resolved cases where a code could not be automatically assigned. The following questions required coding on both the long- and short-form questionnaires:

- relationship to Person 1
- home language
- mother tongue.

The following questions required coding for the long-form sample only:

- place of birth
- citizenship
- non-official languages
- ethnic origin
- population group
- First Nation/Indian band
- place of residence one year ago
- place of residence five years ago
- place of birth of parents
- major field of study
- location of study
- industry
- occupation
- place of work
- language of work.

A total of about 69 million write-ins were coded from the 2016 Census questionnaires. Overall, about 85% were coded automatically, although the autocoding rate varied considerably from one question to the next.

3.6 Classification and non-response adjustments for unoccupied and non-response dwellings

The Dwelling Classification Survey (DCS) was used to estimate the rate of enumerator error in classifying private dwellings in mail-out and list/leave census CUs as occupied or unoccupied. This information was used to make adjustments to the census database. The DCS selected a random sample of 1,730 mail-out and list/leave CUs. Enumerators revisited these CUs in June and July 2016 to reassess the occupancy status as of Census Day of each private dwelling for which no response was received. The DCS estimated that 15.0% of the 1,187,392 private dwellings classified as unoccupied were actually occupied and that 36.9% of the 284,966 private dwellings with no response that were classified as occupied or that had an unknown occupancy status were actually unoccupied. Estimates based on the DCS sample were used to adjust the occupancy status for individual dwellings. This resulted in an increase of 2.6% in the number of occupied private dwellings and a decrease of 6.2% in the number of unoccupied dwellings at the Canada level.

After this adjustment of the occupancy status by the DCS, occupied private dwellings with total non-response had the number of usual residents (if not known) and all the responses to the census questions imputed. The responses were borrowed from another responding household within the same CU. This process, called whole household imputation (WHI), imputed 99.9% of the total non-response households. Using a single donor under WHI was more efficient computationally and was less likely to produce implausible results than using several donors as part of the main edit and imputation process. Nevertheless, the other 0.1% of the total non-response households where no donor household was found under the WHI process was imputed as part of the main edit and imputation process.

The WHI process has another component that is separate from the use of the DCS estimates to adjust the census database. The non-DCS areas—CUs that have interviewer-administered census questionnaires (i.e., Indian reserve, canvasser and collective CUs)—require a different imputation strategy. In these areas only, all unoccupied dwellings are assumed to be truly unoccupied and all non-responding dwellings are assumed to be truly occupied. This implies that unoccupied dwellings are assumed to be classified correctly and **no imputations are done**. Private dwellings with an occupancy status classified as unknown are also assumed to be unoccupied. On the other hand, private dwellings with no response that were classified by enumerators as being occupied are all assumed to be occupied, and the geographically nearest neighbour is used as the donor household for these dwellings. No restrictions were placed on the household size for these imputations, as was done in the DCS areas. At the Canada level (for DCS and non-DCS areas), 2.6% of occupied private dwellings were imputed through the WHI process.

More details on the DCS and the WHI process will be available in the *Coverage Technical Report, Census of Population, 2016* (http://www12.statcan.gc.ca/census-recensement/2016/ref/98-303/index-eng.cfm), Statistics Canada Catalogue no. 98-303-X, which will be released in 2019.

3.7 Obtaining income data

For the first time, in 2016, administrative data were the only source of information on income for the Census Program. This not only reduced response burden, but also increased the quality and quantity of the income data available. The information on individuals' income was compiled from administrative data for the entire population aged 15 and older, rather than from a sample, as was done in 2011 and 2006. Regular, recurring taxable and non-taxable income received during the 2015 calendar year was included. One-time receipts, such as lump-sum withdrawals from registered retirement savings plans and other savings plans, lump-sum insurance settlements, lump-sum pension benefits, capital gains or losses, inheritances, and lottery winnings, were excluded.

The information on census respondents could be linked to two types of Canada Revenue Agency (CRA) files, depending on whether respondents were (1) taxfilers, for whom all income information could be extracted from income tax files, including T1 general returns, tax slips and government programs administered by the CRA, or (2) non-filers, for whom the only information available came from tax slips and government programs administered by the CRA. In 2016, the information for 94.8% of the population aged 15 and older in private households was linked to a CRA administrative file. Specifically, the information for 85.2% of the population was linked to a taxfiler's file, and the information for 9.6% of the population was linked to a non-filer's file.

For more information on how income data were obtained, see the *Income Reference Guide, Census of Population, 2016* (http://www12.statcan.gc.ca/census-recensement/2016/ref/guides/004/98-500-x2016004-eng.cfm), catalogue no. 98-500-X2016004.

3.8 Non-response

A non-response status may differ during the collection and processing phases. The main differences arise because the occupancy status can change between collection and processing, and because the household must answer a minimum number of questions to be considered a respondent in the processing phase. Unless otherwise specified, the term "non-response" refers to non-response in the data processing phase. The same applies when response is referred to rather than non-response.

For the 2016 Census long-form questionnaire, two types of households were considered non-respondent:

- households from the sample that answered only the questions common to both types of questionnaires, i.e., only the short-form questions
- households that did not answer any questions.

This refers to total non-response, which is processed differently depending on the collection method and the type of household.

Finally, partial non-response is when the long-form questionnaire is partially completed. This type of response is processed by imputation. An overview of this method is presented in the next section.

3.9 Edit and imputation

The data collected in any survey or census contain some omissions or inconsistencies. For example, a respondent may be unwilling to answer a question, fail to remember the right answer or misunderstand the question. Other errors, such as incorrect coding, can also occur.

The final clean-up of data, done in the edit and imputation process, was fully automated using the Canadian Census Edit and Imputation System (CANCEIS) (Statistics Canada 2014) for all census topics. Two imputation methods were applied. The first method, called “deterministic imputation,” involved assigning specific values under certain conditions when problems were clear and unambiguous to resolve. Detailed edit rules were applied to identify these conditions, and the variables involved in the rules were assigned predetermined values. The second method, called “minimum-change nearest-neighbour donor imputation,” applied a series of detailed edit rules that identified any missing or inconsistent responses. When a record with missing or inconsistent responses was identified, another record that met the edit rules and had most characteristics in common with the record with an error was selected. Data from this donor record were borrowed and used to make the minimum number of changes to the variables to resolve all cases of missing or inconsistent responses.

3.10 Weighting

The 2016 Canadian Census Program consisted of a census of population and a sample survey for which one-quarter of Canadian private households were selected. Households not sampled for the survey received a short-form questionnaire, while sampled households received a long-form questionnaire. In addition to the short-form questions, the long-form questionnaire gathered sociocultural information, as well as information on daily activities, mobility, place of birth, education, labour market activity, etc. Weighting was used to represent the entire population based on the information gathered from the sample.

The first step in the weighting process was to assign a design weight to each household that reflected its probability of being sampled. These weights then underwent an initial adjustment for coverage and total non-response. This adjustment was applied to the weights of respondent households. Finally, a second adjustment, referred to as final calibration, was made to establish closer agreement between the estimates obtained from respondent households in the sample and the census counts for a number of characteristics from the short-form questionnaire or from administrative data sources. The weighting methodology is described in detail in Chapter 4. All private households attached to collective dwellings and all private households in a canvasser CU were selected for the long-form sample and received a design weight of 1. They were then excluded from the coverage and non-response adjustment processes, as well as from the final calibration process.

Long-form sample households with a non-zero weight at the end of the weighting process were the respondent households, along with the households who were assigned a design weight of 1, i.e., private households attached to collective dwellings and all private households in canvasser CUs. These households made up the set of households that contributed to the long-form estimates.

3.11 Final response rates

Table 3.11.1 presents the final response rates for private households in the 2016 Census of Population, for Canada and for each province and territory, followed by non-weighted and weighted response rates for the long-form sample based on the definition of non-response given in section 3.8.

The final response rate is the ratio of the numerator to the denominator, where

- the numerator is the number of private dwellings for which a questionnaire was completed[3]
- the denominator is the number of private dwellings classified as occupied, according to the census database.

The final classification of a dwelling's occupancy status is based on an analysis of the data gathered by field staff, data provided by respondents and the results of a study into the quality of occupancy status in the DCS (see section 3.6). The response rates indicated in Table 3.11.1 differ from the collection response rates, which were previously published and were mentioned in section 1.3, in that they take data processing and dwelling occupancy verification into account in identifying non-respondent households. These response rates are therefore considered final.

Weighted response rates were produced for the long-form sample. They are based on the following weights as the numerator and denominator:

- the numerator is the design weight
- the denominator is the weights calibrated to the census total for the geographic area (which is the number of private households in occupied dwellings according to the census).

3. Non-respondent sampled households in Wood Buffalo for which administrative data were used (see Appendix D), as well as private dwellings attached to collective dwellings, which by default received only a short-form questionnaire, were considered as having completed their questionnaire for the purposes of short-form response rates. Since they did not complete the long-form questionnaire, they were considered non-respondents for the purposes of calculating long-form questionnaire response rates.

Table 3.11.1
Final response rates for private households from the 2016 Census of Population and the long-form sample

Provinces and territories	Response rate – short-form questionnaire	Non-weighted response rate – long-form questionnaire only	Weighted response rate – long-form questionnaire only
	percent		
Canada	97.4	96.1	95.9
Newfoundland and Labrador	97.4	96.0	95.1
Prince Edward Island	97.5	96.4	96.3
Nova Scotia	97.6	96.6	96.1
New Brunswick	97.6	96.7	96.2
Quebec	97.6	96.7	96.6
Ontario	97.6	96.5	96.3
Manitoba	97.4	95.9	95.8
Saskatchewan	96.7	95.6	95.1
Alberta	97.0	95.3	94.8
British Columbia	96.5	94.8	94.6
Yukon	95.8	91.9	92.8
Northwest Territories	93.9	92.7	93.0
Nunavut	92.7	92.6	92.6

Note: All private households and occupied dwellings are included in the calculation of these response rates, without exception.
Source: Statistics Canada, 2016 Census of Population and 2016 Census long-form sample.

4. Estimation from the census long form sample

Any sampling process requires an associated estimation procedure for scaling sample data up to the population level and for ensuring that survey estimates are representative of the population. The choice of an estimation procedure is generally governed by both operational and theoretical constraints. From the operational viewpoint, the procedure must be feasible within the processing system of which it is a part, and from the theoretical viewpoint, the procedure should minimize the statistical error of the estimates it produces.

The estimation procedure produces a set of weights, and the weight for each sample unit corresponds to the number of units in the population that the sample unit represents. These weights are applied to the sample data to produce millions of estimates from the census long-form sample. Estimates are summary measures such as totals, averages, proportions and medians calculated from the sample for various characteristics of interest.

4.1 Considerations in the choice of an estimation procedure

4.1.1 Operational considerations

Mathematically, an estimation procedure can be described by an algebraic formula, or estimator, that shows how the estimate for the population is calculated as a function of the observed sample values and other information from the sample design or external data sources. Most of the time, this estimator is a simple function of weights and of the variable of interest for the responding units. Using a unique set of weights to produce all estimates guarantees a certain level of consistency among the different estimates of the survey.

Therefore, the approach taken for the census long-form sample (and in most sample surveys) was to split the estimation procedure into two steps: (a) the calculation of weights (known as the weighting procedure) and (b) the use of weights to produce estimates, such as the estimation of a particular population count by summing the weights of those persons or households with the characteristic of interest. Most of the mathematical complexity is contained in step (a), which is performed just once. Meanwhile, step (b) is reduced to a simple process, such as summing weights whenever tabulation is required. Since the weight attached to each sample unit is the same for any tabulation involving that unit, consistency between different estimates based on sample data is assured.

4.1.2 Theoretical considerations

For a given sample design and a given estimation procedure, one can, from sampling theory, make a statement about the chances that a certain interval will contain the unknown population value being estimated. A primary criterion in the choice of an estimation procedure is the minimization of the width of such intervals for a given level of confidence so that these statements about the unknown population values are as precise as possible. A common measure of precision for comparing estimation procedures is known as the standard error. Provided that certain conditions are met, intervals of plus or minus two standard errors from the estimate will contain the true population value for approximately 95% of all possible samples.

As well as minimizing standard error, a second objective in the choice of an estimation procedure for the long-form sample is to ensure, as far as possible, that sample estimates for census characteristics are consistent with the corresponding known census values. Fortunately, these two objectives are usually complementary in the sense that sampling error tends to be reduced by ensuring that sample estimates for certain basic characteristics are consistent with the corresponding population figures. However, while this is true in general, forcing long-form sample estimates for census characteristics to be consistent with corresponding census figures for very small subgroups can have a detrimental effect on the standard error of estimates for the sample characteristics themselves. For example, if in several dissemination areas only a few subjects have a given characteristic, such as birth in a certain country, ensuring consistency between the sample estimates and the census counts for that place of birth would unduly increase the standard error for the rest of the characteristics.

In cases where no information about the population being sampled is available other than that collected for sample units and unit non-response has not occurred, the estimation procedure would be restricted to weighting the sample units inversely to their probability of selection. For example, if a unit had a one-in-four chance of selection, then that selected unit would receive a weight of 4. When unit non-response is observed, the weight must be further adjusted according to the estimated probability of response of the unit, for example. In practice, some supplementary knowledge about the population (e.g., its total size and possibly its breakdown by a certain variable—perhaps by province and territory) is often available. Such information can be used to improve the estimation formula so as to produce estimates with a greater chance of being close to the unknown population value. In the case of the census long-form sample, a large amount of very detailed information about the population being sampled is available from the census short-form data at every geographic level. This wealth of population information is used in the coverage, non-response and calibration adjustments to improve the estimates made from the long-form sample.

Nevertheless, the long-form sample estimates for census characteristics cannot be made consistent with all the census counts at every geographic level. Differences between sample estimates and census counts become visible when a cross-tabulation of a sample variable and the corresponding census variable is produced. The tabulation of sample-based estimates of totals for particular characteristics will not necessarily agree with the equivalent census count tabulations for those characteristics.

Adjusting the weights, by the minimal amounts possible, to achieve perfect agreement between long-form estimates and census counts for certain characteristics and subgroups is known as “calibration.”

4.2 Weighting areas

The various adjustments to design weights were made independently by weighting area. The geographic areas used for this purpose were aggregate dissemination areas (ADAs) and super aggregate dissemination areas (SADAs). ADAs are a new dissemination geography created for the 2016 Census. SADAs were created specifically for the weighting procedures by ADA aggregation. The geographic subdivisions used in 2011 and earlier were constructed differently.

4.2.1 Aggregate dissemination areas

ADAs are a new dissemination geography created for the 2016 Census. Their purpose is to enable the dissemination of a greater quantity of data at a detailed geographic level across the country. In total, Canada was divided into 5,386 ADAs, and households were selected for the long-form sample in 5,143 ADAs. Of the 243 ADAs without sampled households, 235 consisted solely of out-of-scope households. The other eight ADAs had only a handful of in-scope households, and none of them were selected.

ADAs satisfy the following delineation criteria:

1. ADAs cover the entire country and, where possible, have a population count of 5,000 to 15,000 (based on the population counts from the previous census).
2. 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.
3. ADAs are based on one of three 2016 Census dissemination geographic areas: dissemination areas (DAs), census subdivisions (CSDs) or census tracts (CTs):
    - Within CMAs and CAs with CTs, adjacent CTs are combined to meet the ADA population criterion.
    - In areas without CTs (areas outside CMAs and the largest CAs) where CSDs have a population of fewer than 15,000, adjacent CSDs are combined to meet the ADA population criterion.
    - In areas without CTs where CSDs have a population of over 15,000, adjacent DAs are combined within these CSDs to meet the ADA population criterion.

4. Every CSD that consists of an Indian reserve and a small number of other areas where the canvasser method is required constitute distinct ADAs.

"For more information about Aggregate Dissemination Areas, 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."

Table 4.2.1.1 shows the degree to which ADAs with households in the long-form sample were properly adjusted to CSDs. The first scenario occurred in most cases, since ADAs were designed above all to respect the boundaries of CTs and CSDs. Scenario 4 is the only one where CSD boundaries were not respected. CTs were not included in the table because they were all in the first scenario except one, which was in scenario 3.

Table 4.2.1.1
Number of census subdivisions within the boundaries of ADAs with households in the long-form sample, 2016 Census

		Census subdivision	
Scenario	**Description**	number	percent
1	The CSD was small enough to be fully contained in an ADA, and this ADA only had complete CSDs. No CSDs in the ADA were part of another ADA.	4,512	92.40
2	The CSD was small enough to be fully contained in an ADA, but another CSD in the same ADA was part of a different ADA.	81	1.66
3	The CSD was large enough to contain full ADAs. No ADAs were also part of another CSD.	261	5.35
4	The CSD was part of two or more ADAs.	29	0.59
Total		**4,883**	**100.00**

CSD = Census subdivision.
ADA = Aggregate dissemination area.
Source: Statistics Canada, 2016 Census long-form sample.

Table 4.2.1.2 shows the distribution of ADAs with households in the long-form sample by province or territory.

Table 4.2.1.2
Number of ADAs with households in the long-form sample, by province or territory

Region	Number of ADAs
Newfoundland and Labrador	81
Prince Edward Island	21
Nova Scotia	146
New Brunswick	124
Quebec	1,118
Ontario	1,655
Manitoba	216
Saskatchewan	256
Alberta	516
British Columbia	916
Yukon	28
Northwest Territories	40
Nunavut	26
Canada	**5,143**

ADA = Aggregate dissemination area.
Source: Statistics Canada, 2016 Census long-form sample.

Table 4.2.1.3 shows the number of ADAs by the number of in-scope households in the census. The majority of ADAs with households in the long-form sample had from 2,000 to 4,999 households. A considerable number of ADAs had small populations.

Table 4.2.1.3
Distribution of ADAs with households in the long-form sample, by number of in-scope households

In-scope households	Number of ADAs	Percent
0 to 499	976	18.98
500 to 999	117	2.27
1,000 to 1,999	366	7.12
2,000 to 2,999	1,339	26.04
3,000 to 3,999	1,229	23.90
4,000 to 4,999	664	12.91
5,000 to 5,999	300	5.83
6,000 to 6,999	98	1.91
7,000 to 7,999	32	0.62
8,000 to 8,999	14	0.27
9,000 to 9,999	3	0.06
10,000+	5	0.10
Total	**5,143**	**100.00**

ADA = Aggregate dissemination area.
Source: Statistics Canada, 2016 Census of Population.

Table 4.2.1.4 presents the number of ADAs by range of numbers of households that responded to the 2016 Census long-form questionnaire. For the ADAs with the fewest respondents, a specific type of processing was applied to have enough households for weighting purposes (see section 4.5). Overall, ADAs had more respondents than the weighting areas used in 2011.

Table 4.2.1.4
Distribution of ADAs with households in the long-form sample, by number of respondent households for the long-form questionnaire

Respondent households	Number of ADAs	Percent
0 to 99	605	11.76
100 to 199	265	5.15
200 to 299	158	3.07
300 to 399	142	2.76
400 to 499	311	6.05
500 to 599	537	10.44
600 to 699	628	12.21
700 to 799	604	11.74
800 to 899	518	10.07
900 to 999	400	7.78
1,000 to 1,099	304	5.91
1,100 to 1,199	229	4.45
1,200 to 1,299	162	3.15
1,300 to 1,399	101	1.96
1,400 to 1,499	73	1.42
1,500+	106	2.06
Total	**5,143**	**100.00**

ADA = Aggregate dissemination area.
Source: Statistics Canada, 2016 Census long-form sample.

4.2.2 Super aggregate dissemination areas

SADAs were created specifically for weighting 2016 Census data, so that certain weighting procedures for which a large number of observations is desirable could be conducted.

SADAs were created according to the following rules (in order of priority):

1. SADAs are created by combining ADAs (mandatory).
2. SADAs respect provincial and territorial borders (mandatory).
3. SADAs have a population of 50,000 to 150,000 persons (except for CDs with a population of 40,000 to 50,000 persons that constitute their own SADA.
4. SADA population excluding persons living in canvasser CUs).
5. SADAs respect the boundaries of CDs.
6. SADAs respect the boundaries of CMAs and CAs.
7. SADAs respect the boundaries of CSDs.
8. SADAs are single contiguous entities.
9. SADA are as compact as possible.

The first two rules were mandatory, and rules 3 to 9 were followed where possible. A total of 409 SADAs were created.

Table 4.2.2.1 shows the distribution of SADAs by province or territory.

Table 4.2.2.1
Number of SADAs, by province or territory

Region	Number of SADAs
Newfoundland and Labrador	8
Prince Edward Island	2
Nova Scotia	13
New Brunswick	8
Quebec	97
Ontario	150
Manitoba	15
Saskatchewan	14
Alberta	44
British Columbia	55
Yukon	1
Northwest Territories	1
Nunavut	1
Canada	**409**

SADA = Super aggregate dissemination area.
Note: In the case of the three territories, the SADA corresponds to the territory.
Source: Statistics Canada, 2016 Census long-form sample.

Table 4.2.2.2 shows the degree to which SADAs were properly adjusted to CDs and CMAs. SADAs respected the boundaries of the majority of CDs (scenarios 1 and 3) and the boundaries of three-quarters of CMAs. The other CMAs were part of at least two SADAs (scenario 4).

Table 4.2.2.2
Number of census divisions and census metropolitan areas within SADA boundaries, 2016 Census

		Census divisions		Census metropolitan areas	
Scenario	**Description**	number	percent	number	percent
1	The CD or CMA was small enough to be fully contained within a SADA, and the SADA included only complete CDs or CMAs. No CDs or CMAs in the SADA were part of another SADA.	249	84.98	2	5.71
2	The CD or CMA was small enough to be fully contained within a SADA, but another CD or CMA in the same SADA was also part of another SADA.	2	0.68	0	0.00
3	The CD or CMA was large enough to contain complete SADAs. No SADAs were also part of another CD or CMA.	40	13.65	25	71.43
4	The CD or CMA was part of two or more SADAs.	2	0.68	8	22.86
Total		**293**	**100.00**	**35**	**100.00**

CD = Census division.
CMA = Census metropolitan area.
SADA = Super aggregate dissemination area.
Source: Statistics Canada, 2016 Census of Population.

Table 4.2.2.3 shows the number of SADAs by the number of in-scope persons.

Table 4.2.2.3
Distribution of SADAs with households in the long-form sample, by number of in-scope individuals

In-scope individuals	Number of SADAs	Percent
30,000 to 39,999	2	0.49
40,000 to 49,999	26	6.36
50,000 to 59,999	23	5.62
60,000 to 69,999	45	11.00
70,000 to 79,999	106	25.92
80,000 to 89,999	67	16.38
90,000 to 99,999	46	11.25
100,000 to 149,999	94	22.98
Total	**409**	**100.00**

SADA = Super aggregate dissemination area.
Source: Statistics Canada, 2016 Census of Population.

4.3 Design weights

The design weight for each household in the long-form sample was calculated differently, depending on the collection method of the CU where the corresponding dwelling was located.

If the method of collection was:

- mail-out, the design weight was equal to the inverse of the survey fraction, giving a weight of 4
- list/leave, the design weight was equal to the ratio of the number of private dwellings enumerated to the number of private dwellings sampled in the CU, and this gave a weight of approximately 4 for 98% of the selected households, although the weights varied from 1 to 7
- canvasser, the design weight was 1.

Households living in private dwellings attached to collective dwellings were an exception to the rule. As mentioned in section 2.2, all of these households were included in the sample. They were considered take-all, so their design weight was 1.

4.3.1 Weights for households counted in the sample

Sampled households with a design weight of 1 did not have their weight adjusted. These households kept their weight of 1 after the weighting procedures were completed (coverage and non-response, as well as calibration to census counts). They either were located in canvasser CUs or were private households that were attached to a collective dwelling.

Total non-response and partial non-response for these households were addressed by imputation. Once the missing data were imputed, these households were considered to be respondents for estimation purposes (although they were considered to be non-respondents for the calculation of response rates in section 3.11).

4.4 Coverage and total non-response adjustment

The several ways of treating non-response in surveys can be divided into two main categories: imputation and reweighting. The former is usually applied for the treatment of item missing values and the latter for the treatment of total non-response. A household was considered to be a respondent to the long form when it answered at least one of the long-form questions. With the high response rate to the long form, any non-response adjustment method would have had, for the most part, only a modest impact on the final survey weights and estimates. Coverage and total non-response for households in Indian reserve and canvasser enumeration CUs were compensated for with imputation procedures and, for the most part, with whole household imputation (WHI) as described in section 3.6. In the rest of the country, reweighting procedures were used. The rest of this chapter describes those weighting procedures.

The main purpose of coverage and non-response adjustments is to minimize the impact of any potential biases from lack of complete coverage (or from duplicates) and from unit non-response. For the adjustment to actually reduce the potential bias, a rich set of information about the non-respondents is very useful. Otherwise, the non-response adjustment that can be applied is limited, and the potential bias will not be greatly lessened. Geographical information was known for every non-responding household and long-form sample non-respondent (i.e., respondents in the long-form sample who answered the short-form questions but not the long-form questions). The information on non-respondents was thus somewhat limited. Fortunately, before the coverage and non-response adjustments, the process of WHI occurred. An important part of WHI is to impute the short-form characteristics for all non-respondents to the short form. This included long-form sample non-respondents. This additional information served as the basis for the long-form sample non-response adjustment.

The method used to adjust for coverage and total non-response in the long-form sample was a reweighting calibration-based procedure applied to the design weights. The procedure can be divided into the following main steps:

1. selection of calibration constraints for steps 2 and 3
2. non-linear calibration coverage adjustment
3. estimation of a non-response propensity based on non-linear calibration for non-response
4. application of a score method based on the propensity of step 3.

Steps 1 to 4 were applied independently in each SADA. In other words, the non-response adjustment was applied by SADA. See section 4.2 for the definition and information about ADAs and SADAs.

The first step consisted of a forward selection of calibration constraints in the SADA. It was performed as follows:

- The set of potential constraints was derived from the variables common to both the short form and the long form, as well as from some administrative data obtained with record linkage strategies (where all units of the long-form population undergo the linkage procedures). The requirements of the non-linear calibration method used in the second and third steps meant that only constraints at the SADA level, and the number of households and persons in each ADA of the SADA, were considered.
- In each SADA, two mandatory constraints were selected first: the number of households in the SADA (TOTHHLD) and the number of persons in the SADA (TOTPERS).
- The ADA-specific constraints—number of households (HHADA) and number of persons (PPADA)—were evaluated for selection.
- All other potential SADA constraints were evaluated; priority was given to the ones that split the SADA population as closely as possible into halves.

The selection process excluded constraints that occurred in fewer than 250 households in the SADA and constraints that were redundant or almost redundant in terms of collinearity with those constraints or with constraints already selected. Constraints that were redundant with constraints already selected were excluded since they did not add any new information. Given those filters, the order of priority used in the evaluation of constraints ensured that the constraints selected complemented each other and corrected for any potential coverage differential between the long form and the short form, as well as for census total non-response.

The second step applied a coverage non-linear calibration adjustment to the whole sample in the SADA (i.e., respondents and non-respondents). The long-form sample weighted counts, for the constraints selected in the first step, were made to coincide with the corresponding population counts. The purpose of this step was to correct for any potential coverage differential between the long-form sample and its complement (i.e., the set of households receiving only the short form). One way in which overcoverage can occur is if some individuals are counted in two different households. The coverage for the two populations could also be different if, for example, occupied dwellings were more likely to be incorrectly treated as unoccupied dwellings for the long form than for the short form. Another objective of this step was to isolate as much as possible the sampling error. Without this step, the non-response calibration carried out in the next step would confound the non-response error with the sampling error. This step makes the sample estimates coincide with the population estimates. In addition, the same control totals are used in both calibration procedures. As a result, the non-response propensity estimation done next does not have to correct (directly or indirectly) for the sampling error. Combining a correction for the sampling error and for the non-response error in the next step would have been inappropriate. The calibration procedure would have failed if the weight of any respondent was required to decrease to match the census counts, because the estimated propensity would have been greater than 1. Moreover, the score method applied in the last step required an estimate of the response propensity alone. To the extent that the variable of interest was related to the selected constraints, the sampling variance was also reduced by this step.

After these two steps, the main non-response adjustment took place. The weights, adjusted in the previous step, of non-respondents were set to 0 and the weights of respondents were increased so that the weighted sums in the SADA coincided with the corresponding population counts for the selected constraints. A logistic link function between the response propensity and the characteristics used in calibration enabled the implicit estimation of the response propensity. Folsom and Singh (2000) proposed this non-linear calibration method as a way of adjusting for non-response while ensuring that both estimates coincided with selected population counts and that the estimated response probabilities were between 0 and 1. This last condition does not necessarily hold when linear calibration is used for non-response adjustment. To the extent that the response propensity was related to the selected constraints, this step reduced the potential non-response bias without increasing the variance.

The inverse of the estimated response probabilities obtained in the previous step could be directly used to adjust the weights for non-response. However, the score method was used for the last step of the non-response adjustment to smooth the estimated probabilities from the previous step. This further ensured the quality of the non-response adjustment and avoided too extreme adjustments. For each ADA, homogeneous weighting classes were formed according to the estimated response probabilities. In each class, the weighted harmonic mean of the response probabilities was calculated. The harmonic mean was used because it is less affected by outliers in the estimated response probabilities. The inverse of this mean was applied to the weights of respondents in the class as the non-response adjustment.

In summary, the coverage and total non-response adjustment was a product of two quantities: the coverage adjustment and the inverse of the score-method harmonic mean.

4.5 Final calibration

Final calibration is a linear calibration that was done to minimize the sampling variability of estimates derived from long-form questionnaire responses, while ensuring consistency between estimated totals and Census of Population totals. This weighting step was necessary, since ensuring consistency between estimated totals and Census of Population totals was important for a large number of variables and geographic areas, i.e., satisfying calibration constraints.

Only the weights for households in mail-out or list/leave CUs were calibrated, since these households were sampled. Exceptions to this rule were households in these CUs that lived in a private dwelling attached to a collective dwelling. Since all these households were included in the long-form sample and all the long-form questionnaire responses for these households were imputed, no calibration was done. The final weights for these households were therefore equal to 1. The weights produced by the calibration process were the final weights used to calculate the long-form estimates, and these weights applied to households as well as families and persons. In other words, all families and persons from the same household received the household weight. For this final adjustment, the variability of the calibrated weights needed to be limited to avoid having an excessive portion of the weight applied to a single household or person. Therefore, weights were constrained to range from 1 to 20.

Calibration constraints were defined at the person, household and census family levels. In 2016, the notion of constraints was also expanded by the addition of two levels to the hierarchy of geographic units, i.e., ADAs and SADAs. These two levels were added to maximize the overall consistency between estimated totals and Census of Population totals, while minimizing the number of calibration constraints. This should help to reduce the variability of estimates. Appendix C lists all of the ADA and SADA constraints that were taken into consideration during the calibration process. Characteristics available from the census, administrative sources and the long-form questionnaire and for which consistency was attempted included age, sex, marital status, common-law status, household size, dwelling type and official language spoken.

Constraints were chosen in both types of geographic areas simultaneously, independent of each other. Calibration was then performed using all of the selected constraints. In 2016, the addition of calibration constraints for two geographic levels (ADAs and SADAs), the removal of mother tongue constraints, and the addition of constraints based on linked administrative data meant that the number of possible constraints was different than for the 2011 Census. A total of 270 constraints were defined for ADAs and 200 for SADAs. Various factors drove the choice of geographic level for calibration constraints. This choice was made in collaboration with subject-matter experts. For example, some constraints were defined only for SADAs, since they would not have been populated enough at the ADA level. Other constraints, such as age groups, were chosen in a way that ensured they were not only populated enough but also not too similar when assessed by the selection process.

To facilitate their calibration, small ADAs were combined before the selection of calibration constraints to ensure a minimum of 60 long-form respondent households per ADA. Small ADAs that fell entirely within a CSD were initially combined with other ADAs in the same SADA. Next, small ADAs in CDs were combined with other ADAs in the same SADA. Finally, the remaining small ADAs were combined with an ADA from an adjacent SADA. The ADA grouping procedure produced 4,180 groups of ADAs with 60 or more respondent households.

The first step in the process to select calibration constraints was to categorize each of the constraints into one of three groups:

Mandatory constraints: These constraints had to be used in the calibration because the census counts had to agree with the long-form estimates at the geographic levels that are usual aggregates of ADAs and SADAs (e.g., Canada, provinces and territories). The number of persons and the number of households in the ADAs and SADAs were the two mandatory constraints.

Low-response constraints: Constraints evaluated for a population of 200 or fewer households were not used in the calibration because they can make survey estimates unstable.

All other constraints: These constraints were examined further to see whether they should be used in the calibration.

The second step was to determine which constraints from the third group should be used in the calibration process, in addition to the mandatory constraints. The constraints from the third group were added one by one, by repeatedly choosing the constraint that divided the population of the SADA or ADA in two as evenly as possible. Constraints that were too linearly dependent were excluded. To avoid introducing a bias in the point estimates and to avoid increasing their variance, the number of selected constraints was limited. Evaluations determined that this number had to be smaller than the square root of the number of respondent households involved in the constraint.

After the calibration constraints to be used were selected, a final edit was done to check whether the set of constraints chosen at the ADA and SADA levels was free of collinearity.

The calibration itself was then carried out for the final set of constraints from the second step. The weights adjusted for coverage and non-response were modified as little as possible, so that the weighted estimates would be equal to census totals for these constraints. Statistics Canada's Generalized Estimation System (GES) was used to carry out the calibration.

Sample estimates can differ from census counts for a few reasons, particularly for small areas, even after the calibration step. A few of these reasons are given below.

- Constraints excluded during the constraint selection process: As described above, possible constraints could be excluded for having low counts, for being linearly dependent (or overly dependent) on other chosen constraints or for being linearly dependent (or overly dependent) on low-response constraints. This led to some differences between census counts and long-form estimates for these variables when a perfect linear dependency with the chosen constraints was not present.
- Sub-weighting area: In 2016, the ADA was the smallest weighting area for which agreement was attempted between the census counts and the long-form estimates. Any entity smaller than an ADA, such as the majority of DAs, is referred to as a sub-weighting area. These sub-weighting areas could have discrepancies between the census counts and the long-form estimates.

4.6 Details on the selection of constraints[4]

Constraints were selected twice during the weighting process: during the coverage and non-response adjustment, which requires the use of non-linear calibration techniques, and during the final calibration. The variables making up the constraints were essentially the same, but the inclusion or exclusion of constraints varied between the two weighting steps, since their respective objectives were different. Basically, constraints were not selected using the exact same criteria, and the weighting areas varied depending on the weighting step.

This section explains how the constraints were selected during these weighting steps and indicates how often certain constraints were excluded during the coverage and non-response adjustment and during the final calibration. Constraint selection was carried out independently in the 408 SADAs that had sampled households with an adjusted weight, as well as in the 4,238 corresponding ADAs. One SADA did not undergo a weight adjustment since it had only counted households (the SADA corresponding to Nunavut). This explains why the following tables present results for 408 SADAs, while the total number of SADAs indicated in section 4.2 was 409. During the coverage and non-response adjustment, 4,646 weighting areas were defined, i.e., 408 SADAs and 4,238 ADAs, while 4,588 weighting areas were defined at the final calibration step. The number of weighting areas differed depending on the weighting step because, during the final calibration, small ADAs were combined into "calibration ADAs" to ensure a minimum of 60 households per ADA, for a total of 4,180 ADAs at the calibration step.

4.6.1 Process for the coverage and non-response adjustment

The coverage and non-response adjustment procedure uses calibration to adjust the survey weights. The rationale is that, if estimates based on respondents match as much as possible the census counts for auxiliary variables, then the non-response bias of estimates associated to those variables will be reduced. See section 4.4 for more information on the coverage and non-response adjustment.

As mentioned in section 4.4, the coverage and non-response adjustment was carried out independently in the SADA, where several variables were used to define the constraints, and in the ADA for two variables. A sequential procedure was used to select the constraints. In each SADA, two mandatory constraints were selected first: the number of households in the SADA (TOTHHLD) and the number of persons in the SADA (TOTPERS). After that,

4. The selection of constraints specific to the Wood Buffalo area was an exception, since this area had its own process. That is why the selection of constraints for this census subdivision was excluded from this chapter and is presented in Appendix D.

the ADA-specific constraints—number of households (HHADA) and number of persons (PPADA)—in each of the ADAs were considered. Then, all the other potential SADA constraints were considered, with priority given to the ones that split the SADA (approximately) in half. The non-mandatory constraints were entered sequentially into the adjustment as long as they did not fall into the criteria listed in Table 4.6.3.1. The final list of calibration constraints that were considered for the coverage and non-response adjustment is presented in Appendix C.

A constraint that is excluded frequently usually has a larger difference between the census count and the non-response adjusted sample estimate than a constraint that is used more often. This can be seen by comparing Appendix C with Table 5.2.1. Appendix C lists all the potential variables or constraints, the number of times each constraint was used for calibration and the number of times that constraints were excluded for one of the reasons enumerated in Table 4.6.3.1. A constraint was excluded from the coverage and non-response adjustment under five criteria: “No population,” “Small population,” “Linearly dependent,” “High collinearity” and “Explanatory redundancy.” A constraint may have been excluded from calibration for one of the reasons above and yet been calibrated at the end of the process. This occurs for example when the constraint is collinear with the selected constraints. In this case, Appendix C shows that constraint as “Calibrated.”

4.6.2 Procedures for the final calibration adjustment

The purpose of the final calibration is to adjust the household weights so that the long-form estimates are as close as possible to the census counts for many common characteristics. In addition to producing agreement between the estimates, an appropriate choice of constraints reduces variance. Appendix C provides the complete list of possible constraints, and section 4.5 describes the calibration performed on the long-form estimates. The criteria applied to the selection of constraints are similar to those applied to the selection of constraints for the coverage and non-response adjustment, with a few differences as presented in Table 4.6.3.1.

Calibration was performed simultaneously for SADAs and calibration ADAs. In each SADA, calibration constraints were defined at the SADA and ADA levels. All constraints were evaluated in each SADA and excluded only if necessary. The total number of persons (TOTPERS) and the total number of households (TOTHHLD) were the only two mandatory constraints. This meant that they could not be excluded for any of the ADAs.

Statistics Canada’s GES was sent a total of 132,777 preselected constraints at the national level, so that it could perform the final linear calibration. This represents an average of 27 constraints per ADA and an average of 31 constraints per SADA. The mandatory constraints were selected in all the weighting areas. The process selected income constraints the most often, particularly the household income constraint and the low-income household constraint. The constraints selected the least often were primarily the year of immigration and the country of origin.

4.6.3 Comparison of the procedures for the two adjustment and selection steps

Criteria were applied in the selection of constraints at each weighting step. These criteria are indicated in Table 4.6.3.1 by weighting step.

Table 4.6.3.1
Criteria applied in selecting coverage, non-response and final calibration adjustment constraints

Criteria	Adjustment for coverage and non-response	Final calibration
No population according to the census counts: If the constraint had no population in the weighting area, then the estimate after adjustment must also be 0 for that constraint. These constraints are not classified as excluded but rather as ineligible to the adjustment process.	Applied at the SADA/ADA level.	Applied at the SADA/ADA level.
Small population according to the census counts: If a constraint involves less than a certain number of households in the population of the weighting area, then it is considered small and the constraint is excluded. Including such a constraint would unduly increase the variance. However, constraints with small population can be implicitly calibrated and in this case are included in the total number of calibrated constraints.	Applied at the SADA/ADA level. The number of households in the population is larger than 0 but less than 250 in the weighted area.	Applied at the SADA/ADA level. The number of households in the population is more than 0 but less than 200 in the weighted area.
Linearly dependent: If the value of a constraint can be calculated by combining the values of other constraints, one of these constraints is not necessary and must be deleted during the adjustment process because of its linear dependency. However, constraints that are excluded because of their linear dependency are implicitly calibrated. They are therefore included in the total number of calibrated constraints.	Applied at the SADA level. The selection of constraints can be compared to the selection of explanatory variables in a linear regression. The variance inflation factor[1] (VIF) and the condition number[2] are thus used to detect high collinearity.	Applied at the SADA/ADA level. Two dependency checks are conducted to identify linearly dependent constraints. The first check is done when the constraints at the SADA/ADA level are selected, and the second check includes all the constraints chosen at both levels of the geographic hierarchy (SADAs and ADAs).
High collinearity: If a constraint value can be almost calculated by the combination of other constraint values, then at least one of those constraints must be avoided in the adjustment process. Such a constraint is not perfectly calibrated.	Applied at the SADA level. The selection of constraints can be compared to the selection of explanatory variables in a linear regression. The variance inflation factor[1] (VIF) and the condition number[2] are thus used to detect high collinearity.	Applied at the SADA/ADA level. Two linear dependency checks are conducted to identify constraints that are close to being linearly dependent. The first check is done when the constraints at the SADA level and the ADA level are selected, and the second check includes all the constraints chosen at both levels of the hierarchy simultaneously (SADAs and ADAs).

Table 4.6.3.1
Criteria applied in selecting coverage, non-response and final calibration adjustment constraints

Criteria	Adjustment for coverage and non-response	Final calibration
Explanatory redundancy: If a constraint explains the non-response (almost) as well as other constraint(s) already selected, then the non-response calibration procedure would fail. This is equivalent to saying that if a constraint does not add any information about the non-response mechanism, beyond what is explained by the already-selected constraints, then it should not be included.	Applied at the SADA. A sequential procedure is applied (a form of logistic regression) to test the convergence of the logistic regression.	N/A

1. The VIF quantifies the increase in variance of regression coefficients due to collinearity.
2. The condition number quantifies the degree to which a matrix is close to singularity.
SADA = Super aggregate dissemination area.
ADA = Aggregate dissemination area.

Appendix C indicates the status of each constraint selected in at least one of the geographic areas once the selection of constraints was carried out at the weighting step. The geography column indicates the geographic level to which the constraint was applied. When a constraint was applied to both geographic levels, the totals reflect both levels with no distinction. A constraint could have been excluded from the calibration process and still have been calibrated. In that case, the constraint was considered to be calibrated.

4.6.4 Analysis of calibration during the coverage and non-response adjustment

This section summarizes the number of constraints selected and excluded. The reasons for not selecting constraints are also summarized. Additionally, the section sheds some light on the reasons why some constraints are frequently excluded.

Persons born in certain places tend to be more concentrated in certain parts of the country, to the point that many SADAs have little or no population with a given place of birth. As a result, the constraints for place of birth were often not selected because they had small census counts. Similarly, the constraint involving the French official language (OLN_FR) has little or no population in certain regions of the country. Consequently, it was often excluded.

The constraints most often excluded because of high collinearity were:

- “Persons in the ADA” (PPADA)
- “Children in a census family” (CHILD)
- “Census families with children” (CHILDFAM)
- “One-person households” (HHSIZE1)
- “Female” (FEMALE)
- “Male” (MALE)
- “Census families without children” (NOCLDFAM)
- “Females aged 14 years and younger” (FEMALELT15)
- “Persons in an economic family” (INEFAM)
- “Persons in a household that are not part of an economic family” (NOINEFAM)
- “Males aged 14 years and younger” (MALELT15)
- “Persons aged 10 to 14 years” (AGE14)

- “Five-or-more-person households” (HHSIZEGE5)
- “Persons not in a census family” (NOTINFAM)
- “Persons aged 5 to 9 years” (AGE9)
- “Persons aged 0 to 4 years” (AGE4)
- “Six-or-more-person households” (HHSIZEGE6)
- “Adults in a census family” (ADULTCF)
- “Persons in a couple (spouse, partner)” (COUPLE)
- “Two-person households” (HHSIZE2).

The procedure excluded these constraints automatically, since they could be determined very accurately from a combination of the mandatory constraints, “Households” (TOTHHLD) and “Persons” (TOTPERS), together with other constraints that were selected often, such as some constraints for age, marital status, household size, sex by age and persons in census family. These constraints might also have been explained too well by a combination of the variables selected and of the small constraints.

The actual differences between the census counts and the non-response adjusted estimates are examined in section 5.2.

Table 4.6.4.1 shows the number of times that each reason for dropping or removing a SADA-level constraint occurred. The total number of constraints excluded is the sum of the “Small population,” “High collinearity” and “Explanatory redundancy” categories. The “No population” category is not included in the total because it does not actually represent excluded constraints. The average number of constraints excluded per SADA is the total for each category divided by 408, the number of SADAs where coverage and non-response adjustments were done.

Table 4.6.4.1
Summary statistics on SADA-level constraints in 2016 coverage and non-response adjustment

			Excluded			
Constraint	Calibrated	No population	Small population	High collinearity	Explanatory redundancy	Total excluded
Total constraints	30,328	2,865	29,742	11,550	49	41,341
Average number of constraints per SADA	74.3	7.0	72.9	28.3	0.1	101.3

SADA = Super aggregate dissemination area.
Note: Two SADAs were excluded: the Wood Buffalo SADA and a second one that consists of on-reserve households only.
Source: Statistics Canada, 2016 Census long-form sample.

On average, 74.3 SADA-level constraints were calibrated per SADA. An average of 72.9 constraints were discarded per SADA because of small population and 28.3 because of high collinearity.

4.6.5 Analysis of the final calibration

Other than cases where the population was nil, each time that a constraint was excluded, the calibration process did not attempt to make census counts and long-form estimates agree for that constraint in that weighting area. The gap between the census count and the long-form estimate was usually larger for a constraint that was excluded frequently than for a constraint that was excluded less often.

Table 4.6.5.1 shows how often a constraint was removed or excluded at the weighting area level according to each criterion. The total number of constraints excluded is equal to the sum of the values for the various removal criteria. The average number of constraints excluded per weighting area is simply equal to the total for the category divided by the number of weighting areas.

Table 4.6.5.1
Summary statistics on constraint selection status at the weighted area level in the final weight adjustment, 2016

Constraints	Weighted area	Calibrated	No population	Excluded		Total excluded	Number of weighted areas
				Small population	High collinearity		
Number of constraints	ADA	210,307	112,442	283,947	99,319	383,266	4,180
	SADA	39,262	2,698	27,292	10,588	37,880	408
Average number of constraints by SADA/ADA	ADA	50.3	27.1	67.9	23.8	91.7	4,180
	SADA	96.2	6.6	66.9	26.0	92.0	408

SADA = Super aggregate dissemination area.
ADA = Aggregate dissemination area.
Note: Two SADAs were excluded: the Wood Buffalo SADA and a second one that consists of on-reserve households only.
Source: Statistics Canada, 2016 Census long-form sample.

5. Evaluation of the weighting procedures

As described in Chapter 4, the first step in weighting the long-form sample was to assign design weights to households. Weights were assigned differently depending on the collection method of the CU where the household was living. Adjustments were applied only to private households with a design weight greater than 1. These private households were not attached to a collective dwelling in a mail-out or list/leave CU. All of the results presented in this chapter were calculated for this subset of households. The final weight for the other private households therefore corresponds to the design weight and remains at the initial value of 1.

In short, each household was assigned a design weight that was determined by the long-form sample design. Some adjustments were then necessary to address coverage and total non-response. Non-linear calibration was performed during this adjustment to estimate the parameters for non-response models. After being adjusted for coverage and total non-response, the weights were adjusted further in the final calibration process to produce the final weights. The final weights enabled generally better agreement between the census counts and the long-form estimates.

The next few sections examine the distribution of the weights and, for various characteristics, the discrepancies between the census counts and the sample estimates at various aggregate geographic levels.

5.1 Distribution of the weights

Chart 5.1.1 and tables 5.1.1 and 5.1.2 illustrate the distribution of the design weights, the weights adjusted for coverage and non-response, and the weights adjusted in the final calibration. The weights are grouped by 0.5 length intervals, with the exception of the first and last intervals. The chart shows the percentage of times the weights appear in each interval. Nearly all the design weights ranged from 3.75 to 4.25. This is because of the long-form sample design, in which approximately one in four households received a long-form questionnaire in most areas. The impact on the weights of the coverage and non-response adjustments and the final calibration adjustments can also be seen. A very noticeable difference is that the percentage of households that had a design weight from 3.75 to 4.25 was significantly reduced from 100% to 67%, ending at close to 30%. The final weights were more evenly distributed within categories 1 to 20.

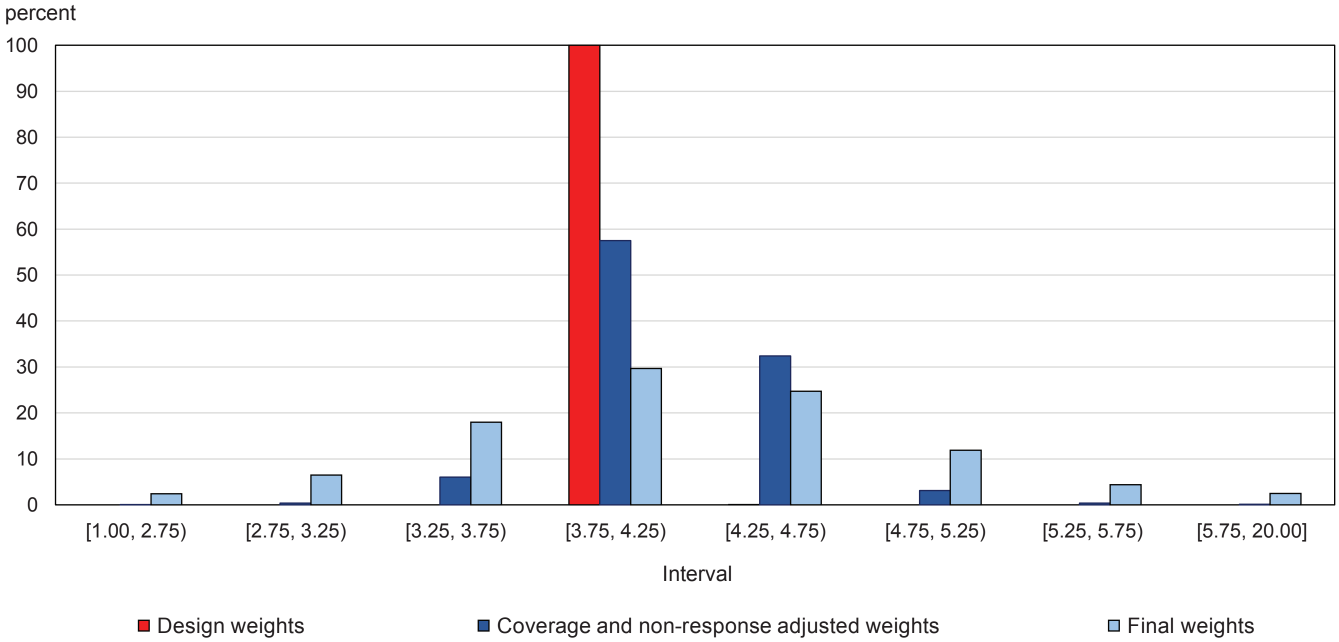
Logically, the non-response adjustment process should tend to increase the weights to compensate for the non-responding units. This did occur for most cases. The changes between the design weights and the coverage and non-response adjusted weights can be observed in Table 5.1.1. This table shows that most of the units that left the [3.75, 4.25) range moved to the [4.25, 4.75) or [4.75, 5.25) ranges. However, the coverage and non-response adjustment process moved the weights from the 3.75+ range to the [1, 2.75), [2.75, 3.25) or [3.25, 3.75) ranges for a few units. The main reason is that the procedure included an adjustment for overcoverage and undercoverage. To the extent that a few population groups may have experienced overcoverage, the weights would have been reduced in those areas.

An important element to notice in Table 5.1.1 is that the non-responding units originally had positive weights, since they were selected for the sample. The non-response adjusting process assigned them a weight of 0 and redistributed their original weights among responding units. The non-responding units correspond to the line labelled “0 (non-respondents)” in Table 5.1.1 and were removed from Table 5.1.2, since they were not used in the calibration process. Table 5.1.2 presents the changes between the coverage and non-response adjusted weights and the calibrated weights.

According to Table 5.1.2, most weights experienced only a small modification during the calibration process. In fact, 76.6% of cases either stayed in the same range or moved only one range up or down. The most stable range was 1 to 2.75, where 64% of the households with a non-response adjusted weight between 1 and 2.75 stayed in that category after calibration. The second most stable category was 5.75+, where 61.8% of households with a non-response adjusted weight between 5.75 and 9.5 stayed in that category (although the calibration range goes up to 20 rather than 9.5).

Finally, whereas the non-response adjusted weights varied from 1 to 9.5, the range of the final weights was from 1 to 20. This variability in the weights is far smaller than the variability in the final weights in the 2011 National Household Survey (see Statistics Canada, 2015), where the maximum weight was capped at 100.

Chart 5.1.1
Distribution of design weights, coverage and non-response adjusted weights, and final weights



Notes: All households with a design weight of 1 were excluded from the weighting process. These households either were located in an Indian reserve or a canvasser enumeration collection unit or were private households attached to a collective dwelling.
The "[" symbol means the number is included in the interval and the ")" symbol means it is not included in the interval.
Source: Statistics Canada, 2016 Census long-form sample.

Table 5.1.1
Distribution of design weights and coverage and non-response adjusted weights

Coverage and non-response adjusted weights	Design weights		
	[3.75, 4.25)	[4.25, 7.00]	Total
0 (non-respondents)	122,807	19	**122,826**
[1.00, 2.75)	1,548	0	**1,548**
[2.75, 3.25)	12,363	0	**12,363**
[3.25, 3.75)	201,534	2	**201,536**
[3.75, 4.25)	1,921,456	60	**1,921,516**
[4.25, 4.75)	1,082,895	239	**1,083,134**
[4.75, 5.25)	103,892	85	**103,977**
[5.25, 5.75)	12,783	14	**12,797**
[5.75, 20.00]	3,730	5	**3,735**
Total	**3,463,008**	**424**	**3,463,432**

Notes: All households with a design weight of one are excluded from the weighting process. These households either come from Indian reserve and canvasser enumeration CUs or are private households attached to collective.
The "[" symbol means the number is included in the interval and the ")" symbol means it is not included in the interval.
Source: Statistics Canada, 2016 Census long-form sample.

Table 5.1.2
Distribution of coverage and non-response adjusted weights and final weights

	Coverage and non-response adjusted weights								
Final weights	**[1.00, 2.75)**	**[2.75, 3.25)**	**[3.25, 3.75)**	**[3.75, 4.25)**	**[4.25, 4.75)**	**[4.75, 5.25)**	**[5.25, 5.75)**	**[5.75, 9.50)**	**Total**
[1.00, 2.75)	990	3,750	18,082	42,843	12,597	1,402	196	71	79,931
[2.75, 3.25)	340	3,805	38,693	138,180	33,513	2,196	205	37	216,969
[3.25, 3.75)	142	2,910	63,351	414,026	115,038	5,456	454	86	601,463
[3.75, 4.25)	42	1,266	50,385	649,869	275,758	12,551	878	135	990,884
[4.25, 4.75)	27	413	21,470	444,950	334,973	22,686	1,522	219	826,260
[4.75, 5.25)	7	151	6,648	165,518	196,585	25,532	2,354	364	397,159
[5.25, 5.75)	0	48	1,986	47,211	75,313	17,880	2,601	514	145,553
[5.75, 20.00]	0	20	921	18,919	39,357	16,274	4,587	2,309	82,387
Total	**1,548**	**12,363**	**201,536**	**1,921,516**	**1,083,134**	**103,977**	**12,797**	**3,735**	**3,340,606**

Notes: All households with a design weight of one are excluded from the weighting process. These households either come from Indian reserve and canvasser enumeration CUs or are private households attached to collective.
The "[" symbol means the number is included in the interval and the ")" symbol means it is not included in the interval.
Source: Statistics Canada, 2016 Census long-form sample.

5.2 Discrepancies between census counts and long-form estimates, Canada

Chapter 4 describes the methods used to calculate the final household weights, and section 5.1 shows some of the relationships between design weights, non-response adjusted weights and final weights. The coverage and non-response adjustment reduced the discrepancies between the census counts and the corresponding long-form estimates for the constraints considered (see Appendix C). Following those adjustments, calibration further reduced or eliminated those discrepancies for certain variables (constraints). However, some discrepancies remain, since constraints are sometimes excluded discarded (see section 4.6). The relative difference between census counts and long-form estimates also called discrepancy is defined as:

$$\text{Discrepancy} = \frac{(\text{long-form estimate} - \text{census count})}{\text{census count}} \times 100\%$$

The numerator in the above expression (long-form estimate – census count) is referred to as the difference between census counts and long-form estimates. By dividing this value by the census count, the difference between census counts and long-form estimates relative to the size of the census count can be seen. In other words, the ratio represents the percentage that the characteristic was overestimated (a positive value) or underestimated (a negative value).

Table 5.2.1 shows the 2016 Canada-level differences between census counts and long-form estimates for the constraints considered for the design weights, the coverage and non-response adjusted weights and the final weights. The census count for the total persons characteristic (TOTPERS) is lower than the published figure of the 2016 Census (35,151,728). This difference is the result of the weighting process, which used only private households (not those living in dwellings attached to collective dwellings) and excluded Indian reserve and canvasser enumeration CUs to use a common population in the comparison.

Table 5.2.1 also shows percentage relative differences (called discrepancies) for estimates based on final weights. Looking at these discrepancies sheds more light on the differences. Over 88% of the cases in Table 5.2.1 had a discrepancy from -1% to 1%, and over 97% of them from -5% to 5%.

Chart 5.2.2 shows, for all the constraints, the difference between the census counts and each of the three estimates: design-weighted (red), non-response adjusted (purple) and final (blue). The x-axis represents the population size of the constraint, and the three lines show for each constraint:

- the difference between the sum of the design weights and the census count
- the sum of the non-response adjusted weights and the census count
- the sum of the final weight and the census count.

The constraints are sorted, from left to right, by increasing population size.

Chart 5.2.3 shows the percentage discrepancies between census counts and final estimates for all the constraints by population size. For the medium-sized and large constraints, the discrepancies are all small. Only certain small constraints have relatively large discrepancies. For example, the difference between the census count and the design weight estimate for the “Single person, never married” (COMMONLAWNO_SINGLE) constraint, with a population size of 13,624,243, is about -171,000 (red). In contrast, the non-response adjustment reduces that difference to only about -6,000 (purple), and the final calibration further reduces the difference to about -2,000 (blue).

The most important observation from Chart 5.2.2 is that the coverage and non-response adjustment carries a big improvement over the design-weighted estimates. Although it is not apparent in the chart, the coverage adjustment does most of the job. The difference between census counts and long-form estimates for design weights tended to be (much) greater than the difference between census counts and long-form estimates for the non-response adjusted weights. This, in turn, tended to be greater than the corresponding difference using the final weights. This shows the importance of the non-response adjustment and calibration processes. As mentioned in section 4.6, a difference between the census count and long-form estimate could occur in a SADA or ADA for a characteristic if its constraint is excluded during calibration. In other words, the process did not control on the excluded constraint for a given area. If the constraint is excluded in many areas, these differences could partially cancel each other out, or they could cumulate to create a large difference at the Canada level. Total persons (TOTPERS) and total households (TOTHHLD) were the only mandatory constraints for which agreement between census counts and long-form estimates had to be guaranteed for all ADAs. As a result, the final weight difference and discrepancy for these characteristics were 0. However, all other constraints had to be excluded in some areas.

Section 4.6 pointed out the constraints that were excluded frequently and where high differences or discrepancies might lie. The effect of dropping those constraints can be seen in Table 5.2.1. The SADA-only constraints that were most often excluded[5] were:

- “Persons in an economic family” (INEFAM)
- “Persons in a household that are not part of an economic family” (NOINEFAM)
- “Persons in a couple (spouse, partner)” (COUPLE)
- “Persons in a two-or-more-person household that are not part of an economic family” (NOINEFAMHHSIZEGT1)
- “Adults in a census family” (ADULTCF)
- “Persons not in a census family” (NOTINFAM)
- “Children in a census family” (CHILD)
- “Persons in a two-or-more-person household that are not part of a census family” (NOTINFAMHHSIZEGT1)
- “Lone parents” (LONEPAR).

5. Excluded in more than 60% of the SADAs.

Not surprisingly, the largest differences (more than 2,000 in absolute value) were obtained precisely for these constraints. The largest difference was obtained for “Persons in an economic family” (INEFAM) (14,493) and for its complement, “Persons in a household that are not part of an economic family” (NOINEFAM). However, because the census counts were so high, the percentage discrepancies for all these constraints were not too high (ranging from -0.93% to 0.06%). In the 2011 National Household Survey, the maximum difference was 166,801 (see Statistics Canada, 2015).

The largest percentage discrepancies were obtained for some of the places of birth. Since many places of birth are not found in high numbers in many parts of the country, these variables were often excluded during calibration. This led to some differences in the census counts and long-form estimates. Furthermore, because many places of birth have relatively low census counts, the relative differences are magnified. This resulted in greater percentage discrepancies. The largest percentage discrepancy, 12.87%, was for the POBG2_49 constraint, which corresponds to “Mongolia.” This constraint was rarely selected because it had either no population or a small population in the area. Other places of birth with high discrepancies, all of which were also rarely selected, were “Botswana, Lesotho and Swaziland” (POBG2_9) with 10.53%; “Canada” (POBG2_13) with 8.56%; “Malta” (POBG2_44) with 6.22%; “Guadeloupe, Martinique and Saint Martin (French part)” (POBG2_15) with 6.1%; “Andorra, Gibraltar and Spain” (POBG2_2) with 4.75%; and “Burma (Myanmar)” (POBG2_12) with 4.53%.

All the household and family characteristics had small percent discrepancies, with the largest being an underestimation of 0.34%, for households of size 6 or more (HHSIZEGE6). The largest difference for this type of characteristics was observed for households of size 1, with an underestimation of 2,025 households. Other household characteristics with relatively large differences[6] (but still small discrepancies) are “Two-person households” (HHSIZE2), “Census families without children” (NOCLDFAM), “Census families” (TOTCFAM), “Six-or-more-person households” (HHSIZEGE6) and “Five-or-more-person households” (HHSIZEGE5). The closest estimate for a non-mandatory constraint of this type belonged to “Households living in a single-detached house” (SNGLDET), with a difference of 56 and a discrepancy of 0.001%.

Other than places of birth, the person-level characteristics with the largest discrepancies were the constraints “Persons in a two-or-more-person household that are not part of an economic family” (NOINEFAMHHSIZEGT1), with a discrepancy of -0.93%, and “Persons in a two-or-more-person household that are not part of a census family” (NOTINFAMHHSIZEGT1), with a discrepancy of -0.32%. Other than the ones mentioned above, the person-level characteristics with an overestimation or underestimation of at least 1,000 persons were:

- “Single persons (never married) not in a common law relationship” (COMLAWNO_SINGLE)
- “Single persons (never married) under 15 years and not in a common law relationship” (COMLAWNO_SINGLE_LT15)
- “Persons in a common law relationship” (COMLAW_YE)
- “Married persons” (MARRIED)
- “Place of birth - Belarus, Moldova, Russian Federation and Ukraine” (POBG2_28).

However, the discrepancies for all of these were small, the largest being 0.74%, for the last.

6. More than 1,200 in absolute value.

The closest estimates for non-mandatory person-level constraints belonged to the Age 35 to 44 (AGE35_44) and Age 45 to 54 (AGE45_54) groups and to persons whose employment income is above the 75th percentile in the SADA (EMPIN_SADA_P100). These characteristics were estimated without sampling error at the Canada level. Other person-level constraints with very small differences and discrepancies[7] included:

- "Persons aged 55 to 59 years" (AGE59)
- "Persons with an annual employment income above the 50th percentile for the SADA" (EMPIN_SADA_GT50)
- "Persons with an annual employment income above the 50th percentile and equal to or below the 75th percentile for the SADA" (EMPIN_SADA_P75)
- "Persons aged 20 to 24 years" (AGE24)
- "Persons with an annual employment income equal to or below the 50th percentile for the SADA" (EMPIN_SADA_LE50)
- "Persons aged 15 to 29 years" (AGE15_29)
- "Persons aged 50 to 64 years" (AGE50_64).

Finally, among the SADA-only constraints, the largest difference (-640) and percentage discrepancy (-0.02%) was obtained for the constraint "Households with an annual income at or below the 25th percentile for the SADA" (HHINC_SADA_P25). All these constraints were selected relatively often. The constraint that was selected the least often, "Persons with an annual employment income above the 25th percentile and equal to or below the 50th percentile for the SADA" (EMPIN_SADA_P50), was selected in 62.7% of the ADAs (as EMPIN_P50).

Table 5.2.1
Census counts and long-form estimate differences and discrepancies – Canada

		Design weights		Coverage and non-response adjusted weights		Final weights		
Characteristic	Census counts	estimates	difference	estimates	difference	estimates	difference	discrepancy (%)
ADULTCF	17,873,064	17,796,322	-76,742	17,880,659	7,595	17,879,104	6,040	0.03
AGE00_14	5,679,559	5,629,183	-50,376	5,684,239	4,680	5,679,522	-37	0.00
AGE4	1,845,178	1,828,386	-16,792	1,848,647	3,469	1,845,517	339	0.02
AGE9	1,962,324	1,945,126	-17,198	1,962,857	533	1,962,244	-80	0.00
AGE14	1,872,057	1,855,671	-16,386	1,872,734	677	1,871,761	-296	-0.02
AGE15_24	4,155,227	4,098,212	-57,015	4,152,363	-2,864	4,155,209	-18	0.00
AGE15_29	6,387,808	6,297,713	-90,095	6,383,708	-4,100	6,387,805	-3	0.00
AGE19	1,970,121	1,949,762	-20,359	1,970,676	555	1,970,110	-11	0.00
AGE24	2,185,106	2,148,450	-36,656	2,181,687	-3,419	2,185,099	-7	0.00
AGE25_34	4,513,475	4,453,702	-59,773	4,512,230	-1,245	4,513,547	72	0.00
AGE29	2,232,581	2,199,501	-33,080	2,231,345	-1,236	2,232,597	16	0.00
AGE30_49	9,047,387	8,960,912	-86,475	9,045,676	-1,711	9,047,405	18	0.00
AGE34	2,280,894	2,254,201	-26,693	2,280,885	-9	2,280,950	56	0.00
AGE35_44	4,454,400	4,412,465	-41,935	4,452,867	-1,533	4,454,400	0	0.00
AGE39	2,243,631	2,223,020	-20,611	2,242,432	-1,199	2,243,735	104	0.00
AGE44	2,210,769	2,189,445	-21,324	2,210,436	-333	2,210,665	-104	0.00

7. Difference less than 10 in absolute value and discrepancy less than 0.005%.

Table 5.2.1
Census counts and long-form estimate differences and discrepancies – Canada

Characteristic	Census counts	Design weights		Coverage and non-response adjusted weights		Final weights		
		estimates	difference	estimates	difference	estimates	difference	discrepancy (%)
AGE45_54	4,936,819	4,905,005	-31,814	4,936,661	-158	4,936,819	0	0.00
AGE49	2,312,093	2,294,247	-17,846	2,311,923	-170	2,312,055	-38	0.00
AGE50_64	7,436,076	7,406,957	-29,120	7,435,705	-371	7,436,084	8	0.00
AGE54	2,624,726	2,610,758	-13,968	2,624,738	12	2,624,764	38	0.00
AGE55_64	4,811,350	4,796,199	-15,151	4,810,967	-383	4,811,321	-29	0.00
AGE59	2,568,136	2,559,441	-8,695	2,569,510	1,374	2,568,127	-9	0.00
AGE64	2,243,214	2,236,757	-6,457	2,241,457	-1,757	2,243,194	-20	0.00
AGE65PL	5,443,624	5,445,343	1,719	5,444,038	414	5,443,637	13	0.00
AGE74	3,302,938	3,302,147	-791	3,303,259	321	3,302,967	29	0.00
AGE75PL	2,140,686	2,143,195	2,509	2,140,779	93	2,140,670	-16	0.00
APTLT5	2,533,749	2,515,260	-18,489	2,532,535	-1,214	2,533,675	-74	0.00
CHILD	10,082,408	9,998,775	-83,633	10,102,364	19,956	10,085,159	2,751	0.03
CHILDFAM	5,722,912	5,681,084	-41,828	5,728,040	5,128	5,722,653	-259	0.00
COMLAW_YE	3,440,562	3,416,138	-24,424	3,447,069	6,507	3,442,087	1,525	0.04
COMLAWNO_DIV	1,749,631	1,741,004	-8,627	1,749,867	236	1,749,963	332	0.02
COMLAWNO_OTHERS	3,793,163	3,778,466	-14,697	3,792,762	-401	3,793,576	413	0.01
COMLAWNO_SEP	688,126	685,664	-2,462	687,481	-645	688,056	-70	-0.01
COMLAWNO_SINGLE	13,624,243	13,453,334	-170,909	13,618,465	-5,778	13,622,342	-1,901	-0.01
COMLAWNO_SINGLE_GE15	7,944,684	7,824,151	-120,533	7,934,226	-10,458	7,942,819	-1,865	-0.02
COMLAWNO_WID	1,355,406	1,351,797	-3,609	1,355,414	8	1,355,557	151	0.01
COMLAWYE_MARRIED	16,577,048	16,508,308	-68,740	16,582,139	5,091	16,578,536	1,488	0.01
COUPLE	16,304,196	16,244,300	-59,896	16,317,065	12,869	16,313,204	9,008	0.06
EMPIN_SADA_GT50	10,115,337	10,071,665	-43,672	10,115,480	143	10,115,329	-8	0.00
EMPIN_SADA_LE50	10,116,607	10,006,750	-109,857	10,115,995	-612	10,116,600	-7	0.00
EMPIN_SADA_P0	13,762,510	13,661,692	-100,818	13,761,890	-620	13,762,525	15	0.00
EMPIN_SADA_P0_GE15	8,082,951	8,032,509	-50,442	8,077,651	-5,300	8,083,003	52	0.00
EMPIN_SADA_P25	5,058,711	5,008,873	-49,838	5,058,197	-514	5,059,199	488	0.01
EMPIN_SADA_P50	5,057,896	4,997,877	-60,019	5,057,797	-99	5,057,401	-495	-0.01
EMPIN_SADA_P75	5,057,676	5,030,940	-26,736	5,058,134	458	5,057,668	-8	0.00
EMPIN_SADA_P100	5,057,661	5,040,726	-16,935	5,057,347	-314	5,057,661	0	0.00
FEMALE	17,258,035	17,153,083	-104,952	17,262,580	4,545	17,258,007	-28	0.00
FEMALEGE15	14,489,305	14,407,839	-81,466	14,489,472	167	14,489,532	227	0.00
FEMALELT15	2,768,730	2,745,244	-23,486	2,773,108	4,378	2,768,475	-255	-0.01
HHADACSD (CSDxADA)	13,932,820	13,372,132	-560,688	13,932,820	0	13,932,820	0	0.00
HHINC_SADA_GT50	6,966,238	6,936,358	-29,880	6,966,557	319	6,966,487	249	0.00

Table 5.2.1
Census counts and long-form estimate differences and discrepancies – Canada

Characteristic	Census counts	Design weights estimates	Design weights difference	Coverage and non-response adjusted weights estimates	Coverage and non-response adjusted weights difference	Final weights estimates	Final weights difference	Final weights discrepancy (%)
HHINC_SADA_LE50	6,966,582	6,927,519	-39,064	6,966,263	-319	6,966,333	-249	0.00
HHINC_SADA_P25	3,483,439	3,458,080	-25,359	3,482,595	-844	3,482,799	-640	-0.02
HHINC_SADA_P50	3,483,143	3,469,439	-13,704	3,483,667	524	3,483,534	391	0.01
HHINC_SADA_P75	3,483,232	3,472,828	-10,404	3,483,857	625	3,483,764	532	0.02
HHINC_SADA_P100	3,483,006	3,463,530	-19,476	3,482,700	-306	3,482,723	-283	-0.01
HHSIZE1	3,939,604	3,932,270	-7,334	3,938,772	-832	3,937,579	-2,025	-0.05
HHSIZE2	4,801,532	4,788,730	-12,802	4,802,550	1,018	4,803,338	1,806	0.04
HHSIZE3	2,119,446	2,103,306	-16,140	2,119,554	108	2,120,058	612	0.03
HHSIZE4	1,927,188	1,915,750	-11,438	1,928,438	1,250	1,928,027	839	0.04
HHSIZE5	739,505	732,595	-6,910	742,931	3,426	739,636	131	0.02
HHSIZEGE5	1,145,050	1,123,820	-21,230	1,143,506	-1,544	1,143,819	-1,231	-0.11
HHSIZEGE6	405,545	391,225	-14,320	400,575	-4,970	404,183	-1,362	-0.34
INEFAM	28,720,122	28,545,611	-174,511	28,750,644	30,522	28,734,615	14,493	0.05
IR_LINK_NO	33,637,976	33,390,492	-247,484	33,637,564	-412	33,638,859	883	0.00
IR_LINK_YE	356,478	349,615	-6,863	355,801	-677	355,595	-883	-0.25
LIM_NO	29,187,451	28,997,962	-189,489	29,188,145	694	29,188,235	784	0.00
LIM_YE	4,807,003	4,742,145	-64,858	4,805,220	-1,783	4,806,219	-784	-0.02
LONEPAR	1,568,868	1,552,022	-16,846	1,563,594	-5,274	1,565,900	-2,968	-0.19
MALE	16,736,419	16,587,024	-149,395	16,730,785	-5,634	16,736,447	28	0.00
MALEGE15	13,825,590	13,703,085	-122,505	13,819,654	-5,936	13,825,399	-191	0.00
MALELT15	2,910,829	2,883,939	-26,890	2,911,130	301	2,911,048	219	0.01
MARRIED	13,136,486	13,092,169	-44,317	13,135,070	-1,416	13,136,449	-37	0.00
NOCLDFAM	3,998,054	3,993,087	-4,967	3,994,087	-3,967	3,999,849	1,795	0.04
NOINEFAM	5,274,332	5,194,497	-79,835	5,242,721	-31,611	5,259,839	-14,493	-0.27
NOINEFAMH-HSIZEGT1	1,334,728	1,262,226	-72,502	1,303,949	-30,779	1,322,260	-12,468	-0.93
NOTINFAM	6,038,982	5,945,010	-93,972	6,010,343	-28,639	6,030,190	-8,792	-0.15
NOTINFAMH-HSIZEGT1	2,099,378	2,012,740	-86,638	2,071,570	-27,808	2,092,611	-6,767	-0.32
OLN_BI	6,187,606	6,160,319	-27,287	6,187,142	-464	6,187,821	215	0.00
OLN_EN	23,157,389	22,959,992	-197,397	23,156,234	-1,155	23,157,275	-114	0.00
OLN_FR	4,015,036	3,989,439	-25,597	4,015,372	336	4,014,760	-276	-0.01
OLN_NO	634,423	630,357	-4,066	634,618	195	634,598	175	0.03
POBG2_1	14,784	14,945	161	14,798	14	14,841	57	0.38
POBG2_2	7,790	7,381	-409	7,390	-400	7,420	-370	-4.75
POBG2_3	27,825	27,490	-335	27,315	-510	27,564	-261	-0.94
POBG2_4	61,352	61,578	226	61,928	576	61,993	641	1.04
POBG2_5	6,443	6,217	-226	6,259	-184	6,238	-205	-3.18

Table 5.2.1
Census counts and long-form estimate differences and discrepancies – Canada

Characteristic	Census counts	Design weights estimates	Design weights difference	Coverage and non-response adjusted weights estimates	Coverage and non-response adjusted weights difference	Final weights estimates	Final weights difference	Final weights discrepancy (%)
POBG2_6	34,071	33,889	-182	33,847	-224	33,734	-337	-0.99
POBG2_7	68,871	68,511	-360	69,240	369	68,917	46	0.07
POBG2_8	4,069	4,012	-57	4,142	73	4,114	45	1.10
POBG2_9	1,360	1,452	92	1,516	156	1,503	143	10.53
POBG2_10	32,812	32,871	59	32,625	-187	32,781	-31	-0.10
POBG2_11	100,655	100,325	-330	100,502	-153	100,519	-136	-0.14
POBG2_12	7,474	7,156	-318	7,298	-176	7,135	-339	-4.53
POBG2_13	2,230	2,034	-196	2,032	-198	2,039	-191	-8.56
POBG2_14	1,694	1,785	91	1,779	85	1,754	60	3.57
POBG2_15	1,405	1,508	103	1,505	100	1,491	86	6.10
POBG2_16	27,097	26,029	-1,068	26,732	-365	27,016	-81	-0.30
POBG2_17	22,097	22,521	424	22,907	810	22,708	611	2.77
POBG2_18	2,909	2,876	-33	2,896	-13	2,892	-17	-0.58
POBG2_19	22,252	22,054	-198	22,189	-63	22,073	-179	-0.80
POBG2_20	24,130	24,675	545	24,381	251	24,165	35	0.15
POBG2_21	873,285	868,958	-4,327	873,080	-205	873,254	-31	0.00
POBG2_22	109,015	108,270	-745	108,163	-852	108,570	-445	-0.41
POBG2_23	15,319	14,917	-402	15,094	-225	15,262	-57	-0.37
POBG2_24	42,218	42,067	-151	42,520	302	42,632	414	0.98
POBG2_25	14,794	13,853	-941	14,093	-701	14,366	-428	-2.90
POBG2_26	65,533	66,428	895	65,646	113	65,666	133	0.20
POBG2_27	56,296	56,892	596	56,841	545	56,950	654	1.16
POBG2_28	139,455	140,786	1,331	140,543	1,088	140,480	1,025	0.74
POBG2_29	70,138	69,530	-608	70,139	1	69,877	-261	-0.37
POBG2_30	33,952	34,553	601	33,707	-245	33,709	-243	-0.72
POBG2_31	103,283	102,569	-714	103,462	179	102,855	-428	-0.41
POBG2_32	186,698	186,575	-123	187,555	857	187,251	553	0.30
POBG2_33	112,041	110,677	-1,364	111,862	-179	111,726	-315	-0.28
POBG2_34	19,235	19,630	395	19,665	430	19,480	245	1.27
POBG2_35	72,931	72,295	-636	73,132	201	72,782	-149	-0.20
POBG2_36	74,350	73,931	-419	74,129	-221	74,570	220	0.30
POBG2_37	112,543	111,512	-1,031	112,225	-318	112,106	-437	-0.39
POBG2_38	86,917	87,053	136	87,116	199	87,155	238	0.27
POBG2_39	28,117	27,361	-756	27,542	-575	27,852	-265	-0.94
POBG2_40	124,170	123,731	-439	124,693	523	124,738	568	0.46
POBG2_41	2,234	2,204	-30	2,240	6	2,258	24	1.08
POBG2_42	154,606	153,966	-640	155,098	492	154,564	-42	-0.03
POBG2_43	609,741	607,631	-2,110	609,653	-88	609,822	81	0.01
POBG2_44	2,761	3,017	256	2,965	204	2,933	172	6.22
POBG2_45	60,546	60,919	373	60,721	175	60,639	93	0.15
POBG2_46	51,880	51,081	-799	51,305	-575	51,651	-229	-0.44

Table 5.2.1
Census counts and long-form estimate differences and discrepancies – Canada

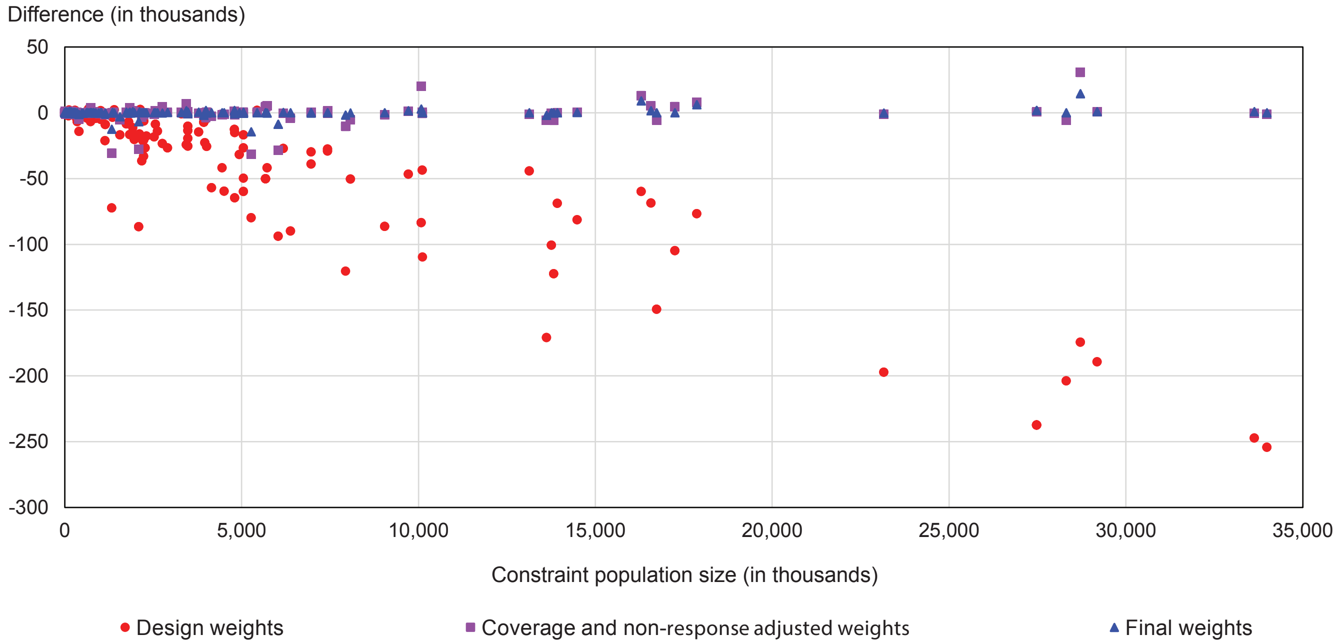
Characteristic	Census counts	Design weights estimates	Design weights difference	Coverage and non-response adjusted weights estimates	Coverage and non-response adjusted weights difference	Final weights estimates	Final weights difference	Final weights discrepancy (%)
POBG2_47	125,745	124,733	-1,012	125,786	41	125,600	-145	-0.12
POBG2_48	348,474	348,730	256	348,747	273	348,490	16	0.00
POBG2_49	1,342	1,500	158	1,466	124	1,515	173	12.87
POBG2_50	1,044	1,064	20	1,068	24	1,078	34	3.23
POBG2_51	14,222	13,744	-478	13,803	-419	13,902	-320	-2.25
POBG2_53	266,119	265,374	-745	264,686	-1,433	265,339	-780	-0.29
POBG2_54	115,745	114,997	-748	115,501	-244	115,690	-55	-0.05
POBG2_55	20,529	20,181	-348	20,361	-168	20,472	-57	-0.28
POBG2_56	74,226	74,585	359	74,351	125	74,304	78	0.10
POBG2_57	27,373	28,272	899	28,058	685	27,813	440	1.61
POBG2_58	13,459	13,286	-173	13,275	-184	13,306	-153	-1.13
POBG2_59	40,929	40,823	-106	40,995	66	40,986	57	0.14
POBG2_60	123,382	125,358	1,976	123,900	518	123,828	446	0.36
POBG2_61	13,814	13,787	-27	13,814	0	13,816	2	0.02
POBG2_62	14,021	14,015	-6	14,103	82	13,920	-101	-0.72
POBG2_63	9,250	9,276	26	9,190	-60	8,962	-288	-3.11
POBG2_64	880,611	880,776	165	880,196	-415	880,351	-260	-0.03
POBG2_65	31,092	30,739	-353	30,820	-272	30,852	-240	-0.77
POBG2_66	292,617	294,336	1,719	292,849	232	292,496	-121	-0.04
POBG2_67	172,316	170,704	-1,612	172,312	-4	172,232	-84	-0.05
POBG2_68	21,518	21,146	-372	21,201	-317	21,317	-201	-0.94
POBG2_69	70,746	69,813	-933	70,611	-135	70,649	-97	-0.14
POBG2_70	38,949	37,477	-1,472	37,941	-1,008	38,369	-580	-1.49
POBG2_71	105,555	105,197	-358	105,258	-297	105,289	-266	-0.25
POBG2_M3	27,477,998	27,240,476	-237,522	27,478,602	604	27,479,831	1,833	0.01
POBG3_2	315,727	313,684	-2,043	315,604	-123	315,346	-381	-0.12
POBG3_3	52,168	51,502	-666	52,616	448	52,702	534	1.02
POBG3_4	144,752	144,355	-397	145,063	311	144,824	72	0.05
POBG3_5	172,246	173,386	1,140	171,963	-283	172,386	140	0.08
POBG3_6	1,026,914	1,021,550	-5,364	1,026,782	-132	1,027,359	445	0.04
POBG3_7	258,618	257,390	-1,228	258,524	-94	258,841	223	0.09
POBG3_8	526,099	524,544	-1,555	525,838	-261	525,741	-358	-0.07
POBG3_9	224,749	223,501	-1,248	225,241	492	224,445	-304	-0.14
POBG3_10	306,076	307,622	1,546	306,124	48	305,802	-274	-0.09
POBG3_12	48,365	47,675	-690	47,680	-685	48,041	-324	-0.67
POBG3_13	263,337	262,581	-756	261,909	-1,428	262,571	-766	-0.29
POBG3_14	288,406	288,209	-197	288,243	-163	288,093	-313	-0.11
POBG3_15	817,934	815,377	-2,557	818,608	674	818,128	194	0.02
POBG3_16	42,289	42,275	-14	42,511	222	42,489	200	0.47
POBG3_17	1,022,321	1,023,931	1,610	1,022,082	-239	1,022,232	-89	-0.01
POBG3_18	298,701	298,685	-16	298,557	-144	298,837	136	0.05
POBG3_19	208,492	209,073	581	209,002	510	208,606	114	0.05

Table 5.2.1
Census counts and long-form estimate differences and discrepancies – Canada

Characteristic	Census counts	Design weights		Coverage and non-response adjusted weights		Final weights		
		estimates	difference	estimates	difference	estimates	difference	discrepancy (%)
POBG3_21	112,196	109,730	-2,466	111,021	-1,175	111,510	-686	-0.61
POBG3_22	212,520	211,823	-697	213,051	531	212,398	-122	-0.06
PPADACSD (CSDxADA)	33,994,454	32,767,748	-1,226,706	33,993,365	-1,089	33,994,454	0	0.00
SNGLDET	7,434,272	7,406,572	-27,700	7,435,570	1,298	7,434,328	56	0.00
TOTCFAM	9,720,966	9,674,172	-46,794	9,722,126	1,160	9,722,502	1,536	0.02
TOTHHLD	13,932,820	13,863,876	-68,944	13,932,820	0	13,932,820	0	0.00
TOTPERS	33,994,454	33,740,107	-254,347	33,993,365	-1,089	33,994,454	0	0.00
TPERGE15	28,314,895	28,110,924	-203,971	28,309,127	-5,768	28,314,932	37	0.00
YRIMD_1900	1,038,231	1,038,998	767	1,037,883	-348	1,038,201	-30	0.00
YRIMD_1981	273,907	275,157	1,250	274,631	724	274,167	260	0.09
YRIMD_1986	491,794	491,464	-330	491,411	-383	491,648	-146	-0.03
YRIMD_1991	726,588	725,927	-661	726,496	-92	726,876	288	0.04
YRIMD_1996	680,866	683,744	2,878	681,300	434	680,314	-552	-0.08
YRIMD_2001	849,532	846,726	-2,806	849,728	196	849,356	-176	-0.02
YRIMD_2006	961,099	956,822	-4,277	960,315	-784	960,960	-139	-0.01
YRIMD_2011	1,150,540	1,141,489	-9,051	1,150,015	-525	1,150,117	-423	-0.04
YRIMD_M5	343,899	339,304	-4,595	342,984	-915	342,983	-916	-0.27
YRIMG1_1981	765,701	766,620	919	766,042	341	765,815	114	0.01
YRIMG1_1991	1,407,454	1,409,672	2,218	1,407,796	342	1,407,190	-264	-0.02
YRIMG1_2001	1,810,631	1,803,548	-7,083	1,810,043	-588	1,810,317	-314	-0.02

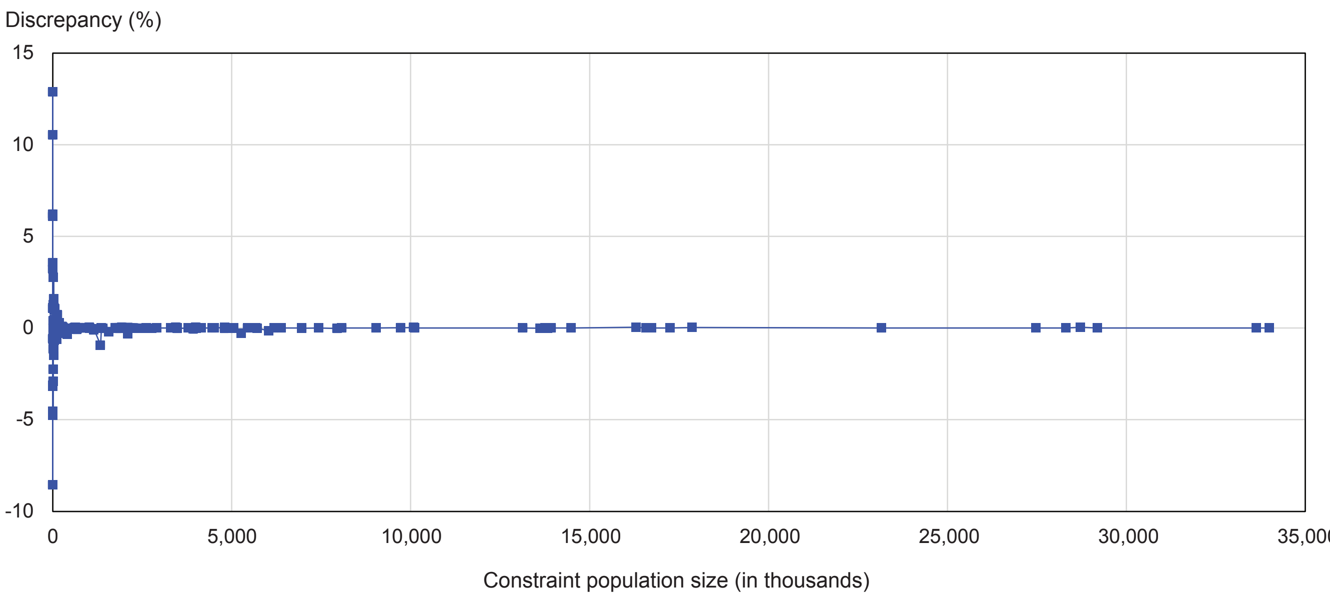
Notes: Estimates are based on variables related to knowledge of official languages—OLN_BI, OLN_EN and OLN_FR—and their weights were calculated using the responses from before the language data update. A study showed that the use of these unadjusted variables as calibration constraints has a negligible impact on the quality of the weighted estimates. For further information on this impact, see Appendix 1.8 of the *Guide to the Census of Population, 2016* (http://www12.statcan.gc.ca/census-recensement/2016/ref/98-304/app-ann1-8-eng.cfm).
All households with a design weight of one are excluded from the weighting process. These households either come from Indian reserve and canvasser enumeration CUs or are private households attached to collective.
Source: Statistics Canada, 2016 Census long-form sample.

Chart 5.2.2
Differences between census counts and design-weighted, coverage and non-response adjusted, and final estimates



Note: All households with a design weight of 1 were excluded from the weighting process. These households either were located in an Indian reserve or a canvasser enumeration collection unit or were private households attached to a collective dwelling.
Source: Statistics Canada, 2016 Census long-form sample.

Chart 5.2.3
Discrepancy between census counts and final estimates, as a percentage of census counts



Note: All households with a design weight of 1 were excluded from the weighting process. These households either were located in an Indian reserve or a canvasser enumeration collection unit or were private households attached to a collective dwelling.
Source: Statistics Canada, 2016 Census long-form sample.

6. Variance estimation

The error in an estimate is the difference between the estimate and the actual value of what is being estimated. There are several sources of error in the long-form sample survey, including sampling error and total non-response error. Sampling error stems from the fact that the estimates are based on observations from a sample and not from the Census of Population. Total non-response error occurs when households selected in the sample do not respond to the survey.

Error has a random component, measured by variance, and a systematic component, measured by bias. Variance measures how much the estimate varies from the average that would result from hypothetical repetitions of the survey process. Variance can be estimated using data from the sample. Bias is the difference between the average estimate that would result from hypothetical repetitions of the survey process and the actual value of the characteristic being estimated. The sampling and estimation methods used in the long-form sample survey produce a negligible bias.

Some estimation methods are more precise than others in estimating a particular characteristic of the population, so they can affect error. The estimated variance can be used to produce several quality indicators that are often used to measure the accuracy of an estimate. For example, it can be used to calculate standard errors, confidence intervals and coefficients of variation. Standard error was the quality indicator produced for the 2016 Census long-form estimates. Standard error corresponds to the square root of the variance.

These measures of variability must be carefully distinguished from other measures of quality that are not, strictly speaking, measures of variability. Examples of such measures are the final response rates presented in section 3.11 and the global non-response rate to the 2016 Census long-form questionnaire. The response rate is an indicator of the risk associated with household non-response error. The global non-response rate is an indicator of the risk of error attributable to household non-response and item non-response. For more information, see Chapter 11 (http://www12.statcan.gc.ca/census-recensement/2016/ref/98-304/chap11-eng.cfm) of the *Guide to the Census of Population, 2016* (Statistics Canada 2017).

Since the long-form sample is geographically stratified into take-some strata (mail-out and list/leave CUs) and take-all strata (canvasser CUs), two variance estimators are used. The first variance estimator is used to estimate the variance in take-some geographic areas (see section 6.3.1), and the second estimator is used to estimate the variance attributable to total non-response in take-all areas (see section 6.3.2). For the remainder of this chapter, the term variance is used to designate the sampling and total non-response variance in take-some geographic areas or the total non-response variance in take-all areas.

Standard errors for various estimates and geographic areas can be viewed or downloaded from the Census Profile Standard Error Supplement, Canada, provinces, territories, census divisions (CDs) and aggregate dissemination areas (ADAs), 2016 Census (http://www12.statcan.gc.ca/census-recensement/2016/dp-pd/adaprof/index.cfm?Lang=E&TABID=0). The sampling and estimation methods used for the 2016 long-form survey are different from those used for the 2011 National Household Survey (NHS). Since the size and impact of non-response were very different in 2011, they affected estimate variability. For more information, refer to the methodological note comparing standard errors for 2016 estimates with those for 2011 and 2006 estimates (http://www12.statcan.gc.ca/census-recensement/2016/ref/se-et-eng.cfm) for some variables.

6.1 Elements to consider in choosing a variance estimation method

A very high number of diverse estimates were produced, and quality indicators for these estimates needed to be established within a reasonable timeframe. As a result, a resampling variance estimator was used, derived from the modified partially balanced repeated replication method (Judkins 1990). The method consists of drawing samples (or replicates) from the original sample. Weights are calculated for each replicate, and the weights undergo the same coverage, non-response and calibration adjustments as the original sample. The resulting weights are called replicate weights. Estimates are then produced for each replicate, and the variance is estimated using replicate estimates and the average of these estimates.

Figure 6.1.1 gives an overview of replication variance estimation when R samples are used.

Figure 6.1.1
Overview of replication variance estimation

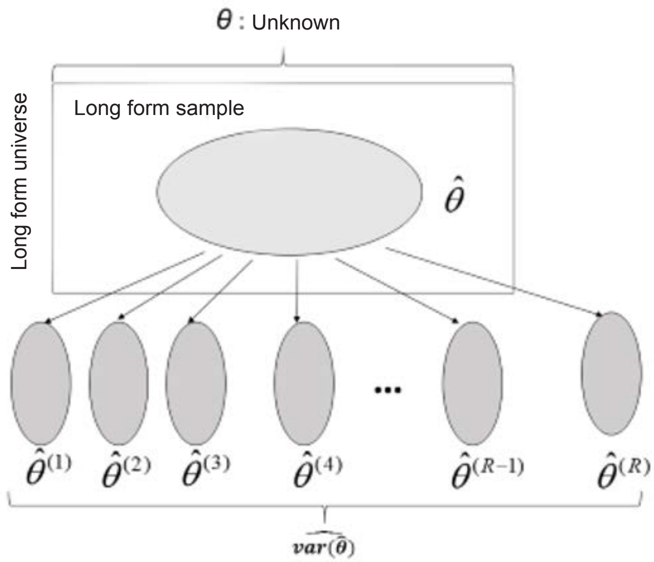


Figure 6.1.1 gives an overview of the replication variance estimation methodology used in the 2016 Census. The replication variance estimation method simulates the selection of several samples to estimate sampling variance.

More specifically, the figure shows the long-form questionnaire universe representing the population of interest and the long-form questionnaire sample. The sample is situated within the universe to indicate whether it corresponds to a subset of the population of interest. This sample is used to estimate a characteristic of the population of interest, such as the number of people who are members of a visible minority. The theta symbol is used to represent the true value of this characteristic. A circumflex on the theta indicates that the value is an estimate of this characteristic. This value is known as theta hat.

The R other samples placed outside the universe are linked to the long-form questionnaire sample with arrows. The arrows indicate that these samples are taken from the long-form questionnaire sample. The characteristic of interest is re-estimated based on these R sub-samples. The theta hat R values, referred to as theta hat one, theta hat two, up to theta hat R, are used to calculate the estimated theta hat variance.

In this figure, we define:

- theta, θ , the true value of the characteristic in the population, which can be a total, an average, a quantile, etc.;
- theta hat, $\hat{\theta}$, the value of θ estimated using the main weights;
- theta hat *r*, $\hat{\theta}^{(r)}$, the value of θ estimated using the replication weights *r*, *r*=1, ..., R;
- theta hat bar, $\bar{\hat{\theta}}$, the average value of the *R* replication estimates $\hat{\theta}^{(r)}$ and
- $\hat{\mathrm{var}}(\hat{\theta})$, the estimated value of the variance of $\hat{\theta}$.

6.2 Variance estimator

The replicate estimator chosen for the long-form sample survey was derived from Fay's balanced half-sample method (Judkins 1990). This method determines the creation of replicates, the calculation of replicate weights and the multiplication factor used to estimate variance.

To produce variance estimates for the long-form sample estimates, two sets of replicate weights were created: the first had 32 replicate weights and the second had 100 replicate weights. The set of 32 replicate weights was produced to estimate the standard errors of standard products that are calculated under operational constraints (i.e., the need to publish a large number of standard errors within a reasonable timeframe). The set of 100 replicate weights was made available to Statistics Canada analysts, research data centre analysts who have access to microdata and users who request custom products, to provide more precise variance estimators.

The replication variance estimator can be calculated in two ways, one of which is more conservative than the other. The first method consists of adding the squared differences between the replication estimates, $\hat{\theta}^{(r)}$, and either the average of the replication estimates, $\bar{\hat{\theta}}$, or the estimate from the primary sample, $\hat{\theta}$, and multiplying this sum by a certain factor. The second method, which uses the estimate from the primary sample, is more conservative. In the Computer-assisted Product Specification System used to publish statistics, the variance estimator is calculated using the average of the replication estimates.

For example, two variance estimators to estimate a total from a set R of replicates are given in the equations below:

$$\hat{\mathrm{Var}}_1\left(\hat{T}\right) = \frac{1}{R/2}\sum_{r=1}^{R}\left(\hat{T}^{(r)} - \bar{\hat{T}}\right)^2 ,$$

$$\hat{\mathrm{Var}}_2\left(\hat{T}\right) = \frac{1}{R/2}\sum_{r=1}^{R}\left(\hat{T}^{(r)} - \hat{T}\right)^2$$

where

$$\hat{T} = \sum\nolimits_{k \in s} w_k y_k ,$$

$$\hat{T}^{(r)} = \sum\nolimits_{k \in s} w_k^{(r)} y_k \text{ , and}$$

$$\bar{\hat{T}} = \sum\nolimits_{r=1}^{R} \hat{T}^{(r)} \big/ \mathrm{R} .$$

The final weight of the sample is represented by w_k, $w_k^{(r)}$ is the final weight of replicate r, y_k is the value of characteristic y for unit *k*, and s is the long-form sample.

The number of degrees of freedom of the variance estimator is approximated by the number of squared differences $\left(\hat{T}^{(r)} - \bar{\hat{T}}\right)^2$ for the variance estimator, i.e., 32 or 100. The number of degrees of freedom gives an idea of the precision of the variance estimator and is used in calculating confidence intervals for long-form estimates. Since the method used to estimate the variance of the estimate can use few replicates, Student's law of probability was used to determine the quantile of the confidence level to be used in calculating the interval. In cases where the number of households participating in the estimate is larger than R, the number of degrees of freedom should be R. In cases where the number of households is less than R but equal to or greater than 20, the number of degrees of freedom corresponds to the number of households. Producing confidence intervals when there are fewer than 20 households is not recommended.

6.3 Replicate weight adjustment

6.3.1 Mail-out and list/leave collection units

As mentioned in section 6.2, replicate weights were calculated for all long-form sample households. The replicates were partially balanced. They were balanced by resampling strata, which were created by combining CUs to obtain 600 to 1,800 households per resampling stratum.

Fay's modified balanced half-sample method, as described by Rao and Shao (1999), requires an epsilon value in the calculation of replicate weights to control the perturbation of the replicate weights. This perturbation results in all sampled households participating in every replicate, unlike other more popular replication methods. This facilitates the calibration of the replicate weights and, occasionally, the calculation of point estimates for each replicate (e.g., the denominator of a ratio estimator for a given replicate will not have a nil value if the corresponding denominator was not nil with the final weight). Adding an epsilon factor to the calculation of replicate weights meant the large survey fraction used to select the long-form sample could be taken into account. The technical details of the variance estimation process were provided by Devin and Verret (2016).

The replicate weights underwent the same adjustments as the primary sample design weight. They were adjusted for coverage and total non-response following the same methodology that was used for the primary sample weight (see section 4.4). The resulting replicate weights were then calibrated to census counts, once again following the same methodology that was used for the primary weight (see section 4.5).

6.3.2 Indian reserve[8] and canvasser enumeration collection units

As described in Chapter 2, all the households in Indian reserve and canvasser enumeration CUs were selected with certainty. As such, they originally had a design weight equal to 1. A coverage adjustment was not needed. All these households were selected for the long form, and therefore differential coverage between the short and long form could not occur. Total non-response in these areas was treated with the process of whole household imputation (WHI), described in Chapter 3. In other words, the data of a non-responding household were replaced by the data of another responding household from the same CU (except for geography variables, which were known for non-respondents). As a consequence, reweighting for households in Indian reserve and canvasser enumeration CUs was not needed.

Calibration was not needed in these areas, because the long form was a census. Consequently, all households in Indian reserve and canvasser enumeration CUs maintained their original weight equal to 1 in the final weighting scheme. For more information on incompletely enumerated Indian reserves and Indian settlements, please refer to Appendix 1.2 (http://www12.statcan.gc.ca/census-recensement/2016/ref/98-304/app-ann1-2-eng.cfm) of the *Guide to the Census of Population, 2016*, Catalogue no. 98-304-X.

Although sampling variability did not occur for households in Indian reserve and canvasser enumeration CUs, WHI variability did occur. Variance estimation in these areas was computed using a similar method to that of the rest of the country, with the following exceptions. First, the response probability by household size combination in each census division was estimated as the number of responding households divided by the number of in-scope households. Then the base replicate weights were created as in the rest of the country, except that all respondents for which the response probability was equal to 1 were placed in every replicate. Respondents with estimated response probabilities less than 1 were not considered certainties and were treated as sampled elements (i.e., they were randomly divided among the replicates). Non-respondent households imputed by WHI were also divided among replicates, and each one was assigned the replicate inclusion indicator corresponding to its donor in a manner similar to that of Shao and Tang (2011). This caused the weights to vary from one replicate to the other

8. The exception to this characteristic was the units in incompletely enumerated Indian reserves and Indian settlements, which were excluded from the target population and whose weight was set to 0, without any further modification to the dataset or weights.

instead of the values. Finally, the replicate weights were calibrated to the number of households and number of persons in the SADA. As a result, the estimated variance of those two quantities was equal to 0 at the SADA level and at more aggregate levels, such as Canada (since those two constraints are mandatory in the rest of the country).

7. Conclusion

The reintroduction of the mandatory long form allowed for several desirable characteristics in terms of sampling and weighting. Mainly, the sampling design was greatly simplified compared with that of the 2011 National Household Survey (NHS) (see Statistics Canada, 2015 for more details on the 2011 NHS sample design). As a direct result, the weighting procedure was also simplified, and the final weights were less variable than those of the 2011 NHS. Additionally, the mandatory nature of the 2016 long form presumably increased the response rate. Reducing the amount of non-response is always important, as this may have a direct effect on the bias of estimates. Therefore, minimizing the non-response rate is important to keep the non-response bias as small as possible.

The combination of the one-in-four sampling fraction and the overall unweighted final response rate of 96.1% meant that the 2016 long-form sample obtained more responses overall than the 2011 NHS (which had a one-in-three sampling fraction and a response rate of 68.6%). The response rates across geographic areas were also more homogeneous. The much higher and more uniform response rate enabled the simplification of the weighting procedures.

All these features, coupled with an improved calibration procedure, meant that the maximum difference between census counts and final-weight estimates (14,493) was about 1/10 of the maximum difference for the 2011 NHS (166,801). The improvement in the calibration procedure was a mixture of choosing fewer calibration variables, larger weighting areas (i.e., SADAs) and a better selection of constraints (such as the prioritization of constraints that split the population approximately in half).

As a consequence of these improvements, the range of the final weights was reduced (compared with the 2011 NHS). This resulted in the reduction of the standard error for most characteristics. For the variables related to the calibration constraints that were selected often, the variance was reduced by virtue of the calibration method. For variables not highly related to the calibration constraints, the variance was still moderated by the limitation of the range of the weights.

Finally, the release of replicate weights implies that users can easily estimate the standard error for any particular variable of interest. For the 2011 NHS, the chosen method of variance estimation—Taylor series linearization combined with a corrective upward adjustment factor (from Monte Carlo simulation)—made computing standard errors for non-standard products difficult.

Appendix A – Glossary

The definitions of the main census terms, variables and concepts mentioned in this document are presented here. Users can also 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 for additional information.

Aggregate dissemination area (ADA) (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo053-eng.cfm): The aggregate dissemination area (ADA) is a new dissemination geography created for the 2016 Census. ADAs cover the entire country and, where possible, have a population between 5,000 and 15,000 based on the previous census population counts. ADAs are created from existing dissemination geographic areas and are formed from census tracts (CTs), census subdivisions (CSDs) or dissemination areas (DAs). ADAs respect provincial, territorial, census division (CD), census metropolitan area (CMA) and census agglomeration (CA) with census tract (CT) boundaries.

The intent of the ADA geography is to ensure the availability of census data, where possible, across all regions of Canada.

Census division (CD) (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo008-eng.cfm): Group of neighbouring municipalities joined together for the purposes of regional planning and managing common services (such as police or ambulance services). These groupings are established under laws in effect in certain provinces of Canada. Census division (CD) is the general term for provincially legislated areas (such as county, *municipalité régionale de comté* and regional district) or their equivalents. In other provinces and the territories where laws do not provide for such areas, Statistics Canada defines equivalent areas for statistical reporting purposes in co-operation with these provinces and territories. Census divisions are intermediate geographic areas between the province or territory level and the municipality (census subdivision).

Census family (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/fam004-eng.cfm): “Census family” is defined as a married couple and the children, if any, of either and/or both spouses; a couple living common law and the children, if any, of either and/or both partners; or a lone parent of any marital status with at least one child living in the same dwelling and that child or those children. All members of a particular census family live in the same dwelling. A couple may be of opposite or same sex. Children may be children by birth, marriage, common-law union or adoption regardless of their age or marital status as long as they live in the dwelling and do not have their own married spouse, common-law partner or child living in the dwelling. Grandchildren living with their grandparent(s) but with no parents present also constitute a census family.

Census subdivision (CSD) (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo012-eng.cfm): Census subdivision (CSD) is the general term for municipalities (as determined by provincial or territorial legislation) or areas treated as municipal equivalents for statistical purposes (e.g., Indian reserves, Indian settlements and unorganized territories). Municipal status is defined by laws in effect in each province and territory in Canada.

Census tract (CT) (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo013-eng.cfm): Census tracts (CTs) are small, relatively stable geographic areas that usually have a population of less than 10,000 persons, based on data from the previous Census of Population Program. They are located in census metropolitan areas and in census agglomerations that had a core population of 50,000 or more in the previous census.

A committee of local specialists (for example, planners, health and social workers, and educators) initially delineates census tracts in conjunction with Statistics Canada. Once a census metropolitan area (CMA) or census agglomeration (CA) has been subdivided into census tracts, the census tracts are maintained even if the core population subsequently declines below 50,000.

Collection unit (CU):[9] Collection units are small geographic units used for the collection of census data. Collection units cover all the territory of Canada.

Collective dwelling (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/dwelling-logements002-eng.cfm): Refers to a dwelling of a commercial, institutional or communal nature. It may be identified by a sign on the premises or by an enumerator speaking with the person in charge, a resident, a neighbour, etc. Included are lodging or rooming houses, hotels, motels, tourist establishments, nursing homes, hospitals, staff residences, military bases, work camps, jails, group homes, and so on.

Dissemination area (DA) (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo021-eng.cfm): A dissemination area (DA) is a small, relatively stable geographic unit composed of one or more adjacent dissemination blocks with an average population of 400 to 700 persons based on data from the previous Census of Population Program. It is the smallest standard geographic area for which all census data are disseminated. DAs cover all the territory of Canada.

Dwelling (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/dwelling-logements004-eng.cfm): A dwelling is defined as a set of living quarters. Two types of dwellings are identified in the census, collective dwellings and private dwellings. The former pertains to dwellings which are institutional, communal or commercial in nature. The latter, private dwellings, refers to a separate set of living quarters with a private entrance either from outside the building or from a common hall, lobby, vestibule or stairway inside the building. The entrance to the dwelling must be one that can be used without passing through the living quarters of some other person or group of persons.

Economic family (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/fam011-eng.cfm): "Economic family" refers to a group of two or more persons who live in the same dwelling and are related to each other by blood, marriage, common-law union, adoption or a foster relationship. A couple may be of opposite or same sex. By definition, all persons who are members of a census family are also members of an economic family. Examples of the broader concept of economic family include the following: two co-resident census families who are related to one another are considered one economic family; co-resident siblings who are not members of a census family are considered as one economic family; and, nieces or nephews living with aunts or uncles are considered one economic family.

Household (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/households-menage007-eng.cfm): Household refers to a person or group of persons who occupy the same dwelling and do not have a usual place of residence elsewhere in Canada or abroad. The dwelling may be either a collective dwelling or a private dwelling. The household may consist of a family group such as a census family, of two or more families sharing a dwelling, of a group of unrelated persons or of a person living alone. Household members who are temporarily absent on reference day are considered part of their usual household.

Private dwelling (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/dwelling-logements005-eng.cfm): "Private dwelling" refers to a separate set of living quarters with a private entrance either from outside the building or from a common hall, lobby, vestibule or stairway inside the building. The entrance to the dwelling must be one that can be used without passing through the living quarters of some other person or group of persons.

The dwelling must meet the two conditions necessary for year-round occupancy:

1. a source of heat or power (as evidenced by chimneys, power lines, oil or gas pipes or meters, generators, woodpiles, electric lights, heating pumps or solar panels)
2. an enclosed space that provides shelter from the elements, as evidenced by complete and enclosed walls and roof, and by doors and windows that provide protection from wind, rain and snow.

9. This definition is not found in the *Dictionary, Census of Population, 2016*, because the dictionary consists mainly of dissemination terms and this is a collection term.

Dwellings that do not meet the conditions necessary for year-round occupancy are marginal dwellings. Private dwellings are classified into regular private dwellings and occupied marginal dwellings. Regular private dwellings are further classified into three major groups: occupied dwellings (occupied by usual residents), dwellings occupied solely by foreign residents and/or by temporarily present persons, and unoccupied dwellings. Marginal dwellings are classified as occupied by usual residents or occupied solely by foreign residents and/or by temporarily present persons. Marginal dwellings that were unoccupied on May 10, 2016, are not counted in the housing stock.

Private dwelling occupied by usual residents (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/dwelling-logements006-eng.cfm): Refers to a private dwelling in which a person or a group of persons is permanently residing. Also included are private dwellings whose usual residents are temporarily absent on May 10, 2016. Unless otherwise specified, all data in housing products are for private dwellings occupied by usual residents, rather than for unoccupied private dwellings or dwellings occupied solely by foreign and/or by temporarily present persons.

Private household (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/households-menage014-eng.cfm): "Private household" refers to a person or group of persons who occupy the same dwelling and do not have a usual place of residence elsewhere in Canada or abroad. The household universe is divided into two sub-universes on the basis of whether the household is occupying a collective dwelling or a private dwelling. The latter is a private household.

For census purposes, households are classified into three groups: private households, collective households and households outside Canada.

Unless otherwise specified, all data in census products are for private households only.

Super aggregate dissemination area (SADA):[10] Super aggregate dissemination areas are created specifically for weighting 2016 Census data. They respect pre-established rules, some of which are mandatory and others optional. SADAs are created by combining aggregate dissemination areas (ADAs) and are contained within provincial and territorial boundaries. All individuals living in census collection units (CUs) are excluded from the SADA population. SADAs are created, inasmuch as possible, with a target population of between 50,000 and 150,000 in mind. Census divisions (CD) with a population of 40,000 to 50,000 comprise their own SADA. In addition, where possible, SADAs respect the boundaries of—in order of priority—census division (CDs), census metropolitan areas (CMAs),census agglomerations (CAs) and census subdivisions (CSD). Lastly, SADAs should be created by combining adjacent ADAs (where possible) and must be as compact as possible.

10. This definition is not found in the *Dictionary, Census of Population, 2016*, because the dictionary consists mainly of dissemination terms and this is a collection term.

Appendix B – The history of sampling in the Canadian census

Sampling was first used in the Canadian census in 1941. A housing schedule was completed for every 10th dwelling. The information from 27 questions on the separate housing schedule was integrated with the data in the personal and household section of the population schedule for the same dwelling. This enabled cross-tabulation of sample and basic characteristics. Also in the 1941 Census, sampling was used at the processing stage to obtain early estimates of earnings of wage-earners, of the distribution of the population of working age and of the composition of families in Canada. In this case, a sample of every 10th enumeration area across Canada was selected and all population schedules in these areas were processed in advance.

The census of housing was again conducted on a sample basis in 1951. This time, every fifth dwelling (those whose identification numbers ended in a 2 or 7) was selected to complete a housing document containing 24 questions. In the 1961 Census, persons aged 15 years and older in a 20% sample of private households were required to complete a population sample questionnaire containing questions on internal migration, fertility and income. Sampling was not used in the smaller censuses of 1956 and 1966.

The 1971 Census saw several major innovations in the method of census-taking. The primary change was from the traditional canvasser method of enumeration to the use of self-enumeration for the majority of the population. This change was prompted by the results of several studies in Canada and elsewhere (Fellegi 1964; Hansen et al. 1959), which indicated that the effect of the enumerator was a major contribution to the variance of census figures in a canvasser census. Consequently, the use of self-enumeration was expected to reduce the variance of census figures by reducing the effect of the enumerator and by giving the respondent more time and privacy in which to answer the census questions—factors that might be expected to yield more accurate responses.

The second aspect of the 1971 Census that differentiated it from any earlier census was its content. The number of topics covered and the number of questions asked were greater than in any previous census. Considerations of cost, respondent burden and timeliness versus the level of data quality to be expected using self-enumeration and sampling led to a decision to collect all but certain basic characteristics on a one-third sample basis in the 1971 Census. In all but the more remote areas of Canada, every third private household received the “long questionnaire,” which contained all the census questions. The remaining private households received the “short questionnaire” containing only the basic questions covering name, relationship to head of household, sex, date of birth, marital status, mother tongue, type of dwelling, tenure, number of rooms, water supply, toilet facilities and certain census coverage items. All households in pre-identified remote enumeration areas and all collective dwellings received the long questionnaire. A more detailed description of the consideration of the use of sampling in the 1971 Census is given in *Sampling in the Census* (Dominion Bureau of Statistics 1968).

The 1976 Census had considerably less content than the 1971 Census. Furthermore, the 1976 questionnaire did not include the questions that cause the most difficulty in collection (e.g., income) or that are costly to code (e.g., occupation, industry and place of work). Therefore, the benefits of sampling in terms of cost savings and reduced respondent burden were less clear than for the 1971 Census. Nevertheless, after estimating the potential cost savings to be expected with various sampling fractions and considering the public relations issues related to a reversion to 100% enumeration after a successful application of sampling in 1971, Statistics Canada decided to use the same sampling procedure in 1976 as in 1971.

Most of the methodology used in the 1971 and 1976 censuses was kept for the 1981 Census, except that the sampling rate was reduced from every third occupied private household to every fifth. Studies done at the time showed that the resulting reduction in data quality (measured in terms of variance) would be tolerable and would not be significant enough to offset the benefits of reduced cost and respondent burden and improved timeliness (Royce 1983). The one-in-five sampling rate was maintained for every census from 1981 to 2006.

In 2011, information previously collected by the mandatory long-form census questionnaire was collected on a voluntary basis, via the National Household Survey (NHS). With this change, every household was required to answer the 10 questions that were contained in the 2011 Census questionnaire, while 30% of households were selected to respond to the NHS. As well, NHS non-responding households were subsampled for follow-up at a rate of one in three. The increased sampling fraction to 30% was implemented in anticipation of a lower response rate to the NHS. For the 2016 Census, the government reinstated the census long-form questionnaire as mandatory, replacing the NHS. The sampling fraction was changed in 2016 to one in four, compared with one in five for the previous census long-form questionnaire in 2006, to mitigate the risk of the response rate not recovering to its previously high levels.

Appendix C – Constraints used in or excluded from the weighting process[11]

The following is a list of the possible constraints defined at the ADA and SADA levels. A total of 270 possible constraints were defined at the ADA level and 200 at the SADA level. Any constraints that were not defined at either the ADA level or the SADA level are not taken into account in the table below. This is why this table does not contain 4,283 or 4,180 ADAs and 408 SADAs, like previous tables.

Table C.1
Statistics on the use of calibration constraints, by constraint

		Coverage and non-response adjustment			Final calibration		
Variable/ constraint	Description	Area	Number of calibrated constraints	Number of excluded constraints	Area	Number of calibrated constraints	Number of excluded constraints
ADULTCF	Adults in a census family	SADA	105	302	SADA	138	270
AGE00_14	Persons aged 0 to 14 years	SADA	288	120	BOTH	4,473	115
AGE4	Persons aged 0 to 4 years	SADA	70	338	BOTH	868	3,720
AGE9	Persons aged 5 to 9 years	SADA	64	344	BOTH	836	3,752
AGE14	Persons aged 10 to 14 years	SADA	52	356	BOTH	820	3,768
AGE15_24	Persons aged 15 to 24 years	SADA	310	98	SADA	390	18
AGE15_29	Persons aged 15 to 29 years	...	...	...	ADA	4,084	96
AGE19	Persons aged 15 to 19 years	SADA	268	140	BOTH	822	3,766
AGE24	Persons aged 20 to 24 years	SADA	252	156	BOTH	885	3,703
AGE25_34	Persons aged 25 to 34 years	SADA	373	35	SADA	401	7
AGE29	Persons aged 25 to 29 years	SADA	324	84	BOTH	1,406	3,182
AGE30_49	Persons aged 30 to 49 years	...	...	...	ADA	4,148	32
AGE34	Persons aged 30 to 34 years	SADA	348	60	BOTH	1,575	3,013
AGE35_44	Persons aged 35 to 44 years	SADA	360	48	SADA	408	0
AGE39	Persons aged 35 to 39 years	SADA	369	39	BOTH	1,455	3,133
AGE44	Persons aged 40 to 44 years	SADA	389	19	BOTH	1,384	3,204

11. See Appendix D for the Wood Buffalo constraints.

Table C.1
Statistics on the use of calibration constraints, by constraint

Variable/ constraint	Description	Coverage and non-response adjustment			Final calibration		
		Area	Number of calibrated constraints	Number of excluded constraints	Area	Number of calibrated constraints	Number of excluded constraints
AGE45_54	Persons aged 45 to 54 years	SADA	406	2	SADA	408	0
AGE49	Persons aged 45 to 49 years	SADA	403	5	BOTH	1,656	2,932
AGE50_64	Persons aged 50 to 64 years	...	...	...	ADA	4,026	154
AGE54	Persons aged 50 to 54 years	SADA	405	3	BOTH	1,918	2,670
AGE55_64	Persons aged 55 to 64 years	SADA	390	18	SADA	402	6
AGE59	Persons aged 55 to 59 years	SADA	385	23	BOTH	1,683	2,905
AGE64	Persons aged 60 to 64 years	SADA	371	37	BOTH	1,491	3,097
AGE65PL	Persons aged 65 years and older	SADA	390	18	BOTH	4,361	227
AGE74	Persons aged 65 to 74 years	SADA	400	8	BOTH	2,128	2,460
AGE75PL	Persons aged 75 years and older	SADA	391	17	BOTH	2,109	2,479
APTLT5	Households living in an apartment in a building with less than five storeys	SADA	406	2	BOTH	2,669	1,919
CHILD	Children in a census family	SADA	7	400	SADA	39	369
CHILDFAM	Census families with children	SADA	16	391	BOTH	3,145	1,443
COMLAW_ YE	Persons in a common law relationship	SADA	252	155	BOTH	2,708	1,880
COMLAWNO_ DIV	Divorced persons not in a common law relationship	SADA	405	2	BOTH	1,181	3,407
COMLAWNO_ OTHERS	Divorced, separated or widowed persons not in a common law relationship	SADA	399	8	BOTH	2,954	1,634
COMLAWNO_ SEP	Separated persons not in a common law relationship	SADA	399	8	BOTH	609	3,979

Table C.1
Statistics on the use of calibration constraints, by constraint

Variable/ constraint	Description	Coverage and non-response adjustment			Final calibration		
		Area	Number of calibrated constraints	Number of excluded constraints	Area	Number of calibrated constraints	Number of excluded constraints
COMLAWNO_ SINGLE	Single persons (never married) not in a common law relationship	SADA	204	203	BOTH	3,496	1,092
COMLAWNO_ SINGLE_LT15	Single persons (never married) under 15 years and not in a common law relationship	...	...	...	BOTH	4,473	115
COMLAWNO_ WID	Widowed persons not in a common law relationship	SADA	398	9	BOTH	763	3,825
COMLAWYE_ MARRIED	Persons married or in a common law relationship	SADA	200	207	BOTH	3,241	1,347
COUPLE	Persons in a couple (spouse, partner)	SADA	111	296	SADA	139	269
EMPIN_GT50	Persons with an annual employment income above the 50th percentile for the ADA	...	...	...	ADA	4,110	70
EMPIN_LE50	Persons with an annual employment income equal to or below the 50th percentile for the ADA	...	...	...	ADA	4,094	86
EMPIN_P0	Persons with no annual employment income, at the ADA level	...	...	...	ADA	4,129	51
EMPIN_P0_ GE15	Persons aged 15 years and older with no annual employment income, at the ADA level	...	...	...	ADA	4,069	111

Table C.1
Statistics on the use of calibration constraints, by constraint

Variable/ constraint	Description	Coverage and non-response adjustment			Final calibration		
		Area	Number of calibrated constraints	Number of excluded constraints	Area	Number of calibrated constraints	Number of excluded constraints
EMPIN_P0_ LT15	Persons aged 14 years and younger with no annual employment income, at the ADA level	...	...	...	ADA	4,083	97
EMPIN_P25	Persons with an annual employment income equal to or below the 25th percentile for the ADA	...	...	...	ADA	2,636	1,544
EMPIN_P50	Persons with an annual employment income above the 25th percentile and equal to or below the 50th percentile for the ADA	...	...	...	ADA	2,619	1,561
EMPIN_P75	Persons with an annual employment income above the 50th percentile and equal to or below the 75th percentile for the ADA	...	...	...	ADA	3,713	467
EMPIN_P100	Persons with an annual employment income above the 75th percentile for the ADA	...	...	...	ADA	3,718	462
EMPIN_SADA_ GT50	Persons with an annual employment income above the 50th percentile for the SADA	SADA	405	3	SADA	407	1

Table C.1
Statistics on the use of calibration constraints, by constraint

Variable/ constraint	Description	Coverage and non-response adjustment			Final calibration		
		Area	Number of calibrated constraints	Number of excluded constraints	Area	Number of calibrated constraints	Number of excluded constraints
EMPIN_SADA_LE50	Persons with an annual employment income equal to or below the 50th percentile for the SADA	SADA	406	2	SADA	407	1
EMPIN_SADA_P0	Persons with no annual employment income, at the SADA level	SADA	407	1	SADA	406	2
EMPIN_SADA_P0_GE15	Persons aged 15 years and older with no annual employment income, at the SADA level	...	...	...	SADA	392	16
EMPIN_SADA_P0_LT15	Persons aged 14 years and younger with no annual employment income, at the SADA level	...	...	...	SADA	390	18
EMPIN_SADA_P25	Persons with an annual employment income equal to or below the 25th percentile for the SADA	SADA	341	67	SADA	192	216
EMPIN_SADA_P50	Persons with an annual employment income above the 25th percentile and equal to or below the 50th percentile for the SADA	SADA	339	69	SADA	192	216

Table C.1
Statistics on the use of calibration constraints, by constraint

Variable/ constraint	Description	Coverage and non-response adjustment			Final calibration		
		Area	Number of calibrated constraints	Number of excluded constraints	Area	Number of calibrated constraints	Number of excluded constraints
EMPIN_SADA_P75	Persons with an annual employment income above the 50th percentile and equal to or below the 75th percentile for the SADA	SADA	386	22	SADA	407	1
EMPIN_SADA_P100	Persons with an annual employment income above the 75th percentile for the SADA	SADA	387	21	SADA	408	0
FEMALE	Females	SADA	39	369	BOTH	4,400	188
FEMALEGE15	Females aged 15 years and older	SADA	386	22	BOTH	3,054	1,534
FEMALELT15	Females aged 14 years and younger	SADA	35	373	BOTH	2,942	1,646
HHADA (ADA)	Households in the ADA	ADA	4,043	195	...	...	...
HHADACSD	Households that fall within the CSD and the ADA	...	...	...	ADA	4,938	1,994
HHINC_GT50	Households with an annual income above the 50th percentile for the ADA	...	...	...	ADA	4,138	42
HHINC_LE50	Households with an annual income at or below the 50th percentile for the ADA	...	...	...	ADA	4,138	42
HHINC_P25	Households with an annual income at or below the 25th percentile for the ADA	...	...	...	ADA	4,022	158

Table C.1
Statistics on the use of calibration constraints, by constraint

Variable/ constraint	Description	Coverage and non-response adjustment			Final calibration		
		Area	Number of calibrated constraints	Number of excluded constraints	Area	Number of calibrated constraints	Number of excluded constraints
HHINC_P50	Households with an annual income above the 25th percentile and at or below the 50th percentile for the ADA	...	...	...	ADA	4,024	156
HHINC_P75	Households with an annual income above the 50th percentile and at or below the 75th percentile for the ADA	...	...	...	ADA	4,014	166
HHINC_P100	Households with an annual income above the 75th percentile for the ADA	...	...	...	ADA	4,014	166
HHINC_SADA_GT50	Households with an annual income above the 50th percentile for the SADA	SADA	407	0	SADA	407	1
HHINC_SADA_LE50	Households with an annual income at or below the 50th percentile for the SADA	SADA	407	0	SADA	407	1
HHINC_SADA_P25	Households with an annual income at or below the 25th percentile for the SADA	SADA	407	0	SADA	407	1
HHINC_SADA_P50	Households with an annual income above the 25th percentile and at or below the 50th percentile for the SADA	SADA	407	0	SADA	407	1

Table C.1
Statistics on the use of calibration constraints, by constraint

Variable/ constraint	Description	Coverage and non-response adjustment			Final calibration		
		Area	Number of calibrated constraints	Number of excluded constraints	Area	Number of calibrated constraints	Number of excluded constraints
HHINC_SADA_ P75	Households with an annual income above the 50th percentile and at or below the 75th percentile for the SADA	SADA	402	5	SADA	407	1
HHINC_SADA_ P100	Households with an annual income above the 75th percentile for the SADA	SADA	402	5	SADA	407	1
HHSIZE1	One-person households	SADA	28	379	BOTH	766	3,822
HHSIZE2	Two-person households	SADA	114	293	BOTH	3,985	603
HHSIZE3	Three-person households	SADA	394	13	BOTH	3,634	954
HHSIZE4	Four-person households	SADA	376	31	BOTH	3,639	949
HHSIZE5	Five-person households	SADA	131	276	BOTH	991	3,597
HHSIZEGE5	Five-or-more-person households	SADA	15	392	BOTH	386	4,202
HHSIZEGE6	Six-or-more-person households	SADA	13	394	BOTH	277	4,311
INEFAM	Persons in an economic family	SADA	51	356	SADA	111	297
IR_LINK_NO	Persons who could not be linked to the Indian Register	SADA	207	200	BOTH	515	4,073
IR_LINK_YE	Persons who could be linked to the Indian Register	SADA	207	200	BOTH	451	4,137
LIM_NO	Persons not in a low income economic family	SADA	399	8	BOTH	3,267	1,321
LIM_YE	Persons in a low income economic family	SADA	399	8	BOTH	3,267	1,321
LONEPAR	Lone parents	SADA	118	289	SADA	149	259

Table C.1
Statistics on the use of calibration constraints, by constraint

Variable/ constraint	Description	Coverage and non-response adjustment			Final calibration		
		Area	Number of calibrated constraints	Number of excluded constraints	Area	Number of calibrated constraints	Number of excluded constraints
MALE	Males	SADA	39	369	BOTH	4,400	188
MALEGE15	Males aged 15 years and older	SADA	295	113	BOTH	3,054	1,534
MALELT15	Males aged 14 years and younger	SADA	53	355	BOTH	2,938	1,650
MARRIED	Married persons	SADA	346	62	BOTH	2,761	1,827
NOCLDFAM	Census families without children	SADA	40	367	BOTH	1,774	2,814
NOINEFAM	Persons in a household that are not part of an economic family	SADA	51	356	SADA	111	297
NOINEFAMH-HSIZEEQ1	Person in a one-person household that is not part of an economic family	...	...	...	SADA	103	305
NOINEFAMH-HSIZEGT1	Persons in a two-or-more-person household that are not part of an economic family	...	...	...	SADA	160	248
NOTINFAM	Persons not in a census family	SADA	75	332	SADA	143	265
NOTINFAMH-HSIZEGT1	Persons in a two-or-more-person household that are not part of a census family	...	...	...	SADA	98	310
OLN_BI	Official languages - English and French	SADA	374	33	BOTH	1,874	2,714
OLN_EN	Official language - English	SADA	78	329	BOTH	559	4,029
OLN_FR	Official language - French	SADA	74	333	BOTH	782	3,806
OLN_NO	Official language - neither	SADA	224	183	BOTH	455	4,133
POBG2_1	Place of birth - Albania	SADA	10	299	BOTH	64	4,524

Table C.1
Statistics on the use of calibration constraints, by constraint

Variable/ constraint	Description	Coverage and non-response adjustment			Final calibration		
		Area	Number of calibrated constraints	Number of excluded constraints	Area	Number of calibrated constraints	Number of excluded constraints
POBG2_2	Place of birth - Andorra, Gibraltar and Spain	SADA	22	366	BOTH	101	4,487
POBG2_3	Place of birth - Australia and New Zealand	SADA	13	383	BOTH	191	4,397
POBG2_4	Place of birth - Austria, Germany and Liechtenstein	SADA	25	382	BOTH	262	4,326
POBG2_5	Place of birth - Estonia, Latvia and Lithuania	SADA	17	332	BOTH	62	4,526
POBG2_6	Place of birth - Belgium and Netherlands	SADA	14	393	BOTH	241	4,347
POBG2_7	Place of birth - Belize, El Salvador, Guatemala and Honduras	SADA	26	381	BOTH	173	4,415
POBG2_8	Place of birth - Bhutan	SADA	3	122	BOTH	10	4,578
POBG2_9	Place of birth - Botswana, Lesotho and Swaziland	SADA	13	254	BOTH	11	4,577
POBG2_10	Place of birth - Brazil	SADA	15	389	BOTH	138	4,450
POBG2_11	Place of birth - Bulgaria and Romania	SADA	45	361	BOTH	179	4,409
POBG2_12	Place of birth - Burma (Myanmar)	SADA	6	285	BOTH	39	4,549
POBG2_13	Place of birth - Canada	SADA	34	335	BOTH	40	4,548
POBG2_14	Place of birth - Anguilla, Bermuda, British Virgin Islands, Cayman Islands, Montserrat and Turks and Caicos Islands	SADA	9	303	BOTH	26	4,562

Table C.1
Statistics on the use of calibration constraints, by constraint

Variable/ constraint	Description	Coverage and non-response adjustment			Final calibration		
		Area	Number of calibrated constraints	Number of excluded constraints	Area	Number of calibrated constraints	Number of excluded constraints
POBG2_15	Place of birth - Guadeloupe, Martinique and Saint Martin (French part)	SADA	4	181	BOTH	17	4,571
POBG2_16	Place of birth - Democratic Republic of the Congo and Republic of the Congo	SADA	16	366	BOTH	85	4,503
POBG2_17	Place of birth - Cameroon, Central African Republic, Chad and Gabon	SADA	16	350	BOTH	71	4,517
POBG2_18	Place of birth - Angola and Sao Tome and Principe	SADA	14	289	BOTH	44	4,544
POBG2_19	Place of birth - Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan and Uzbekistan	SADA	17	356	BOTH	89	4,499
POBG2_20	Place of birth - Chile	SADA	19	384	BOTH	162	4,426
POBG2_21	Place of birth - China, Hong Kong, Macao and Taiwan	SADA	65	342	BOTH	269	4,319
POBG2_22	Place of birth - Colombia, Ecuador and Peru	SADA	69	338	BOTH	256	4,332
POBG2_23	Place of birth - Costa Rica, Nicaragua and Panama	SADA	16	379	BOTH	127	4,461
POBG2_24	Place of birth - Czech Republic, Hungary and Slovakia	SADA	20	380	BOTH	161	4,427
POBG2_25	Place of birth - Burundi and Rwanda	SADA	14	340	BOTH	68	4,520

Table C.1
Statistics on the use of calibration constraints, by constraint

Variable/ constraint	Description	Coverage and non-response adjustment			Final calibration		
		Area	Number of calibrated constraints	Number of excluded constraints	Area	Number of calibrated constraints	Number of excluded constraints
POBG2_26	Place of birth - Eritrea, Kenya, Tanzania, Uganda and Zambia	SADA	33	358	BOTH	159	4,429
POBG2_27	Place of birth - Comoros, Djibouti, Madagascar, Malawi, Mauritius, Seychelles, Somalia and Zimbabwe	SADA	28	378	BOTH	150	4,438
POBG2_28	Place of birth - Belarus, Moldova, Russian Federation and Ukraine	SADA	43	364	BOTH	176	4,412
POBG2_29	Place of birth - Egypt, South Sudan and Sudan	SADA	22	377	BOTH	131	4,457
POBG2_30	Place of birth - Ethiopia	SADA	28	325	BOTH	111	4,477
POBG2_31	Place of birth - France, Luxembourg and Monaco	SADA	27	380	BOTH	197	4,391
POBG2_32	Place of birth - Cambodia, Laos and Viet Nam	SADA	59	348	BOTH	149	4,439
POBG2_33	Place of birth - Cuba, Dominican Republic and Haiti	SADA	34	372	BOTH	185	4,403
POBG2_34	Place of birth - Greece	SADA	16	374	BOTH	131	4,457
POBG2_35	Place of birth - Guyana and Suriname	SADA	45	327	BOTH	140	4,448
POBG2_36	Place of birth - Holy See (Vatican City State), Italy and San Marino	SADA	54	349	BOTH	241	4,347

Table C.1
Statistics on the use of calibration constraints, by constraint

Variable/ constraint	Description	Coverage and non-response adjustment			Final calibration		
		Area	Number of calibrated constraints	Number of excluded constraints	Area	Number of calibrated constraints	Number of excluded constraints
POBG2_37	Place of birth - Bahamas, Jamaica and Puerto Rico	SADA	76	314	BOTH	221	4,367
POBG2_38	Place of birth - Antigua and Barbuda, Aruba, Barbados, Bonaire, Sint Eustatius and Saba, Curaçao, Dominica, Grenada, Saint Kitts and Nevis, Saint Lucia, Sint Maarten (Dutch part), Saint Vincent and the Grenadines, Trinidad and Tobago and United States Virgin Islands	SADA	39	357	BOTH	186	4,402
POBG2_39	Place of birth - Japan	SADA	27	369	BOTH	175	4,413
POBG2_40	Place of birth - North Korea and South Korea	SADA	24	376	BOTH	120	4,468
POBG2_41	Place of birth - Liberia	SADA	12	274	BOTH	27	4,561
POBG2_42	Place of birth - Algeria, Libya, Morocco and Tunisia	SADA	23	382	BOTH	134	4,454
POBG2_43	Place of birth - Brunei Darussalam, Indonesia, Malaysia, Philippines, Singapore and Timor-Leste	SADA	87	320	BOTH	266	4,322
POBG2_44	Place of birth - Malta	SADA	10	270	BOTH	36	4,552
POBG2_45	Place of birth - Mexico	SADA	17	390	BOTH	183	4,405

Table C.1
Statistics on the use of calibration constraints, by constraint

Variable/ constraint	Description	Coverage and non-response adjustment			Final calibration		
		Area	Number of calibrated constraints	Number of excluded constraints	Area	Number of calibrated constraints	Number of excluded constraints
POBG2_46	Place of birth - Bahrain, Qatar, Saudi Arabia, United Arab Emirates and Yemen	SADA	24	356	BOTH	156	4,432
POBG2_47	Place of birth - Lebanon, Syria	SADA	33	373	BOTH	164	4,424
POBG2_48	Place of birth - Afghanistan, Cyprus, Iran, Iraq, Israel, Jordan, Kuwait, Oman, Turkey and West Bank and Gaza Strip (Palestine)	SADA	94	309	BOTH	218	4,370
POBG2_49	Place of birth - Mongolia	SADA	3	161	BOTH	12	4,576
POBG2_50	Place of birth - Mozambique	SADA	18	243	BOTH	20	4,568
POBG2_51	Place of birth - Nepal	SADA	11	315	BOTH	33	4,555
POBG2_54	Place of birth - Poland	SADA	64	339	BOTH	222	4,366
POBG2_55	Place of birth - Oceania Region (excluding Australia and New Zealand)	SADA	13	348	BOTH	57	4,531
POBG2_56	Place of birth - Portugal	SADA	52	343	BOTH	222	4,366
POBG2_57	Place of birth - Argentina, Bolivia, Paraguay and Uruguay	SADA	19	383	BOTH	154	4,434
POBG2_58	Place of birth - Åland Islands, Denmark, Faroe Islands, Finland, Iceland, Norway and Sweden	SADA	14	383	BOTH	158	4,430
POBG2_59	Place of birth - Namibia and Republic of South Africa	SADA	13	373	BOTH	146	4,442

Table C.1
Statistics on the use of calibration constraints, by constraint

Variable/ constraint	Description	Coverage and non-response adjustment			Final calibration		
		Area	Number of calibrated constraints	Number of excluded constraints	Area	Number of calibrated constraints	Number of excluded constraints
POBG2_60	Place of birth - Sri Lanka	SADA	27	342	BOTH	93	4,495
POBG2_61	Place of birth - Switzerland	SADA	18	387	BOTH	165	4,423
POBG2_62	Place of birth - Thailand	SADA	20	379	BOTH	138	4,450
POBG2_63	Place of birth - Armenia, Azerbaijan and Georgia	SADA	9	324	BOTH	48	4,540
POBG2_64	Place of birth - Bangladesh, India and Pakistan	SADA	38	364	BOTH	161	4,427
POBG2_65	Place of birth - Union of Soviet Socialist Republics, Former	SADA	20	365	BOTH	138	4,450
POBG2_66	Place of birth - Guernsey, Ireland, Isle of Man, Jersey, Sark and United Kingdom	SADA	10	397	BOTH	127	4,461
POBG2_67	Place of birth - United States	SADA	220	187	BOTH	443	4,145
POBG2_68	Place of birth - Venezuela	SADA	16	364	BOTH	111	4,477
POBG2_69	Place of birth - Gambia, Ghana, Nigeria and Sierra Leone	SADA	13	371	BOTH	106	4,482
POBG2_70	Place of birth - Benin, Burkina Faso, Côte d'Ivoire, Guinea, Mali, Mauritania, Niger, Senegal and Togo	SADA	9	372	BOTH	73	4,515

Table C.1
Statistics on the use of calibration constraints, by constraint

Variable/ constraint	Description	Coverage and non-response adjustment			Final calibration		
		Area	Number of calibrated constraints	Number of excluded constraints	Area	Number of calibrated constraints	Number of excluded constraints
POBG2_71	Place of birth - Bosnia and Herzegovina, Croatia, Kosovo, Republic of Macedonia, Montenegro, Serbia and Slovenia	SADA	55	346	BOTH	221	4,367
POBG3_1	Place of birth - Canada	SADA	34	335	BOTH	40	4,548
POBG3_2	Place of birth - Caribbean and Bermuda	SADA	154	253	BOTH	374	4,214
POBG3_3	Place of birth - Central Africa	SADA	27	375	BOTH	142	4,446
POBG3_4	Place of birth - Central America	SADA	128	279	BOTH	305	4,283
POBG3_5	Place of birth - Eastern Africa	SADA	120	287	BOTH	300	4,288
POBG3_6	Place of birth - Eastern Asia	SADA	205	202	BOTH	739	3,849
POBG3_7	Place of birth - Eastern Europe (excluding Union of Soviet Socialist Republics, Former)	SADA	180	227	BOTH	381	4,207
POBG3_8	Place of birth - West Central Asia and the Middle East	SADA	191	215	BOTH	470	4,118
POBG3_9	Place of birth - Northern Africa	SADA	89	318	BOTH	262	4,326
POBG3_10	Place of birth - Northern Europe	SADA	247	160	BOTH	461	4,127
POBG3_12	Place of birth - Oceania	SADA	47	356	BOTH	215	4,373
POBG3_14	Place of birth - South America	SADA	176	231	BOTH	364	4,224
POBG3_15	Place of birth - Southeast Asia	SADA	249	158	BOTH	652	3,936
POBG3_16	Place of birth - Southern Africa	SADA	17	370	BOTH	158	4,430

Table C.1
Statistics on the use of calibration constraints, by constraint

		Coverage and non-response adjustment			Final calibration		
Variable/ constraint	Description	Area	Number of calibrated constraints	Number of excluded constraints	Area	Number of calibrated constraints	Number of excluded constraints
POBG3_17	Place of birth - Southern Asia	SADA	218	185	BOTH	647	3,941
POBG3_18	Place of birth - Southern Europe	SADA	182	225	BOTH	407	4,181
POBG3_19	Place of birth - Union of Soviet Socialist Republics, Former	SADA	112	295	BOTH	307	4,281
POBG3_20	Place of birth - United States	SADA	220	187	BOTH	443	4,145
POBG3_21	Place of birth - Western Africa	SADA	68	339	BOTH	211	4,377
POBG3_22	Place of birth - Western Europe	SADA	195	212	BOTH	449	4,139
PPADA (ADA)	Persons in the ADA	ADA	42	4,196	...	...	...
PPADACSD	Persons with geographic overlap between census subdivision and ADA	...	...	...	ADA	4,881	2,051
SNGLDET	Households living in a single-detached house	SADA	392	16	BOTH	3,704	884
TOTCFAM	Census families	SADA	333	74	BOTH	1,672	2,916
TOTHHLD	Households	SADA	408	0	BOTH	4,588	0
TOTPERS	Persons	SADA	408	0	BOTH	4,588	0
TPERGE15	Persons aged 15 years and older	SADA	288	120	BOTH	4,474	114
YRIMD_1900	Immigrants who landed prior to 1981	SADA	360	47	SADA	376	32
YRIMD_1981	Immigrants who landed from 1981 to 1985	SADA	229	178	SADA	277	131
YRIMD_1986	Immigrants who landed from 1986 to 1990	SADA	224	183	SADA	276	132
YRIMD_1991	Immigrants who landed from 1991 to 1995	SADA	238	169	SADA	293	115

Table C.1
Statistics on the use of calibration constraints, by constraint

Variable/ constraint	Description	Coverage and non-response adjustment			Final calibration		
		Area	Number of calibrated constraints	Number of excluded constraints	Area	Number of calibrated constraints	Number of excluded constraints
YRIMD_1996	Immigrants who landed from 1996 to 2000	SADA	231	176	SADA	290	118
YRIMD_2001	Immigrants who landed from 2001 to 2005	SADA	259	148	SADA	309	99
YRIMD_2006	Immigrants who landed from 2006 to 2010	SADA	258	149	SADA	310	98
YRIMD_2011	Immigrants who landed from 2011 to 2016	SADA	300	107	SADA	333	75
YRIMD_M5	Persons with no year of immigration	...	...	...	SADA	239	169
YRIMG1_1900	Immigrants who landed prior to 1981	SADA	360	47	SADA	376	32
YRIMG1_1981	Immigrants who landed from 1981 to 1990	SADA	310	97	SADA	338	70
YRIMG1_1991	Immigrants who landed from 1991 to 2000	SADA	320	87	SADA	350	58
YRIMG1_2001	Immigrants who landed from 2001 to 2010	SADA	346	61	SADA	369	39
YRIMG1_2011	Immigrants who landed from 2011 to 2016	SADA	300	107	SADA	333	75

... not applicable
SADA = Super aggregate dissemination area.
ADA = Aggregate dissemination area.
BOTH = SADA/ADA.
Source: Statistics Canada, 2016 Census long-form sample.

Appendix D – Wood Buffalo

Introduction

On May 1, 2016, a wildfire began southwest of Fort McMurray, Alberta. On May 3, the wildfire swept through the community, destroying many homes and buildings and forcing the largest wildfire evacuation in Alberta's history. Statistics Canada suspended census data collection (referred to as "field data collection") in the evacuated areas. At that time, many responses had already been received from the evacuated areas.

Statistics Canada used a set of measures to ensure that the 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 areas were derived from a combination of sources: direct field data collection and administrative data sources. Statistics Canada worked closely with provincial and local authorities in Alberta to obtain access to administrative records to help validate the data derived from administrative data sources available at Statistics Canada.

This appendix provides additional details about the methodology used in the Wood Buffalo CSD as the result of the wildfire. In particular, it covers weighting and estimation procedures. More details on the collection procedures can be found in Appendix 1.4 (http://www12.statcan.gc.ca/census-recensement/2016/ref/98-304/app-ann1-4-eng.cfm) of the *Guide to the Census of Population, 2016*, Catalogue no. 98-304-X.

Reference date

For the 2016 Census, the reference date for data reporting was May 10, 2016. For residents of the areas evacuated during the wildfire, the reference date was May 1, 2016. This new date was deemed to better represent the situation in the Wood Buffalo CSD. Consequently, a slight discrepancy may have arisen in responses received after the fire began but before field data collection resumed in August. This is considered to be a minor issue.

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 related to a person's name, date of birth, sex and marital status.

Since the administrative data files did not contain the same language content that was 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 proxies for the 2016 language questions. Responses to 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.

Data collection

In August 2016, data collection was reinstated in Wood Buffalo, and census representatives went door to door to complete census questionnaires. Efforts were focused on collecting data for the one in four dwellings included in the long-form questionnaire sample. To further improve data quality, field data collection was performed for dwellings in the areas for which no administrative data were available and for collective dwellings. In areas where enumerators prepared a list of dwellings and delivered census materials (list/leave areas), field data collection was done for all dwellings.

Processing

If a census response was obtained for the residents of a dwelling, the data collected from that respondent took precedence over any available administrative data. For cases without a census response, data from administrative sources were considered a response to the same extent as a direct response obtained through traditional collection methods during data processing and for the calculation of response rates. The same procedures as described in Chapter 3 were applied to both data direct from respondents and data from administrative sources.

Data quality

For 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 provided satisfactory results.

Estimation

So that a calibration adjustment did not cause an estimate bias, the sample and census data needed to be conceptually comparable. In Wood Buffalo, the collection response rate for the short-form questionnaire (2A) was lower than for the long-form questionnaire (2A-L). This occurred because the objective was, when possible, to complete 2A questionnaires using administrative data sources and to complete 2A-L questionnaires with collection data.

For the 2A form, 27% of the data came from direct responses, 67% came from administrative data sources and 6% were imputed through the Dwelling Classification Survey (see Chapter 2). For the 2A-L form, 79% of the data came from direct responses, 8% came from 2A questionnaire administrative sources, 6% came from a mix of the 2A form administrative data sources and the imputed 2A-L forms, 6% came from 2A form imputations through the DCS, and 1% came from 2A form responses only.

Some census concepts are more difficult to derive from administrative data sources, namely families, marital status, common law status and household income characteristics. Therefore, the difference in contribution from administrative data sources to the 2A form and to the 2A-L form resulted in sizeable differences between census counts and sample estimates (before weight adjustments), as expected.

When weighting diagnostics were performed, the weighting distribution was observed to be atypical or bimodal for Wood Buffalo. This issue was found to be the result of more households being of size 1 in the administrative database (a database built by Statistics Canada from administrative sources used to populate the 2A form) than expected when compared with the census long-form data.

Sizeable discrepancies were also found for characteristics derived from the Indian Register and immigration administrative data linked to census data. In these cases, the differences stemmed in part from the fact that these data can be linked to the administrative database much more easily than to returned questionnaires. This is because the Indian Register and immigration administrative data were used to form the administrative database.

To resolve these issues, certain constraints were deactivated for Wood Buffalo to prevent the introduction of biases in the estimates. The deactivated constraints relate to characteristics involving families, marital status, common law status, household size and household income, and to characteristics derived from the Indian Register and immigration administrative data linked to census records.

The following table shows the differences between the census counts and long-form estimates for those characteristics with large discrepancies. Note that the census counts were largely gathered from administrative sources. This is particularly the case for variables related to the 2A-L form, such as year of immigration, where the so-called census counts were strictly administrative data counts.

Table D.1
Census counts and long-form estimate differences and discrepancies – Wood Buffalo

Variable/Contrainte	Description	Census counts (mostly from administrative sources)	Estimates (2A-L)	Relative difference (%)	Difference
ADULTCF	Adults in a census family	32,505	35,215	-8.34	-2,710
COMLAW_YE	Persons in a common law relationship	5,986	7,754	-29.53	-1,768
COMLAWNO_SEP	Separated persons not in a common law relationship	1,678	1,080	35.63	598
HHINC_P25	Households with an annual income at or below the 25th percentile for the ADA	6,341	5,466	13.79	875
LIM_YE	Persons in a low income economic family	3,134	2,462	21.44	672
LONEPAR	Lone parents	2,547	2,113	17.05	434
NOCLDFAM	Census families without children	6,167	7,034	-14.05	-867
NOINEFAM	Persons in a household that are not part of an economic family	13,714	10,202	25.61	3,512
NOTINFAM	Persons not in a census family	17,342	12,875	25.76	4,467
HHSIZE1	One-person households	5,879	4,646	20.98	1,233
HHSIZEGE6	Six-or-more-person households	1,450	1,096	24.40	354
IR_LINK_YE	Persons who could be linked to the Indian Register	2,261	1,853	18.05	408
YRIMG1_1981	Immigrants who landed from 1981 to 1990	634	548	13.55	86
YRIMG1_1991	Immigrants who landed from 1991 to 2000	1,934	1,612	16.65	322
YRIMG1_2001	Immigrants who landed from 2001 to 2010	5,692	5,092	10.54	600

ADA = Aggregate dissemination area.
Source: Statistics Canada, 2016 Census long-form sample.

This table lists the constraints both used for and excluded from the calibration process in Wood Buffalo. A value of “Yes” indicates that the variable was used in the process, while a value of “No” indicates that the variable was excluded.

Table D.2
Calibration constraints excluded in Wood Buffalo

Variable/Constraint	Excluded in Wood Buffalo
ADULTCF	No
AGE00_14	Yes
AGE14	Yes
AGE15_24	Yes
AGE19	Yes
AGE24	Yes
AGE25_34	Yes
AGE29	Yes
AGE34	Yes
AGE35_44	Yes
AGE39	Yes
AGE4	Yes
AGE44	Yes
AGE45_54	Yes
AGE49	Yes
AGE54	Yes
AGE55_64	Yes
AGE59	Yes
AGE64	Yes
AGE65PL	Yes
AGE74	Yes
AGE75PL	Yes
AGE9	Yes
APTLT5	Yes
CHILD	No
CHILDFAM	No
COMLAW_YE	No
COMLAWNO_DIV	No
COMLAWNO_OTHERS	No
COMLAWNO_SEP	No
COMLAWNO_SINGLE	No
COMLAWNO_WID	No
COMLAWYE_MARRIED	No
COUPLE	No
EMPIN_GT50	Yes
EMPIN_LE50	Yes
EMPIN_P0	Yes
EMPIN_P100	Yes
EMPIN_P25	Yes
EMPIN_P50	Yes

Table D.2
Calibration constraints excluded in Wood Buffalo

Variable/Constraint	Excluded in Wood Buffalo
EMPIN_P75	Yes
FEMALE	Yes
FEMALEGE15	Yes
FEMALELT15	Yes
HHADACSD (ADA)	Yes
HHINC_GT50	No
HHINC_LE50	No
HHINC_P100	No
HHINC_P25	No
HHINC_P50	No
HHINC_P75	No
HHSIZE1	No
HHSIZE2	No
HHSIZE3	No
HHSIZE4	No
HHSIZE5	No
HHSIZEGE5	No
HHSIZEGE6	No
INEFAM	No
IR_LINK_NO	No
IR_LINK_YE	No
LIM_NO	No
LIM_YE	No
LONEPAR	No
MALE	Yes
MALEGE15	Yes
MALELT15	Yes
MARRIED	Yes
NOCLDFAM	No
NOINEFAM	No
NOTINFAM	No
OLN_BI	Yes
OLN_EN	Yes
OLN_FR	Yes
OLN_NO	Yes
POBG2_1	No
POBG2_10	No
POBG2_11	No
POBG2_12	No
POBG2_13	No
POBG2_14	No
POBG2_15	No
POBG2_16	No
POBG2_17	No

Table D.2
Calibration constraints excluded in Wood Buffalo

Variable/Constraint	Excluded in Wood Buffalo
POBG2_18	No
POBG2_19	No
POBG2_2	No
POBG2_20	No
POBG2_21	No
POBG2_22	No
POBG2_23	No
POBG2_24	No
POBG2_25	No
POBG2_26	No
POBG2_27	No
POBG2_28	No
POBG2_29	No
POBG2_3	No
POBG2_30	No
POBG2_31	No
POBG2_32	No
POBG2_33	No
POBG2_34	No
POBG2_35	No
POBG2_36	No
POBG2_37	No
POBG2_38	No
POBG2_39	No
POBG2_4	No
POBG2_40	No
POBG2_41	No
POBG2_42	No
POBG2_43	No
POBG2_44	No
POBG2_45	No
POBG2_46	No
POBG2_47	No
POBG2_48	No
POBG2_49	No
POBG2_5	No
POBG2_50	No
POBG2_51	No
POBG2_54	No
POBG2_55	No
POBG2_56	No
POBG2_57	No
POBG2_58	No
POBG2_59	No

Table D.2
Calibration constraints excluded in Wood Buffalo

Variable/Constraint	Excluded in Wood Buffalo
POBG2_6	No
POBG2_60	No
POBG2_61	No
POBG2_62	No
POBG2_63	No
POBG2_64	No
POBG2_65	No
POBG2_66	No
POBG2_67	No
POBG2_68	No
POBG2_69	No
POBG2_7	No
POBG2_70	No
POBG2_71	No
POBG2_8	No
POBG2_9	No
POBG3_1	No
POBG3_10	No
POBG3_12	No
POBG3_14	No
POBG3_15	No
POBG3_16	No
POBG3_17	No
POBG3_18	No
POBG3_19	No
POBG3_2	No
POBG3_20	No
POBG3_21	No
POBG3_22	No
POBG3_3	No
POBG3_4	No
POBG3_5	No
POBG3_6	No
POBG3_7	No
POBG3_8	No
POBG3_9	No
PPADACSD (ADA)	Yes
SNGLDET	Yes
TOTCFAM	No
TOTHHLD	Yes
TOTPERS	Yes
TPERGE15	Yes
TPERLT15	Yes
YRIMD_1900	No

Table D.2
Calibration constraints excluded in Wood Buffalo

Variable/Constraint	Excluded in Wood Buffalo
YRIMD_1981	No
YRIMD_1986	No
YRIMD_1991	No
YRIMD_1996	No
YRIMD_2001	No
YRIMD_2006	No
YRIMD_2011	No
YRIMG1_1900	No
YRIMG1_1981	No
YRIMG1_1991	No
YRIMG1_2001	No
YRIMG1_2011	No

Notes:
Yes indicates the variable was used in the process.
No indicates it was excluded.

Bibliography

Devin, N., and F. Verret. 2016. “The development of a variance estimation methodology for large-scale dissemination of quality indicators for the 2016 Canadian census long form sample.” *JSM Proceedings*. Survey Research Methods Section, American Statistical Association. Chicago, United States.

Dominion Bureau of Statistics. 1968. *Sampling in the Census*. Internal report. Statistics Canada. Ottawa, Ontario.

Fellegi, I.P. 1964. “Response variance and its estimation.” *Journal of the American Statistical Association*. Vol. 59, no. 308. p. 1016–1041.

Folsom, R.E., Jr., and A.C. Singh. 2000. “The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification.” *Proceedings of the Survey Research Methods Section*. American Statistical Association. p. 598–603.

Hansen, M.H., W.N. Hurwitz and M.A. Bershad. 1959. “Measu-rement errors in censuses and surveys.” *Bulletin of the International Statistical Institute*. Vol. 38. p. 359–374.

Judkins, D.R. 1990. “Fay’s method for variance estimation.” *Journal of Official Statistics*. Statistics Sweden. Vol. 6, no. 3. p. 223–239.

Rao, J.N.K., and J. Shao. 1999. “Modified balanced repeated replication for complex survey data.” *Biometrika*. Oxford University Press. Vol. 86, no. 2. p. 403–415.

Royce, D. 1983. *The Use of Sampling in the 1981 Canadian Census.* Internal report. Statistics Canada. Ottawa, Ontario.

Shao, J., and Q. Tang. 2011. “Random group variance estimators for survey data with random hot deck imputation.” *Journal of Official Statistics*. Statistics Sweden. Vol. 27, no. 3. p. 507–526.

Statistics Canada. 2014. Canadian Census Edit and Imputation System (CANCEIS), version 5.2. Basic user guide. Ottawa, Ontario.

Statistics Canada. 2015. *Sampling and Weighting Technical Report, 2011 National Household Survey (NHS)*. Catalogue no. 99-002-X2011001. Ottawa, Ontario. Version revised March 24, 2015. http://www12.statcan.gc.ca/nhs-enm/2011/ref/reports-rapports/sw-ep/index-eng.cfm (accessed February 28, 2018).

Statistics Canada. 2017. *Guide to the Census of Population, 2016*. Catalogue no. 98-304-X2016001. Ottawa, Ontario. Version updated January 30, 2018. http://www12.statcan.gc.ca/census-recensement/2016/ref/98-304/index-eng.cfm (accessed February 28, 2018).