Catalogue no. 99-002-X2011001
ISBN : 978-1-100-25393-0

Technical Report

Sampling and Weighting Technical Report

National Household Survey, 2011

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 infostats@statcan.gc.ca

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

• Statistical Information Service	1-800-263-1136
• National telecommunications device for the hearing impaired	1-800-363-7629
• Fax line	1-877-287-4369

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 “About us” > “The agency” > “Providing services to Canadians.”

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 ($p < 0.05$)

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Table of contents

Page

Introduction

The 2011 National Household Survey (NHS) is a new voluntary survey distributed to about 30% of Canadian census households. It collects information previously collected by the long-form census questionnaire which was distributed to one in five households every five years. The 2011 Census questionnaire consists of the same eight questions that appeared on the 2006 Census short form questionnaire as well as two additional questions on language. Contrary to the NHS, participation to the census is mandatory.

The NHS collected social and economic information that communities need to plan services such as child care, schooling, family services, housing, roads and public transportation, and skills training for employment. The goal of the NHS is to provide data for small geographical areas and small population groups.

In order to help users assess the usefulness of NHS data for their purposes, this 2011 National Household Survey Technical Report on Sampling and Weighting details the conceptual framework and definitions used in conducting the NHS, as well as the data collection and processing procedures employed. Because those procedures are linked to the data collection and processing procedures employed by the census, the 2011 Census data collection and processing procedures are also described. The report then outlines the method of sampling and weighting used in the 2011 NHS as well as its effect on the results. Because some information is collected on a sample basis and weighted to the full population level, bias and discrepancies may be observed in the final estimates. This report identifies some of these observed differences and explains the probable causes.

Because of its potential impact on response rates, the new voluntary nature of the NHS has brought more complex sampling and weighting methods than those used for the 2006 Census.

In general, the motivation to use sampling stems from a desire to reduce respondent burden, reduce costs, obtain results sooner, or a combination of all three. A disadvantage of sampling is that the results based on a sample may not be as precise as those based on the whole population. Nonetheless, when the loss in precision (which may be quite small when the sample is large) is tolerable in terms of the intended uses of the data, sampling is often cost-effective.

The 2011 Census of Population and NHS used sampling in a variety of ways. It was used in:

- Ensuring that the quality of the enumerator's work in collecting questionnaires met certain standards.
- Controlling the quality of coding responses during processing.
- Evaluating the quality of the census and NHS data.

The primary use of sampling, however, was to select dwellings to be included in the NHS sample. This report describes this specific use of sampling and evaluates its effect on the quality of NHS data.

Chapter 1 details the collection methods used in the 2011 Census and NHS. Chapter 2 explains the data processing procedures. Chapter 3 describes how sampling is applied in the NHS. Chapter 4 gives an overview of the procedures used in applying weights to the NHS sample data in order to obtain estimates for the population. Chapter 5 provides information on NHS non-response bias indicators while Chapter 6 describes the weighting areas. Chapter 7 presents an evaluation of the weighting procedures and Chapter 8 provides an overview of the variance estimation methodology in the 2011 NHS.

1. Data collection

1.1 Census data collection

The data collection stage of the 2011 Census ensured that each of the 14.5 million dwellings in Canada received 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 together with 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, and family members living with them.

The census also counted Canadian citizens and landed immigrants who were temporarily outside the country on Census Day. This included federal and provincial/territorial government employees working outside Canada, Canadian embassy staff posted to other countries, members of the Canadian Forces stationed abroad and Canadians aboard merchant vessels.

In the Census of Canada, various questionnaires and forms are used to collect data. Form 1 is the Visitation Record (VR). It is used to list every occupied and unoccupied private dwelling or collective dwelling, agricultural operation and agricultural operator in the collection unit. It serves as an address listing for field operations and control purposes for data collection.

Form 2A is the basic short questionnaire used to enumerate all private dwellings. Every household that receives a 2A census questionnaire is asked to list all household members who belong to the census population and answer questions for them.

Form 2C is used to enumerate Canadians posted in other countries, including government employees (federal and provincial/territorial) and their families, and members of the Canadian Forces and their families.

Form 3A is an individual census questionnaire used to enumerate persons in collective dwellings[1] (each person in a collective dwelling must complete a separate Form 3). 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 2A questionnaire.

Wave methodology is a new approach to data collection first used in the 2011 Census. Households are contacted at appropriate times to remind them to participate in the census and persuade them to complete the questionnaire. In each wave, households are provided with the information they need to respond. Based on the fact that every Canadian household is required by law to answer the census questions, the method is designed to encourage people to respond promptly online, while mitigating the risk of a decline in overall response and the need for costly field follow-up.

This new methodology varies with the collection method used to distribute the census materials for a given region. These collection methods are described in the next section. In 2011, Canadian households had the option of responding online, on a paper questionnaire (mail-back) or by contacting the Census Help Line.

1.1.1 Collection methods: Delivery of census questionnaires

The three questionnaire delivery methods used in the Canadian census are mail-out, list/leave and interview. To make census collection as efficient as possible, Canada is divided into small geographic units known as collection units (CUs). In the 2011 Census, there were some 46,000 CUs in Canada.

1.1.1.1 Mail-out

For mail-out CUs, the postal system is used to deliver the census materials. This method ensures effective, coordinated distribution, without the need to recruit and train a large contingent of enumerators. Mail-out CUs are typically in urban areas. While mail-out CUs now include about 80% of Canadian dwellings, they cover only a tiny fraction of the country’s land area.

1. A collective dwelling is a dwelling used for commercial, institutional or communal purposes (for example hospitals, hotels and prisons).

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 18% of Canadian dwellings are in list/leave CUs, which cover a large portion of the country's land area.

1.1.1.3 Interview

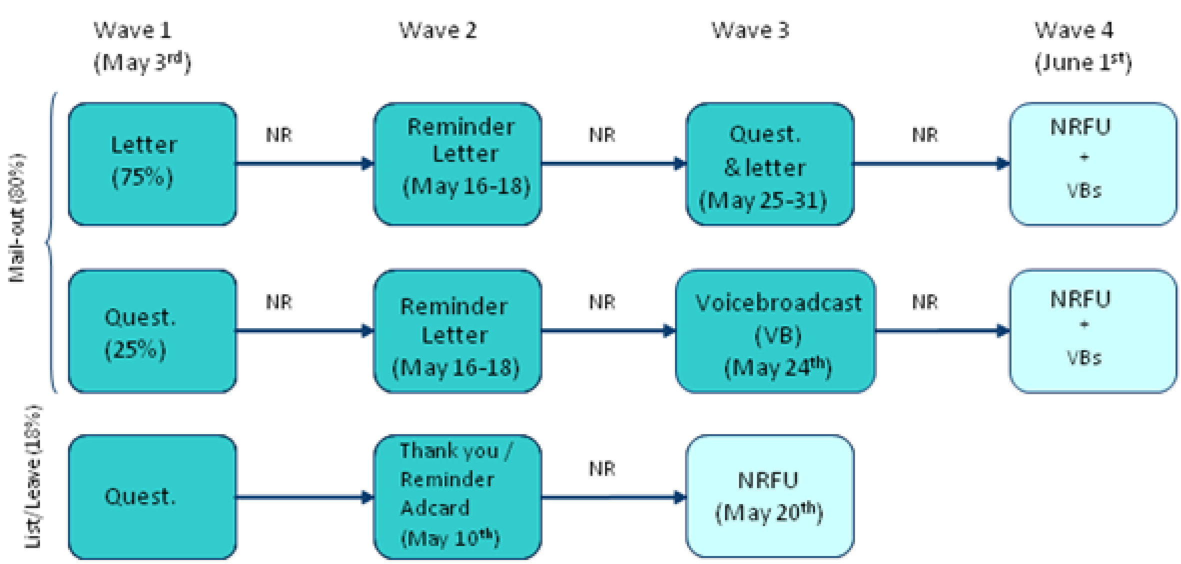
Interview CUs are usually in remote or difficult to access places and in Aboriginal communities. To limit the number of trips that enumerators have to make to those places for follow-ups—trips that are often expensive and logistically complicated—they do more than prepare dwelling lists and deliver census materials; they also complete a questionnaire with each household on the spot. Interview CUs cover just over half of Canada's land area, but only about 2% of its dwellings.

1.1.2 Wave methodology in the census

Wave methodology was designed to encourage online response while offering an alternative for households that do not wish to complete their questionnaire on the Web. This approach has many advantages for response rates, questionnaire registration, question flow and data capture.

Wave methodology is applied differently in different CU groups. Three main groups of CUs were defined, and a different wave methodology was developed for each one. Because of the nature of interview CUs, however, no wave methodology was developed for them. The sections that follow and Figure 1.1.2.1 provide an overview of the wave methodology used in the 2011 Census.

Figure 1.1.2.1 Overview of the wave methodology used in the 2011 Census



Note: NR = Non-respondents; NRFU = Non-response follow-up.

Source: Statistics Canada, 2011 Census.

1.1.2.1 Mail-out collection units – Wave 1 letter

First, a set of mail-out CU was identified so that the households most likely to respond online could be targeted and those least likely to respond to the census could be avoided. That set of CUs covers about 75% of the dwellings in mail-out areas. For that group of CUs, Wave 1 involved sending out only one letter asking households to complete the questionnaire online using the secure access code (SAC) provided or call an automated system on a toll-free line to have a paper questionnaire mailed to them. The Wave 1 letters were delivered by the postal system one week before Census Day (i.e., on May 3, 2011).

Wave 2 consisted of a reminder letter sent to all Wave 1 non-respondent households. The letter reminded the households that they were required by law to complete the census. Like the Wave 1 letter, it also provided the SAC and the toll-free telephone number. It was delivered to households between May 16 and 18, i.e., as early as six days after Census Day.

In Wave 3, a paper questionnaire was sent to non-respondent households. It was delivered to them between May 25 and May 31, i.e., as early as 15 days after Census Day. The households could still respond online using a 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, 2011, an enumerator would contact the household by telephone or in person to complete the questionnaire. It was also noted in the letter 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, 2011, consisted of field non-response follow-up (NRFU) and an automated reminder call. NRFU is described in Section 1.1.3 of this document.

1.1.2.2 Mail-out collection units – Wave 1 questionnaire

The second group of CU on which a variant of wave methodology was used is the set of other mail-out CUs. That set of CUs covers about 25% of the dwellings in mail-out areas. Households in those CUs were considered less likely to respond after receiving only a letter. Wave 1 for those CUs was the mailing of a paper questionnaire. The questionnaire provided a SAC, so that the household had the option of responding online. As in the case of the first group, Wave 1 took place one week before Census Day (i.e., on May 3, 2011). Wave 2 for the second group was the same as for the first group. In Wave 3, the group's non-respondent households for which a telephone number was available in the census frame received an automated reminder call on May 24. Wave 4 was the same as the first group.

1.1.2.3 List/leave collection units

The third group of CU on which another variant of wave methodology was used is the set of all list/leave CUs. In Wave 1, enumerators delivered a paper questionnaire to all dwellings in those CUs on or about May 3. The questionnaire also provided a SAC, so that the household had the option of responding online. In Wave 2, all dwellings in those CUs received a thank you / reminder card in the mail on May 10, whether they had responded or not, because it was generally impossible in those areas to send mail to civic addresses. Wave 3, 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 final wave in the wave methodology was non-response follow-up (NRFU). In that wave, enumerators telephoned and visited households that had not responded. Each non-respondent household for which a telephone number was available received an automated reminder call at the beginning of the NRFU period. The message reminded non-respondents of their legal obligation to respond to the census.

The enumerators had information from the Field Management System (FMS) to help them manage their work. A computerized system accessible over the Internet, the FMS was developed for the 2011 Census to facilitate the gathering of collection progress information.

1.1.4 Verification of dwellings' occupancy status

Before NRFU, field operations were also carried out for the dwelling occupancy verification (DOV). The purpose of DOV, which began on May 13, 2011, was to identify a significant number of dwellings that were unoccupied on Census Day or cancelled (addresses that are not private or collective dwellings) before NRFU started. Identifying such dwellings close to Census Day should make occupancy classification more accurate and perhaps easier to perform. DOV also reduces the NRFU workload, since any 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 operation, the Dwelling Classification Survey, is carried out after NRFU. It assesses and determines the occupancy status of dwellings for which no completed questionnaire has been received (unoccupied dwellings, non-respondent dwellings or unresolved cases). The survey's results are used to adjust the Census of Population counts during processing (see Section 2.7 or the *Coverage Technical Report*, 2011 Census, Catalogue no 98-303-X).

1.2 Collection methodology of the 2011 National Household Survey

The target population of the National Household Survey (NHS) consists of persons enumerated by the census who live in private dwellings.

Two questionnaires are used to collect the data: Forms N1 and N2. The two forms contain the same questions, but Form N2 is used in interview areas. The same 10 census questions are also on the NHS forms.

The decision was made to conduct the NHS during the same period as the census in order to take advantage of census resources and infrastructure such as collection management systems and employees. However, that strategy had the potential to impose a heavier burden on employees and respondents. NHS collection operations therefore had to be managed in such a way as to minimize any impact on the results of the census, since the latter takes precedence. Persons who responded to the census on the Internet (mail-out CUs) were given the opportunity to complete the NHS online immediately after finishing the census questionnaire, and those who used a paper census questionnaire were contacted about completing the NHS on or after June 7, when the paper NHS questionnaires were delivered.

The NHS collection strategy also had to take account of resource limitations, time constraints and the survey's voluntary nature, which prevented interviewers from pushing too hard and reduced the number of possible contacts with non-respondents.

1.2.1 NHS collection in interview areas

In interview areas, NHS data collection with Form N2 was carried out at the same time as census enumeration. The responses to the census questions, which were also in Form N2, were copied from one form to the other, as long as the respondent agreed to participate in the NHS.

In non-interview areas, the wave collection methodology described in the next section was used.

1.2.2 Wave methodology in the NHS

A wave methodology was also used for NHS collection, in conjunction with the wave methodology used in the census. The NHS wave methodology was based not only on the CU's collection method but also on the census response status and response mode (online or paper). With the dwellings selected for the NHS, the three groups described below were established.

The first group of dwellings, referred to as Survey group 1, consisted of online census respondents. The dwellings in Survey group 2 were census respondents who used the paper questionnaire. Survey group 3 consisted of census non-respondents.

The NHS wave methodology is different for each survey group; details are presented below. Table 1.2.2.1 also provides an overview. As in the case of the census, because of the nature of the interview CUs, wave methodology was not used for them.

Table 1.2.2.1 Overview of the wave methodology in the 2011 NHS

Survey group 1 – Online census respondents (mail-out CUs)	
Wave 1	Online offer immediately after completing the census questionnaire
Wave 2	Reminder letter received on or about June 7
Wave 3	None
Wave 4	Telephone or field non-response follow-up (starting only when census response was acceptable, on June 8 in some areas where census collection was almost complete)
Survey group 2 – Paper questionnaire census respondents (mail-out CUs)	
Wave 1	Questionnaire received on or about June 7
Wave 2	Reminder letter received on or about June 14
Wave 3	Questionnaire received on or about July 6 – Cancelled
Wave 4	Telephone or field non-response follow-up (as in the case of Survey group 1)
Survey group 3 – Census non-respondents (mail-out CUs)	
Wave 1	Questionnaire received on or about June 7 – Cancelled
Wave 2	None
Wave 3	None
Wave 4	Telephone or field non-response follow-up, at the same time as census non-response follow-up, starting on June 8
List/leave CUs	
Wave 1	Online offer immediately after completing the census questionnaire (Survey group 1 only)
Wave 2	Questionnaire delivered on or about June 7 to all selected dwellings (except online NHS respondents)
Wave 3	None
Wave 4	Telephone or field non-response follow-up, starting on June 8 (at the same time as census non-response follow-up, for Survey group 3)

Source: Statistics Canada, 2011 National Household Survey.

1.2.2.1 Mail-out – Survey group 1

Online census respondents selected in the NHS sample were invited to complete the NHS questionnaire immediately after finishing the census. After submitting their online questionnaire (using the original secure access code), these respondents were automatically routed to a transition page informing them that they had been selected for the NHS. The page also contained basic information about the survey, including an explanation of its importance and a note concerning its voluntary nature. Respondents who wanted to continue were then redirected to the NHS questionnaire, and the application automatically displayed the first few questions, which were the same as the census questions that the respondents had just answered. This was Wave 1 for Survey group 1.

Wave 2 for Survey group 1 was the mailing of a reminder letter to non-respondent households. The letter, delivered to households on or about June 7, made reference to the NHS's importance.

No further materials were sent out after Wave 2 (i.e., no Wave 3), and the next stage was field non-response follow-up (NRFU). Starting on June 8, follow-up was phased in, CU by CU, depending on the level of response to the census and the possibility of redeploying enumerators to help out with census collection in areas that were behind schedule.

Field procedures for the conversion of NHS non-response were similar to the census procedures. Census employees, infrastructures and systems were also used. When operationally feasible, attempts to convert refusals were made in person by team leaders and enumerators who had received special training. The definition of a refusal was the same as in the census: a clearly expressed refusal to respond to the survey. However, in view of the NHS's voluntary nature, Statistics Canada's usual refusal rules for voluntary surveys were applied: i.e., termination of follow-up efforts after two firm refusals (including telephone refusals).

In addition, some interviewers from Statistics Canada's computer-assisted telephone interviewing centres were used to supplement NHS NRFU. Those interviewers were deployed mostly in areas where NHS collection was behind schedule and there were not enough field employees.

It is worth noting that dwellings in Survey group 1 never received a paper NHS questionnaire. Since those dwellings chose to respond to the census online, it was assumed that they would do likewise for the NHS if they agreed to respond. Mailing out a paper questionnaire would have had little impact on the response rate and would have increased collection costs significantly.

1.2.2.2 Mail-out – Survey group 2

There was no indication in the paper census questionnaire that the dwelling might receive an additional, voluntary survey later. As a result, dwellings that used a paper census questionnaire in May did not have their initial contact with the NHS until about four weeks after Census Day (on or about June 7), when they received a paper NHS questionnaire in the mail. A secure access code for online response was printed on each questionnaire. This was Wave 1 for that group.

For Wave 2, a reminder letter was delivered to non-respondent households one week later, on or about June 14. Like the reminder letter sent to Survey group 1, this letter reminded non-respondents of the importance of completing the voluntary survey.

In the original plans, Wave 3 involved mailing a second paper questionnaire to non-respondent households. That questionnaire, accompanied by another reminder letter, was to have been delivered on or about July 6. However, because of the uncertainty generated by the possibility of a work stoppage at Canada Post and the opportunity to start field follow-up earlier, Wave 3 was cancelled.

Field NRFU (Wave 4) began when census collection was sufficiently advanced. The same procedures as those described above for Survey group 1 were followed.

1.2.2.3 Mail-out – Survey group 3

The wave methodology used for census non-respondents was simpler than the methodology used for the other survey groups. Since the respondents of dwellings in Survey group 3 did not react positively to any of the requests to respond to the census, which is compulsory, sending those letters or questionnaires or leaving additional voice messages urging them to complete a voluntary survey would not have had a significant effect. The plan was to send out only the NHS questionnaire for delivery on or about June 7, but that wave was also cancelled because of the possible work stoppage at Canada Post, so that priority could be given to preparing and distributing Survey group 2 questionnaires, which was taking place at the same time.

Hence, the only contact with Survey group 3 households was Wave 4, NRFU. That wave, which started on June 8, was carried out during the same field visit as census follow-up.

1.2.2.4 List/leave – Survey groups 1, 2 and 3

The dwellings in list/leave CUs received their census questionnaires in early May. As in the case of the wave methodology used for the mail-out CU survey groups, dwellings that responded to the census online using the secure access code printed on the questionnaire were invited to complete the NHS questionnaire online immediately after finishing the census questionnaire. In addition, NHS questionnaires were delivered by enumerators on or about June 7 to dwellings that used a paper census questionnaire or did not respond to the census, and to dwellings that refused the NHS offer after completing the census questionnaire online.

No other materials were delivered to the door, and telephone or field NRFU started on the same dates as for mail-out areas. For Survey group 3, NRFU for the NHS was also carried out during the same field visit as NRFU for the census.

2. Census and National Household Survey data processing

2.1 Introduction

This chapter discusses the processing of all the completed questionnaires, which encompasses everything from the reception of the questionnaires through the creation of an accurate and complete census database and a National Household Survey (NHS) database. Described below are the steps of questionnaire registration, questionnaire imaging and data capture, editing, error correction, failed edit follow-up, coding, dwelling classification and non-response adjustments, imputation, and weighting.

Automated processes, implemented for the 2011 Census and NHS, had to be monitored to ensure that all Canadian residences were enumerated once and only once and to indicate which of those residences were to be included in the NHS. 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 a daily basis with information about each dwelling's status in the census and NHS process flow (i.e., delivered, received, processed, etc.). Reports were generated and made accessible online to the managers to ensure that census and NHS operations were efficient and effective.

2.2 Receipt and registration

Responses received through the Internet or help line telephone interview were received directly to a centralized data processing centre called the Data Operations Centre (DOC) and their receipt 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. Each day, Canada Post would send a daily file listing all census and NHS questionnaires received at each regional processing plant, by date of receipt.

The registration of each returned questionnaire was flagged on the Master Control System (MCS) at Statistics Canada. A list of all the dwellings for which a questionnaire had not been received was generated by the MCS and then transmitted to field operations for follow-up. Registration updates were sent to field operations on a daily basis to prevent follow-up on households which had already completed their questionnaire.

2.3 Imaging and keying from images

In 2011, the forms imaged were: the three census questionnaires (2A, 2C, 3A), the Census of Agriculture questionnaire (F6), and the two NHS questionnaires (N1, N2). The image quality improved relative to 2006 with the replacement of black and white scanners with color scanners. The following steps were part of the imaging process:

- **Document preparation:** mailed-back questionnaires were removed from envelopes and foreign objects, such as clips and staples, were detached in preparation for scanning. The questionnaires were batched by form type. Forms that were in a booklet format were separated into single sheets by cutting off the spine.
- **Scanning:** converted the questionnaires 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, send it directly to key entry, or send it to be rescanned.

- **Automated recognition:** this step attempted to automatically recognize hand-written responses and marks on the questionnaire.

- **Key entry:** operators entered responses that automated recognition could not determine with sufficient accuracy.

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

2.4 Coverage edits

Coverage edits were applied to both census and NHS questionnaires. At this stage, a number of automated edits were performed on respondent data. These edits were designed to detect cases where invalid persons may have been created either due to respondent error or data capture error. Examples include data erroneously entered in the wrong person column, crossed off data that was captured in error, or data provided for the same person more than once, usually due to the receipt of duplicate forms (e.g., a husband or wife completed the Internet version and their spouse filled in the paper form and mailed it back). The edits were also designed to detect the possible absence of usual residents, when data are not provided for every household member listed at the beginning of the questionnaire.

About 45% of edit failure cases were resolved deterministically by the system. The remainder were forwarded to processing clerks for resolution. An interactive system enabled the clerks to examine the captured data and compare them with the image if available (online questionnaires would not have an image). Edit failures were resolved by deleting invalid or duplicate persons and adding missing ones (i.e., creating blank person records), as necessary and appropriate.

2.5 Completion edits and failed edit follow-up

Completion edits and failed edits follow-up only apply to census questionnaires. Following the coverage edits, another set of automated edits was run on census questionnaires to detect cases where there were either too many missing responses, or there were indications that data may not have been provided for all usual residents in the household. Households failing these edits were sent for follow-up. An interviewer telephoned the respondent to resolve any coverage issues and to fill in the missing information, using a computer-assisted telephone interviewing application. The data were then sent back to the DOC for reintegration into the system for subsequent processing.

2.6 Coding

Both the census and NHS questionnaires contained questions for which answers could be checked off against a list, as well as questions requiring a written response from the respondent in the boxes provided. These written responses underwent automated coding to assign each one a numerical code, using 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 for both the census and NHS: Relationship to Person 1, Home language, and Mother tongue. The following questions required coding for NHS only: Place of birth, Citizenship, Non-official languages, Ethnic origin, Population group, First Nation/Indian band, Religion, Place of residence 1 year ago, Place of residence 5 years ago, Place of birth of parents, Major field of study, Location of study, Industry, Occupation, Place of work and Language of work.

About 15 million write-ins were coded from the 2011 Census questionnaires, while about 46 million were coded from the NHS questionnaires. Overall about 87% were coded automatically, although the autocoding rate varied considerably from one variable to the next.

As the responses for a particular variable were coded, the data for that variable were sent to the edit and imputation phase.

2.7 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 dwellings in the self-enumerated collection areas of the census as occupied or unoccupied. Based on this information, adjustments were made to the census database. The DCS selected a random sample of 1,729 self-enumerated CUs that were revisited in July and August 2011 to reassess the occupancy status as of census day for each dwelling for which no response had been received. The DCS found that 13.8% of the 1,099,156 dwellings classified as unoccupied were actually occupied and that 30.8% of the 317,976 dwellings with no responses that were classified as occupied or with occupancy status classified as unknown 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 3.3% in the number of occupied dwellings, and a decrease of 5.0% in the number of unoccupied dwellings at the Canada level.

After this adjustment of the occupancy status by the DCS, occupied dwellings with total non-response had the number of usual residents (if not known) and all the responses to the census questions imputed by borrowing the unimputed responses from another household within the same CU. This process, called whole household imputation (WHI), imputed 99% of the total non-response households. Utilizing 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 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.

More details on the DCS and the whole household imputation procedure can be found in the *Coverage Technical Report*, 2011 Census, Catalogue no. 98-303-X.

2.8 Edit and imputation

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

The final clean-up of data, done in the edit and imputation process, was for the most part fully automated. Two types of imputation were applied. The first type, called 'deterministic imputation', involved assigning specific values under certain conditions when the resolution of the problem is clear and unambiguous. Detailed edit rules were applied to identify these conditions, and then the variables involved in the rules would be assigned a pre-determined value. The second type of imputation, 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 is identified, another record with most characteristics in common with the record in error was selected. Data from this donor record were borrowed and used to make the minimum number of changes to the variables in order to resolve all missing or inconsistent responses. The Canadian Census Edit and Imputation System (CANCEIS) (see CANCEIS version 5.2 Basic User Guide) was the automated system used for nearly all deterministic and minimum-change nearest-neighbour donor imputation in the 2011 Census and National Household Survey (NHS).

2.9 Weighting

In 2011, the census questionnaire consisted of the same eight questions that appeared on the 2006 Census short-form questionnaire plus two additional questions on language. These questions were asked of 100% of the population. All remaining information was collected by the National Household Survey, which was distributed to about 30% of households. Weighting was used to project the information gathered from the 30% sample to the entire population.

Sampling and Weighting Technical Report

The sampling approach used for the 2011 NHS was different from what was used for the 2006 Census long form. Therefore, the weighting methodology was also different. The first step in the weighting process was to assign basic weights that reflect the probability of the household being sampled. These weights were then adjusted for total non response. A final adjustment was done by the smallest possible amount needed to ensure closer agreement between the sample estimates and the census counts for a number of characteristics related to age, sex, marital status, common-law status, language, and household size. The weighting methodology is described in detail in Chapter 4.

3. Sampling[2]

3.1 The National Household Survey sample

The National Household Survey (NHS) is a sample survey designed to collect detailed demographic, social and economic information about the Canadian population. The sample was drawn from the 2011 Census of Population dwelling list. Only occupied private dwellings and their corresponding households were in-scope for the NHS. Thus, unlike the census, all collective dwellings and households outside Canada (e.g., diplomats, military personnel) were out of scope for the NHS. At the time that the NHS sample was selected, it was not always known which addresses were linked to out of scope dwellings, meaning that some out of scope dwellings erroneously received an NHS questionnaire. Once a dwelling was determined to be out of scope, no further collection or processing activities were carried out. The NHS questionnaire was distributed to about 30% of private households.[3]

The proportion of dwellings to be selected for the NHS was determined at the collection unit (CU) level. The census collection method (e.g., mail-out, list/leave, canvasser) used for any given CU was also used to help determine its sampling fraction for the NHS.

All private dwellings in CUs whose census information was collected via canvasser were selected for the NHS. This included dwellings on Indian reserves. Canvasser areas included approximately 200,000 occupied private dwellings. In list/leave areas, private dwellings were sampled at a rate of one in three.

In mail-out areas, private dwelllings were selected according to sampling fractions that were calculated at the provincial/territorial level. These fractions were derived in order to reach the desired fixed national sample size of 4.5 million dwellings, while obtaining equal provincial/territorial sampling fractions for the combined list/leave and mail-out area. Table 3.1.1 presents the different sampling fractions for the provinces and territories in which questionnaires were mailed out.[4]

Table 3.1.1 Provincial/Territorial sampling fractions for mail-out CUs

Province and territory	Mail-out sampling fraction
Newfoundland and Labrador	0.1973
Prince Edward Island	0.2091
Nova Scotia	0.2488
New Brunswick	0.2678
Quebec	0.2768
Ontario	0.2752
Manitoba	0.2640
Saskatchewan	0.2473
Alberta	0.2725
British Columbia	0.2772
Yukon	0.2621
Northwest Territories	…
Nunavut	…

… not applicable

Source: Statistics Canada, 2011 National Household Survey.

In mail-out and list/leave areas where self-enumeration was used, sampled households were selected based on a stratified systematic random sampling design (stratified by province and territory and collection mode). In the list/leave

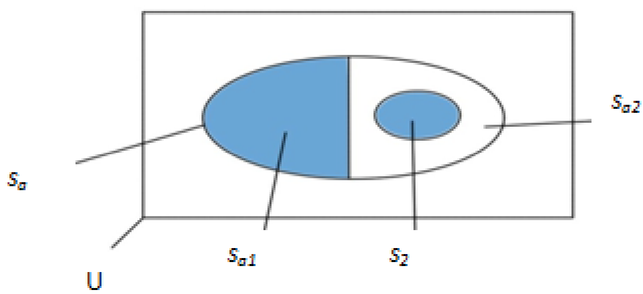
2. For history on the sampling in Canadian Censuses, please refer to Appendix B.
3. A private household refers to a person or group of persons (other than foreign residents) who occupy a private dwelling and do not have a usual place of residence elsewhere in Canada.
4. There were no mail-out CUs in the Northwest Territories (N.W.T.) and Nunavut. In the N.W.T., in list/leave areas, private dwellings were sampled at a rate of one in three while private dwellings in CUs whose NHS information was collected via canvasser were all selected. In Nunavut, NHS information was collected via canvasser for 100% of the private dwellings.

CUs, one out of three dwellings was selected. Dwellings in the mail-out CUs were selected using provincial and teritorial sampling fractions that were calculated so as to (1) generate the desired total sample size at the Canada level, and (2) produce the same sampling fraction for each province and territory (mail-out and list/leave collection modes combined).

3.2 NHS subsample

It was determined prior to the collection of NHS data that the resources available for NHS non-response follow-up (NRFU) would not allow for follow-up on all non-respondents. It was therefore decided that as of a certain date, a subsample of the remaining non-responding cases would be selected. This would ensure that field staff would have a manageable number of cases with which to work, and it would reduce the risk of non-response bias[5] by allowing staff to focus on certain areas or dwellings. All cases that were not selected in the NRFU subsample would therefore be excluded from further collection activities. This approach is a form of sampling known as ‘two-phase sampling’. The dwellings originally sampled for the NHS form the first phase sample; the subsampled dwellings for the NRFU form the second phase sample. Figure 3.2.1 illustrates the design. Rectangle U represents the dwellings of the census. The NHS sample is shown by the large ellipse s_a, where s_{a1} represents the dwellings that responded to the NHS by July 14, 2011. The remaining portion of the oval, which represents dwellings that did not respond by that date, was then split into two parts. The subsample of non-respondents for NRFU is represented by the small oval s_2 and the non-respondents not selected for NRFU are in s_{a2}.

Figure 3.2.1 Sampling design of the NHS and NRFU



Source: Statistics Canada, 2011 National Household Survey.

The two-phase sampling approach was proposed initially by Hansen and Hurwitz (1946) and has proven to be an efficient method to be used when the second phase collection cost per unit is higher. It is also effective when a more concentrated effort is used to target a smaller number of non-respondents for NRFU. The goal for the NHS was to put the maximum reasonable effort on fewer selected follow-up cases rather than diluting the effort by spreading the available resources over all non-response cases. At the same time, the NRFU sub-sampling methodology can take advantage of a combination of frame information and paradata in order to allocate the resources in an efficient manner, leading to a reduction in non-response bias.

The following cases were deemed ineligible for NRFU subsampling:

- Those that were identified as respondents at the time of subsampling.
- Those for which NRFU appointments already had been made because they were considered to have a relatively greater chance of becoming full responses in the near future.
- Those in areas using the canvasser method of enumeration.
- Those that were added to the NHS sample after the initial sample selection since no questionnaires had been sent.
- Cases corresponding to an unused Visitation Record line or dwellings that could be identified prior to subsampling as a dwelling with only temporary or foreign residents (TR/FR).

5. ‘Bias’ refers to the difference between the average of the estimates, over all possible samples, and the true population value.

3.3 Targeted CUs

An advantage of creating a subsample for NRFU is that non-respondents living in certain regions or having certain characteristics that are less likely to respond to the NHS can be targeted to improve their response rate. Therefore, because these non-respondents typically have a lower propensity to respond, it was decided to oversample certain pre-determined CUs to ensure a good representation of specific populations in the NHS subsample. Table 3.3.1 lists five populations (target domains) that were oversampled because based on other surveys or past censuses, they were considered to be at risk of having a lower response rate and they are considered of high interest to potential statistical analysis.

Table 3.3.1 At-risk populations identified to be oversampled in NRFU

Characteristic	Description	Level
Aboriginals (outside reserves)	North American Indian, Métis, Inuit, or any combination of the three	Person
Recent immigrants	Non-permanent residents and Canadians attaining immigration status between 1996 and 2006	Person
Visible minorities	All non-White groups except for Chinese[6] and Aboriginal	Person
Low degree of education	Those in the labour force with no education certificate beyond a high school diploma	Person
Low level of income	Household income between $0[7] and $20,000	Household

Source: Statistics Canada, 2011 National Household Survey.

The oversampling strategy identified target areas that contained a relatively high proportion of persons or households in one or more target domains. Unweighted data from the 2006 Census were used for the creation of the target domains. The dwellings in 2006 were matched with a CU from 2011, and only CUs containing mail-out or list/leave collection methods were considered for oversampling.

3.4 NRFU subsample selection

In order to determine the NRFU subsample, the country was divided into 23,901 strata. Any CU that was targeted for oversampling served as a stratum; otherwise, a stratum was a grouping of usually two adjacent CUs called enumerator zones. All strata were represented in the NRFU subsample. If a stratum had one or two eligible dwellings, those dwellings were automatically added to the subsample. Only 196 dwellings belonged to these strata. For strata with three or more eligible dwellings, the number of dwellings to be selected was calculated to be proportional to the size of the stratum. The subsample size for the strata that were flagged for oversampling was inflated by a factor of 1.6.

Once the adjusted stratum subsample sizes had been calculated, the NRFU subsample was selected on July 14, 2011 using systematic sampling with a fractional sampling interval. In all, 642,442 dwellings were selected for the NRFU subsample, which represented approximately 33.5% of remaining eligible NHS dwellings. Targeted CUs contributed 169,657 dwellings, which was 26.4% of the entire NRFU subsample. The subsample distribution by province and territory is shown in Table 3.4.1.

6. Canadians of Chinese descent were excluded from the visible minority group as this group has traditionally responded at a relatively high rate.
7. Household income is sometimes reported in negative values. For sampling purposes, these households are not considered to be low income.

Table 3.4.1 NRFU subsample distribution by province and territory

Province and territory	Total
Newfoundland and Labrador	11,884
Prince Edward Island	3,160
Nova Scotia	19,832
New Brunswick	16,098
Quebec	144,767
Ontario	244,309
Manitoba	23,770
Saskatchewan	22,666
Alberta	66,889
British Columbia	88,311
Yukon	613
Northwest Territories	143
Total	**642,442**

Source: Statistics Canada, 2011 National Household Survey.

4. National Household Survey estimation

Any sampling process requires an associated estimation procedure for scaling sample data up to the population level and to ensure 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 in the estimates it produces. Section 4.1 describes the operational and theoretical considerations relevant to the choice of estimation procedures. Sections 4.2 to 4.6 focus on the details of the estimation in the 2011 National Household Survey (NHS), including the definition of the NHS universe, design weights, total non-response weight adjustments, and final weight calibration processes.

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 between the different estimates of the survey.

Therefore, the approach taken in the NHS (and in most sample surveys) is 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 characteristic by summing the weights of those persons or households having that characteristic. Most of the mathematical complexity is then 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. It should be noted that 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. The 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. The usual 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 population value for approximately 95% of all possible samples.

As well as minimizing standard error, a second objective in the choice of estimation procedure for the NHS 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 NHS 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.

In cases where there is no information about the population being sampled other than that collected for sample units, and there has not been unit non-response, the estimation procedure would be restricted to weighting the sample units inversely to their probabilities of selection. For example, if a unit had a one-in-three chance of selection, then that selected unit would receive a weight of 3. When unit non-response is observed, the weight also has to be further adjusted using the probability of response of the unit. Also, in practice, one almost always has some supplementary knowledge about the population (e.g., its total size, and possibly its breakdown by a certain variable—perhaps by province and territory). Such information can be used to improve the estimation formula so as to produce estimates with a greater chance of lying close to the unknown population value. In the case of the NHS sample, a large amount of very detailed information about the population being sampled was available in the form of the census data at every geographic level. We can take advantage of this wealth of population information to improve the estimates made from the NHS sample. This will be discussed later in this report.

4.1.3 Additional considerations for the National Household Survey weighting

Just as in previous censuses, every household in the NHS was assigned a weight so that the characteristics of NHS respondents can be weighted and combined to produce estimates for the population. However, there were at least two major issues that made the process more complex in 2011. First, census and NHS collection were separate events. Having two sets of questionnaires being asked at two different times led to complications such as household linkage and data inconsistencies between the two sources. This was mentioned in Section 2.9. Secondly, the NHS had to deal with relatively high household non-response. Chapter 3 discussed the problem of total non-response and the design of the non-response follow-up (NRFU) process to improve response rates.

In a survey, the procedure of weighting the sample units inversely to their probabilities of selection (or probabilities of response in the presence of a total non-response adjustment) should result in small differences between the sample estimates and the census counts for large subgroups of the population. However there could be significant differences for smaller subgroups. These differences were usually made greater by the cases of total non-response in the NHS.

It is difficult to make the NHS sample estimates for census characteristics 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 counts tabulations for those characteristics.

Adjusting the weights, equal to the inverse of the probabilities of response, by small amounts to achieve perfect agreement between estimates and census counts for certain subgroups is known as ‘calibration’. This procedure is further discussed in Section 4.6.

4.2 NHS universe

The census household universe can be broken into three: the private households, the collective households, and the households outside Canada. The NHS household universe corresponds to private households that were eligible for the 2011 Census. Unless specified otherwise, the term ‘in-scope’ will be used in this document to indicate that a household is part of the NHS universe (i.e., private households) while ‘out of scope’ refers to households not in the universe (i.e., collective and outside Canada households).

An exception to the NHS universe involves the private households of five census subdivisions (CSDs) corresponding to five Indian reserves. They were excluded from the universe because of a very low response rate in the NHS. If they had not been excluded, then surrounding areas within the same weighting areas (WA) would have been greatly affected by the non-response and calibration weight adjustments that were necessary to compensate for their low response rates. The exclusion of those five CSDs is an example of some of the differences that can be observed when comparing a census publication to an NHS publication. The number of private households and the population of those five CSDs can be found in Table 4.2.1.

Table 4.2.1 Census subdivisions not in the NHS universe

Province	Census division	Census subdivision	Number of private households	Population
Quebec	La Haute-Côte-Nord	Essipit	111	268
Ontario	Cochrane	Factory Island 1	384	1,414
Ontario	Kenora	Sandy Lake 88	459	1,861
Manitoba	Division No. 19	The Narrows 49	337	826
Saskatchewan	Division No. 14	Opaskwayak Cree Nation 27A (Carrot River)	72	286
Total			**1,363**	**4,655**

Sources: Statistics Canada, 2011 Census and 2011 National Household Survey.

4.3 Design weights

Every dwelling of the first phase sample (i.e., selected for the NHS) was given a first phase basic weight equal to the inverse of the probability of selecting that dwelling in the first phase sample. Every subsampled dwelling in the second phase sample (i.e., selected for non-response follow-up [NRFU]) was also given a second phase basic weight equal to the inverse of the probability of selecting that dwelling in the subsample.

The design weights were generally calculated as follows:

1. Dwellings that were not eligible for NRFU subsampling were assigned a design weight equal to their first phase sample weight. The vast majority of these corresponded to households identified as having responded to the NHS.

2. Dwellings that were eligible for NRFU subsampling and were not selected for NRFU had their design weight set to 0.

3. Dwellings that were subsampled for NRFU were assigned a design weight equal to the product of their first and second phase basic weights.

4.4 Total non-response adjustment

There are different types of non-respondent households in the NHS. Households that were identified as having responded by the time of subsampling were not eligible for subsample selection. However, it was discovered after careful evaluation of their questionnaires that some of them, despite having returned the questionnaire, had not in fact provided responses and so were actually non-respondents. Furthermore, many dwellings selected for NRFU did not contribute information despite attempts to obtain a response. The overall unweighted response rate was 68.6%, and the weighted response rate was 77.2%.

Various strategies for the treatment of total non-response that made use of auxiliary data available for both respondents and non-respondents were studied. The imputation approaches were attractive in the NHS context given that census data were available for the vast majority of non-respondents. In other words, unit non-response to the NHS can then be viewed as an item non-response problem. Unfortunately, the imputation approaches were not always successful because the large number of matching variables made it often impossible to find a perfect donor. A perfect donor would be a respondent household that has the same value as the non-respondent household for every matching census variable. This led to significant increases of occurrence in the data of combinations of census variables and non-census imputed variables that are rare in the population. After thorough analysis, these imputation strategies were discarded and it was decided that non-respondent households that were initially flagged as having responded and households in the subsample that did not respond would be assigned a weight of 0. The weights that they would have had if they had responded were transferred to their nearest neighbours. The method, a re-weighting approach, can be divided into the following three main steps.

The first step was to determine the auxiliary variables that best predicted the households' propensities to respond. Many auxiliary variables from the census and NHS, from linkages to 2010 tax files, the Indian Register, and 1980 to 2011 immigration files, and variables related to Indian reserves (where applicable), were considered in the construction of logistic regression models at the Canada level using forward variable selection. All the auxiliary variables were then assigned a relative weight according to their predictive power in the response model that was selected. Although some auxiliary variables did not enter the final models, every variable was assigned a minimum positive weight.

In the second step, the Canadian Census Edit and Imputation System (CANCEIS) (see CANCEIS version 5.2 Basic User Guide) was used to locate 20 nearby respondent households that best matched each non-respondent household. The chosen households had the same number of members and were usually in the same neighbourhood or a neighbourhood near the non-respondent household. The matching process used the auxiliary variables and their relative weights from the first step. A match score between 0 and 1 was given for each matching variable, where a score of 0 occurred if the values were the same and a value greater than 0 (usually 1) was given if they were different. These scores were multiplied by the CANCEIS imputation weight of the variable and summed over all variables to calculate a distance score between the non-respondent and the respondent. The 20 respondents with the lowest scores were identified.

The third step consisted of transferring the weight of the non-respondent household to each of the 20 respondent households identified in the second step. The amount of weight distributed was proportional to the inverse of the distance between the respondent household and non-respondent household. Therefore, respondent households that better matched the non-respondent household received a greater share of the weight that was being transferred.

4.5 Surprise respondents

Surprise respondents are households that were from dwellings selected in the first phase sample that had not responded before subsample selection, but which responded afterward without being part of the NRFU subsample. The surprise respondents were combined with respondent households in the second phase sample. Their design weight was set to 0 as described in Section 4.3. Instead of leaving these surprise respondents with their initial weight of 0, their weight was forced to 1 and the weight of some respondents in the subsample was reduced accordingly. This was done because a weight of 1 minimised bias while making sure the surprise respondents were self-represented.

Similar to the total non-response adjustment, CANCEIS was used to identify the set of neighbours consisting of the 20 closest respondent households in the second phase sample for each surprise respondent. A weight of 1/20 was then subtracted from each of these neighbour respondents and transferred to surprise respondents.

4.6 Calibration

Once the NHS design weights were adjusted to account for total non-response and surprise respondents, they were adjusted one last time in a calibration process. The process slightly adjusted the weights in order to satisfy a certain number of constraints. While this ensures consistency between NHS estimated totals and census counts for many constraints, the effect of calibration is a reduced sampling variability in the survey estimates. During this final adjustment step, it was important to limit the range of the calibrated weights so that no unreasonable amount of weight was placed on any household or person. Therefore, weights were constrained to be between 1 and 100.

Appendix C lists all the constraints that were considered during the calibration process. They included the same constraints as in 2006 as well as additional constraints involving census and economic family variables and language variables. Characteristics available from both the census and the NHS for which consistency was attempted include information such as age, sex, marital status, common-law status, household size, dwelling type, mother tongue, home language, etc.

Calibration and constraint selection was carried out independently in geographic areas known as weighting areas (WA). There were 5,884 WAs in Canada, most of which consisted of a contiguous area of land within a census division (CD). On average a WA contained about 2,300 households in the census, and most WAs contained between 300 and

699 households that were either initial NHS respondents or households that were selected for NRFU. See Chapter 6 for additional information about the construction of WAs.

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

1. Mandatory constraints – These constraints must be calibrated because they must have agreement between the census count and the NHS estimates at geographic levels which are usual aggregates of WAs (e.g., Canada, provincial and territorial). The number of persons in the WA and the number of households in the WA were the two mandatory constraints.

2. Low response constraints – Constraints that would be applied to less than 30 responses should not be calibrated because they can make survey estimates unstable.

3. All other constraints – These constraints were examined further to see if they should be calibrated.

The second step was to determine the constraints from the third group to be used in the calibration process in addition to the mandatory constraints. The constraints from the third group were added one by one beginning with those that were least linearly dependent with the low response constraints and the constraints already included. Constraints that were too linearly dependent were not included.

Calibration was then performed on the final set of constraints from step 2 by adjusting the household weights as little as possible so that the sum of the weighted estimates equalled the census counts for those constraints. In practice, the calibration process was performed using software developed by Statistics Canada called StatMx, which was programmed entirely in SAS language. See Verret (2013) for more details.

There are a few reasons why sample estimates may be different from census counts, particularly for small areas, even after the calibration step. The main ones are listed below.

1. Constraints dropped during the calibration process: As described earlier, constraints could be dropped for having small counts, by being linearly dependent with other constraints that were chosen, or by being nearly linearly dependent with constraints having small counts. Constraints that were dropped were not controlled upon, and usually led to some differences between census counts and NHS estimates.

2. Sub-WAs: The WA was the smallest geographic area for which the weighting system attempted to have agreement between the census counts and the NHS estimates for as many auxiliary variables as possible. Any place that is smaller than a WA, such as most DAs, is known as a sub-WA. These could have discrepancies between the census counts and the NHS estimates.

The point above is important in areas of high non-response. This is because WAs had to be larger than usual to contain 300 to 699 respondent households. As a result, some small communities or municipalities with low response rates were sub-WAs, meaning that consistency of their estimates with census counts cannot be guaranteed.

5. National Household Survey non-response bias indicators

5.1 Introduction

Because of the voluntary nature of the NHS, Statistics Canada put a great deal of effort into maximizing household participation, particularly using mixed collection methods and the wave collection methodology (see Chapter 1 on Data collection). Despite these efforts and a response rate similar to that of other voluntary surveys conducted by Statistics Canada (68.6%), almost one in three households did not participate. At the processing stage, total non-response was treated using a reweighting methodology. The purpose of reweighting is mainly to reduce the non-response bias in the estimates.

Non-response has two possible effects on the data. First, it contributes to an increase in the total variance of the estimates, since the actual sample size is smaller than the planned size. Second, it can introduce a bias in the estimates when non-respondents differ from respondents in the characteristics measured.[8] This is referred to as non-response bias. Consequently, to assess data quality, it is recommended that efforts be made to determine whether there is a potential non-response bias and, if possible, to derive an indicator of its magnitude.

A non-response bias indicator is a tool attempting to identify the presence of a potential residual bias, i.e., after reweighting. Non-response bias occurs when the characteristics of interest of respondents differ from those of non-respondents. Unlike the sampling variance, it does not necessarily decline with an increase in sample size.

A positive non-response bias indicator indicates that the estimator overestimates the parameter of interest on average, and a negative non-response bias indicator indicates an underestimate. The magnitude of the non-response bias indicator indicates the extent of the potential bias in the estimate. Thus, a high value for the non-response bias indicator may indicate substantial differences between the characteristics of respondents and the characteristics of non-respondents.

5.2 Methodology of non-response bias indicators

The indicators of bias due to unit non-response were produced from a linked file of the 2011 and 2006 censuses. Using family name, address and date of birth, 73% of 2011 Census respondents were linked to their responses from the 2006 Census. This included responses from the 2006 long-form for a large portion of the 2011 NHS sample, including both respondents and non-respondents to the NHS.

The weights of the resulting matched sample were calibrated to estimates of the number of persons in the population common to both the 2006 and 2011 censuses. This calibration was done separately for 2011 NHS matched respondents and matched non-respondents. A second set of weights was calculated for the matched respondents. These weights used the 2011 non-response adjustment factors and were calibrated to the estimates of the total number of persons in the population common to both the 2006 and 2011 censuses. Using the two sets of weights, benchmark totals were obtained from the entire matched sample (i.e., 2011 NHS respondents and non-respondents) using the 2006 Census long-form questionnaire data and adjusted totals were obtained from the matched respondents sample again using the 2006 Census long-form questionnaire data.

The difference between an adjusted total and a benchmark total was used as an indicator of the bias due to total non-response in the 2011 NHS. Algebraically, for category *k* of the 2006 variable x_j this bias indicator can be expressed as $\hat{B}(\hat{X}_{jk}^a) = \hat{X}_{jk}^a - \hat{X}_{jk}$ with the benchmark $\hat{X}_{jk} = \sum_{i \in s_m} w_i x_{ijk}$ and the adjusted total $\hat{X}_{jk}^a = \sum_{i \in s_{m,r}} a_i w_i x_{jk}$, where, for unit i, a_i is a total non-response adjustment, w_i is a calibrated weight, x_{ijk} is the value of variable j for category k, s_m is the matched sample and $s_{m,r}$ is the set of 2011 respondents in the matched sample.

8. Statistics Canada (2009). Statistics Canada. Quality Guidelines, 5th edition, Catalogue no. 12-539XWE.

5.3 Limitations of the NHS non-response bias indicators

There are several limitations to these indicators. The indicator reflects total non-response bias and thus does not account for item non-response. It is an indicator for the population in common between 2006 and 2011. In other words, the potential bias for births, recent immigrants (last 5 years) and non-permanent residents (last 5 years) is neglected. The indicator is based on a matched sample, so unmatched units that were part of the common population are not represented. The weight adjustment that accounts for the unmatched portion is likely imperfect. The indicator is based on the 2006 variables which are proxies to the 2011 variables. It lacks precision for geographic areas with a small matched sample, often the smaller geographic areas. Finally, its reliability is also affected by non-response to the 2006 Census short and long forms as well as to the 2011 Census. In sum, the bias indicators are exactly as the name implies, indicators of bias and not measures or estimates of bias.

5.4 Some NHS non-response bias indicators

For illustrative purposes, the non-response bias indicators shown in Table 5.4.1 were calculated at the Canada level for a few characteristics where there is potentially a higher risk of bias.

- Immigration status (landed immigrant, non-permanent resident)
- Aboriginal identity (North American Indian, Métis, Inuit)
- Visible minority (South Asian, Chinese, Black, Filipino, Latin American)
- Highest certificate, diploma or degree (no certificate, diploma or degree; secondary (high) school diploma or equivalent; trades certificate or college diploma;[9] bachelor's degree).

Table 5.4.1 Bias indicators at the Canada level for the 'Immigration status,' 'Aboriginal identity,' 'Visible minority' and 'Highest certificate, diploma or degree' characteristics

Characteristic	Bias indicator (%)
Immigration status	
Landed immigrant	-1.2
Non-permanent resident	-0.8
Aboriginal identity	
North American Indian	1.7
Métis	0.2
Inuit	6.1
Visible minority	
South Asian	1.3
Chinese	0.9
Black	-3.4
Filipino	7.3
Latin American	-1.1
Highest certificate, diploma or degree	
No certificate, diploma or degree	-0.7
Secondary (high) school diploma or equivalent	-0.9
Trades certificate or college diploma	0.0
Bachelor's degree	1.9

Source: Statistics Canada, 2011 National Household Survey.

9. This category includes people whose highest level of education is a registered apprenticeship or trades certificate or diploma, or a college, CEGEP or other non-university certificate or diploma.

6. Weighting areas

Weighting areas (WA) are the geographical partitions that were used during the National Household Survey (NHS) weighting procedures. In the 2011 NHS, Canada was partitioned into 5,884 WAs. The weighting process attempted to achieve agreement between certain sample estimates and the corresponding census counts for each WA. Some changes were made to the design of the WAs in 2011 because the non-mandatory NHS resulted in different response rates for different parts of the country. Weighting areas were designed in two steps in 2011: the regular WA formation and the canvasser and reserve adjustment.

6.1 Regular WA formation

The first step in creating a set of WAs was to group together dissemination areas (DAs) while adhering to the following conditions:

(a) A WA must respect the boundaries of census divisions (CD) (i.e., a WA can only be found in a single CD).

(b) A WA should, where possible, contain between 300 and 699 households that responded to the NHS or were selected in the subsample (both subsample respondents and non-respondents). For simplicity in this chapter, these households will be known as respondent households.

(c) A WA should, where possible, respect (in order of priority) census subdivision (CSD) boundaries and census tract (CT) boundaries.

(d) A WA should, where possible, be made up of contiguous DAs (i.e., not be in two or more parts or contain any 'holes') and it should be as compact as possible.

The conditions above are very similar to those in 2006, with the main exception being condition (b). In 2006 the WA size was based on the number of households of the census that were subject to sampling, which in most cases was between 1,000 and 3,000 households. If WAs had been created in the same fashion in 2011, the number of respondent households would have varied widely between WAs. Calibration would not have been effective in some of the low-response WAs.

To remedy this problem in 2011, WAs were created based primarily on the number of respondent households, even though it meant that WAs in low-response areas would have to include a higher number of in-scope dwellings than in the past. The 2006 Census was sampled at approximately one in five, meaning that each WA with 1,000 to 3,000 in-scope households would have had approximately between 200 and 600 respondent households in the long-form sample. This interval was increased slightly in 2011 to include between 300 and 699 to allow for additional non-response and to give more census counts and NHS estimate agreement. Increasing the WA by more than this would have caused significantly more CSD boundary conflicts in condition (c).

The fourth condition was generally simple to apply in urban areas. However, the condition of having contiguous DAs had to be relaxed in areas where two or more disjoint pockets of land belonged to the same municipality. The condition was also relaxed if there were several small or low-responding municipalities in a CD that needed to be put into a single WA in order to satisfy condition (b). This most commonly occurred in areas with reserves and in the North.

The algorithm that was used in the first step of the WA formation process in 2011 was the same as in 2006. In both cases, many manual adjustments were done to abnormal WAs by splitting, joining, or realigning WAs to better fit the conditions described earlier. After the manual adjustments in the first step of 2011, 99.3% of all WAs had the desired number of households—which was the same rate as in 2006. Table 6.1.1 gives some statistics on how well the WAs fit together with CSDs and CTs. The descriptions of each category are provided after the table.

Table 6.1.1 Number of census subdivisions and census tracts that respect weighting area boundaries, first step, 2011 Census

Scenario	Description	CSD		CT	
		Number	Percentage (%)	Number	Percentage (%)
1	CSD or CT was small enough to fit entirely within a WA, and the same WA only consisted of whole CSDs or CTs. None of the CSDs or CTs in that WA crossed into a different WA.	3,729	70.6	2,408	44.1
2	CSD or CT was small enough to fit entirely within a WA, but a different CSD or CT within that same WA was shared by a different WA	669	12.7	203	3.7
3	CSD or CT was large enough to contain whole WAs. None of the WAs crossed into a different CSD or CT	674	12.8	2,725	49.9
4	CSD or CT is shared by at least two WAs	181	3.4	129	2.4
Total		**5,253**	**100.0**	**5,465**	**100.0**

Source: Statistics Canada, 2011 Census.

In scenario 1, condition (c) was satisfied. This scenario occurred frequently for CSDs because there were many very small municipalities such as Indian reserve and villages that contributed less population than was required to create a WA. In scenario 2, condition (c) is not satisfied. In scenario 3, condition (c) was satisfied. In scenario 4, condition (c) is not satisfied.

6.2 WA adjustment for canvasser and reserve areas

The second step in creating WAs was new for 2011. In order to create more homogenous WAs, households collected by a canvasser were processed together. These households were removed from their original WAs and placed into a small number of special WAs known as canvasser WAs. In order to generate WAs with 300 to 699 respondent households, the restriction of being in the same CD and occasionally the same province or territory was lifted. Similarly, households on Indian reserves were also processed together with the same aboriginal tribe wherever possible. Households that resided on Indian reserves, excluding land that had been leased for non-aboriginal dwellings, were placed in a small number of special WAs called reserve WAs. Most reserve WAs also contained 300 to 699 respondent households and were permitted to cross CD boundaries and occasionally provincial or territorial boundaries. In fact, 15 of the 5,884 WAs crossed provincial and territorial boundaries and appeared in exactly two provinces or territories, and one WA appeared in British Columbia, Alberta, and the Northwest Territories at the same time. Table 6.2.1 shows the number of WAs that crossed CD boundaries, and particularly the number of CDs that they touched.

Table 6.2.1 Number of WAs associated with multiple census divisions

Number of CDs as part of a WA	Number of WAs
1	5,793
2	58
3	20
4	8
5	2
6	1
7	0
8	1
9	0
10	0
11	1
Total	**5,884**

Source: Statistics Canada, 2011 National Household Survey.

Table 6.2.1 shows that 5,793 of the 5,884 WAs remained in a single CD, while 58 WAs were involved in two CDs. The highest number of CDs touched by a single WA was 11.

All respondent households that were not placed in canvasser or reserve WAs remained in regular WAs. In most cases, these households remained in their original WA. However, because of the household extraction for canvasser and reserve WAs, some original WAs had a reduced number of households. If the original WA was reduced to a small number of respondent households (usually less than 200), its households were reassigned to nearby regular WAs and the original WA was eliminated. However, if the original WA retained a significant number of households (usually at least 200), then the WA was left alone, even if the number of households fell below the desired 300 to 699 range. Table 6.2.2 shows the breakdown of the number of WAs by WA type after step 2 was completed.

Table 6.2.2 Frequency of weighting area type

WA type	Region	Number of WAs
Regular	Newfoundland and Labrador	82
	Prince Edward Island	24
	Nova Scotia	158
	New Brunswick	125
	Quebec	1,501
	Ontario	2,063
	Manitoba	199
	Saskatchewan	162
	Alberta	578
	British Columbia	756
	Yukon	5
	Northwest Territories	10
	Nunavut	0
	Canada	5,663
Canvasser	Canada	39
Reserve	Canada	182
All	**Canada**	**5,884**

Note:
Statistics for canvasser and reserve WAs were only provided at the Canada level because of their ability to cross provincial and territorial boundaries.

Source: Statistics Canada, 2011 National Household Survey.

Table 6.2.3 shows the number of WAs by WA type and intervals of the number of respondent households. As specified in Section 6.1, most WAs should contain 300 to 699 respondent households. The table shows that 5,736 (97.5%) of the final 5,884 WAs were within the range of 300 to 699 in 2011.

Table 6.2.3 Distribution of WAs by the number of respondent households and the WA type

Respondent households	2011 National Household Survey				
	Regular WA	Canvasser WA	Reserve WA	Total	Percentage (%)
0 to 99	2	0	6	8	0.1
100 to199	5	0	8	13	0.2
200 to 299	84	1	5	90	1.5
300 to 399	1,981	16	84	2,081	35.4
400 to 499	1,797	10	38	1,845	31.4
500 to 599	1,212	8	13	1,233	21.0
600 to 699	564	3	10	577	9.8
700 to 799	11	1	8	20	0.3
800 to 899	1	0	5	6	0.1
900 to 999	1	0	4	5	0.1
1,000+	5	0	1	6	0.1
Total	**5,663**	**39**	**182**	**5,884**	**100.0**

Source: Statistics Canada, 2011 National Household Survey.

The set of occupied private dwellings and their corresponding households in the 2011 Census were considered in-scope for the 2011 NHS. However, due to the varying response rates in the NHS, the size of each WA was dependent on the number of respondent households in the NHS and not the number of in-scope households. After determining the WAs in the NHS, the number of in-scope census households was also determined. Table 6.2.4 shows the number of WAs by WA type as well as intervals of the number of in-scope households in the census.

Table 6.2.4 Distribution of WAs by the number of in-scope census households and WA type

In-scope households	2011 National Household Survey				
	Regular WA	Canvasser WA	Reserve WA	Total	Percentage (%)
0 to 499	1	12	102	115	2.0
500 to 999	5	27	72	104	1.8
1,000 to 1,499	217	0	5	222	3.8
1,500 to 1,999	1,595	0	2	1,597	27.1
2,000 to 2,499	1,790	0	0	1,790	30.4
2,500 to 2,999	1,285	0	1	1,286	21.9
3,000 to 3,499	615	0	0	615	10.5
3,500 to 3,999	143	0	0	143	2.4
4,000 to 4,499	10	0	0	10	0.2
4,500 to 4,999	0	0	0	0	0.0
5,000+	2	0	0	2	0.0
Total	**5,663**	**39**	**182**	**5,884**	**100.0**

Sources: Statistics Canada, 2011 Census and 2011 National Household Survey.

The majority of WAs contained between 1,500 and 2,999 in-scope census households, which is in line with the 2006 Census in which WAs were designed to have between 1,000 and 3,000 households.

7. Evaluation of weighting procedures

As described in Chapter 4, the estimation process for the National Household Survey (NHS) involved the assignment of weights. Each household was first assigned a design weight that was determined by the sample design of the NHS. Some adjustments to these weights were then required to address total non-response. These adjusted weights, known as initial weights or pre-calibration weights, were adjusted further in the calibration process to produce final household weights. These final weights allowed for generally better agreement between the census counts and the NHS estimates for common variables between the two surveys. During calibration, the characteristics from Appendix C were used as constraints. Chapter 4 discussed how some constraints were dropped in order to get better census counts/NHS estimates agreement for other characteristics.

This chapter presents and evaluates certain aspects pertaining to the National Household Survey (NHS) weighting procedures. It examines the frequency that certain constraints were dropped during calibration as well as their effect on the NHS estimates. The chapter also examines the distribution of the weights and, for various characteristics, the discrepancies between census counts and sample estimates at various geographic levels.

7.1 Discarding constraints

The purpose of calibration is to adjust the household weights so that the resulting NHS estimates are as close as possible to the census counts for many common characteristics. Calibration also needs to make exact census counts/NHS estimate agreement for any mandatory constraints. See Appendix C for the list of constraints and Section 4.6 for more information on calibration. The addition of language variables for the 2011 Census means that there were many more constraints than in the 2006 Census.

Calibration was performed in 5,884 independently processed weighting areas (WA). In each WA, all 60 constraints entered calibration and were only dropped if necessary. The total persons and total households variables were the minimal set of constraints, meaning that they could not be dropped in any of the WAs. Other constraints were dropped or removed as required for the following reasons:

1. No population – If the constraint had no population in the WA, then the estimate must also be 0 for that constraint. This constraint does not contribute to the calibration process. We do not classify these constraints as being dropped, but rather as being ineligible for calibration.

2. Small sample – If the number of NHS respondents for a constraint in a given WA is more than 0 but less than 30, then using such constraints would reduce the accuracy of aggregated estimates. These constraints were therefore dropped.

3. Linearly dependant and nearly linearly dependant – If a constraint value can be calculated by the combination of other constraint values, then one of those constraints must be dropped because of linear dependence. For example, the value of the marital status constraints Married, Single, Separated, Divorced, and Widowed add to the TOTPERS constraint. At least one of these is not required and can be dropped. Age and language variables contribute to other groups of constraints that lead to linear dependence because they also add up to the TOTPERS constraint. The household size constraints also add up to the number of households' constraint, so at least one of these must be dropped.

   If a constraint is dropped for having small sample (reason 2), its value may be determined by subtracting the combination of other constraint values not yet dropped. In this case, one of the remaining constraints also has to be dropped because of linear dependence. If another constraint is not dropped, it would be equivalent to retaining the small sample constraint dropped in reason (2). For example, suppose that the marital status constraint Widowed is dropped for having less than 30 respondents in a WA. However, this value of Widowed can be retrieved by subtracting the remaining constraints Married, Single, Separated, and Divorced from the TOTPERS constraint. Consequently, an additional constraint needs to be removed. Summing the remaining constraints, Married, Single, Separated, and Divorced, would have approximately equalled the Total Number of Persons constraint.

Linear dependence is equivalent to having redundancy with selected constraints or with selected constraints and low population constraints. Near linear dependence is equivalent to having almost redundant constraints.

4. Outlier – If retaining a constraint causes a household weight to go outside the acceptable limit between 1 and 100, it is dropped so that it does not cause outlier weights.

Except for the first case where there is no population, each time that a constraint is dropped, the calibration process does not attempt to make census counts/NHS estimates agree for that constraint in that WA. A constraint dropped frequently will usually have a larger census count/NHS estimate difference than a constraint dropped much less often. This will be apparent if Table 7.1.1 and Table 7.3.1 are compared.

Table 7.1.1 lists all the WA-level constraints as well as the number of times that the constraint was dropped for each reason. The situations above are listed respectively as the columns No population, Small sample, Linearly dependant, Nearly linearly dependant and Outlier.

Table 7.1.1 Frequency of discarding WA-level constraints in 2011 in final weight adjustment

Variables/ constraints	No population	Small sample	Linearly dependant	Nearly linearly dependant	Outlier
TOTPERS	0	0	0	0	0
TPERGE15	0	3	3,878	1,970	0
MALE	0	4	1	3,114	0
MALEGE15	0	5	0	373	0
AGE4	1	592	390	427	0
AGE9	1	804	301	325	0
AGE14	1	707	356	299	0
AGE19	3	459	278	179	0
AGE24	2	299	88	168	0
AGE29	2	387	88	246	0
AGE34	2	279	196	322	0
AGE39	2	210	326	341	0
AGE44	1	145	358	290	0
AGE49	1	70	400	237	0
AGE54	0	66	119	62	0
AGE59	1	97	10	13	0
AGE64	3	253	5	17	0
AGE74	0	186	187	212	0
AGE75PL	1	1,070	209	190	0
SINGLE	0	4	38	5,804	0
MARRIED	0	5	1	5,611	0
SEPARATED	1	4,192	0	0	0
DIVORCED	1	213	0	1	0
WIDOWED	2	1,392	0	11	0
COMLAW	0	341	0	164	0
COUPLE	0	5	2	5,844	0
CHILD	0	14	2	5,835	0
LONEPAR	1	1,242	0	25	0
INEFAM	0	4	3,030	2,817	0
NOINEFAM	1	44	0	2,778	1
NOTINFAM	0	9	0	306	2
EN_EN_BI	32	2,044	0	12	0

Table 7.1.1 Frequency of discarding WA-level constraints in 2011 in final weight adjustment (continued)

Variables/ constraints	No population	Small sample	Linearly dependant	Nearly linearly dependant	Outlier
EN_EN_EN	117	824	0	2,425	0
FR_EN_BI	119	4,952	0	0	0
FR_FR_BI	195	3,082	0	148	0
FR_FR_FR	1,722	1,091	0	847	0
LAN_ABOR	2,571	1,317	0	60	0
LAN_AFR	1,562	2,532	0	0	0
LAN_ARAB	808	3,255	0	2	0
LAN_CHIN	610	2,994	0	12	0
LAN_GER	217	4,248	0	29	0
LAN_INDO	1,230	2,653	0	18	0
LAN_IRAN	1,830	2,533	0	0	0
LAN_KOR	1,843	2,312	0	1	0
LAN_PILI	1,769	2,696	0	0	0
LAN_ROM	161	3,298	0	3	0
LAN_SLAV	347	3,793	0	0	0
CHILDFAM	0	30	2	5,813	0
NOCLDFAM	0	15	0	616	0
TOTHHLD	0	0	0	0	0
HHSIZE1	1	214	2	5,202	0
HHSIZE2	0	14	0	1,218	0
HHSIZE3	0	223	0	531	0
HHSIZE4	3	714	0	4,060	0
HHSIZE5	2	4,007	0	569	0
HHSIZEG6	9	5,219	14	467	0
SNGLDET	81	404	2	55	0
APTLT5	274	2,111	0	19	0

Source: Statistics Canada, 2011 National Household Survey.

Certain unofficial languages tend to be found in certain parts of the country and not in others, so many WAs will have little or no population of a given language. As a result, the constraints for unofficial languages will often have census counts and estimates equal to 0 or the constraints will be dropped for having a small number of respondents with that characteristic. Similarly, some constraints involving French or English have little or no population in certain regions of the country. Consequently, constraints such as FR_EN_BI, FR_FR_BI, and EN_EN_EN, were often dropped.

In general, there was no major imbalance with the age constraints being dropped, with perhaps the exception of the 75 and over constraint. All age groups were regularly controlled upon, meaning that the census count/NHS estimate differences were not too extreme.

There were a few other constraints that were dropped frequently. The number of people aged 15 and over was almost always dropped for being linearly dependent or nearly linearly dependent. This is because the number of persons aged 15 and over can be determined by the age constraints, AGE19, AGE24, AGE29 up to AGE75PL. Finally, the constraints representing people in economic families or having children were frequently dropped because the constraints representing the number of people not in an economic family or not in a family with children caused them to be nearly linearly dependent.

The actual differences between the census counts and the estimates will be examined in Section 7.3.

Table 7.1.2 shows the number of times that each reason for dropping or removing a WA-level constraint occurred. The total number of constraints dropped is the sum of the Small sample, the Linearly dependant, the Nearly linearly dependant and the Outlier categories. As mentioned earlier, the No population category is not included in the total because it does not actually represent dropped constraints. The average number of constraints dropped per WA is simply the total for the category divided by 5,884, the number of WAs.

Table 7.1.2 Summary statistics on discarding WA-level constraints in 2011 in the final weight adjustment

Constraints	No population	Small sample	Linearly dependant	Nearly linearly dependant	Outlier	Total
Total constraints dropped (WA level)	15,530	69,676	10,283	60,088	3	140,050
Average number of constraints dropped per WA	2.6	11.8	10.2	1.7	0.0	23.8

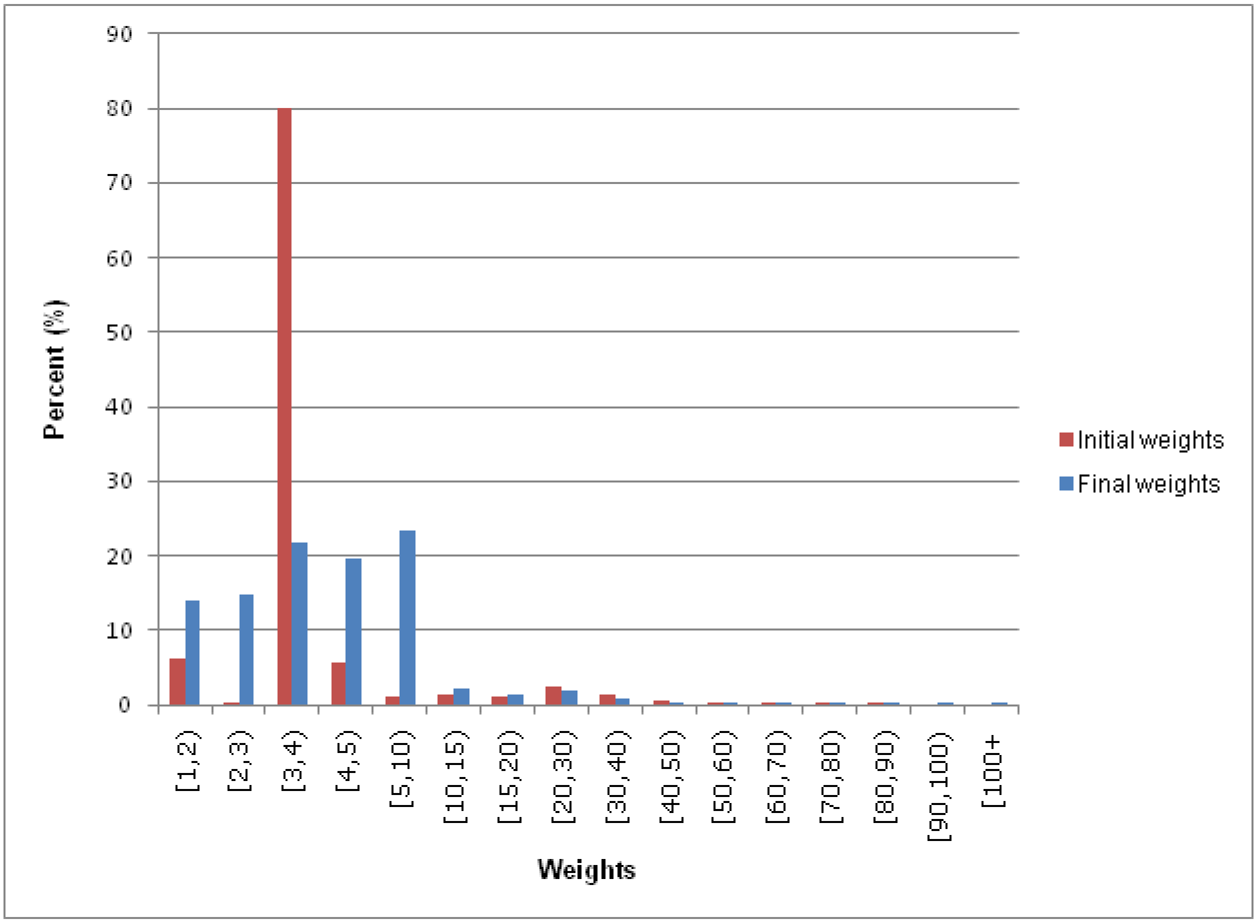
Source: Statistics Canada, 2011 National Household Survey.

7.2 Distribution of weights

Figure 7.2.1 shows the distribution of weights prior to calibration (initial weights) and the weights after calibration (final weights). Weights were grouped into intervals of size 1 for the lower weights, intervals of size 5 for weights between 5 and 20, and intervals of size 10 for the less frequent weights higher than 20. The figure shows the percentage of time that weights within each range occur. It can be seen that initial weights, or pre-calibration weights, were between 3 and 4 for approximately 80% of the households. This is due to the NHS design in which approximately 1 in 3 households received an NHS questionnaire in most areas. Initial weights between 1 and 2 occurred about 6% of the time, as did the range 4 to 5. The remaining 8% of the initial weights were distributed in categories between 2 and 3 and greater than 5. The mean initial weight was 5.0 and the median initial weight was 3.6.

The effect that calibration adjustments had on the weights can also be seen in Figure 7.2.1. A very noticeable difference is that the percentage of households that had initial weights between 3 and 4 was significantly reduced from 80% to approximately 22%. Furthermore, the final weights were more evenly distributed within the categories between 1 and 10. The percentage of households with initial weights greater than 10 did not change significantly however. The mean for final weights was again 5.0, but the median final weight increased to 4.0.

Figure 7.2.1 Distribution of initial weights and final weights



Source: Statistics Canada, 2011 National Household Survey.

The changes between the initial weights and the final weights can be observed in Table 7.2.1. This table shows where the changes between categories occurred. For example, it can be observed that the most stable range was 1 to 2, where 93.4% of the households with an initial weight between 1 and 2 stayed in that category after calibration. The second most stable category was 5 to 10 where 53.7% of households with initial weights between 5 and 10 stayed in that category. This stability is likely attributed to the fact that this category is wider than all the ones below it.

Approximately 55% of the households with initial weights between 2 and 3 had their weight dropped between 1 and 2. However, this was not the case for the households that were initially between 3 and 4; 26.6% of them went into a lower weight group, 47.9% of them went into a higher group, and only 25.5% of them retained a weight between 3 and 4. Some households with very high initial weights had their weights reduced. Categories with weights greater than 20 saw reductions in the number of households after calibration.

Table 7.2.1 Distribution of initial weights and final weights

Initial weights	Final Weights									
	[1,2)	[2,3)	[3,4)	[4,5)	[5,10)	[10,15)	[15,20)	[20,30)	[30+	Total
[1,2)	**153,152**	8,701	1,609	372	177	5	0	0	0	**164,016**
[2,3)	1,217	**747**	164	44	34	7	0	1	0	**2,214**
[3,4)	205,266	361,855	**542,330**	477,016	529,771	11,819	540	30	1	**2,128,628**
[4,5)	10,001	17,565	31,677	**34,953**	54,061	1,848	77	9	0	**150,191**
[5,10)	1,460	1,486	2,647	3,929	**16,777**	4,113	681	142	6	**31,241**
[10,15)	954	474	692	979	10,416	**13,052**	6,028	1,951	121	**34,667**
[15,20)	606	254	354	520	5,168	9,960	**8,885**	5,842	583	**32,171**
[20,30)	968	314	425	492	4,975	10,598	15,317	**22,083**	5,819	**60,991**
[30+	762	161	203	280	2,188	4,346	7,110	18,454	**19,838**	**53,342**
Total	**374,386**	**391,557**	**580,101**	**518,585**	**623,567**	**55,748**	**38,638**	**48,512**	**26,367**	**2,657,461**

Note:
The “[“ symbol means the number is included in the interval and the “)” symbol means it is not included in the interval.

Source: Statistics Canada, 2011 National Household Survey.

7.3 Discrepancies between census counts and NHS estimates – Canada

Chapter 4 describes the methods used to calculate the final household weights and Section 7.2 shows some of the relationships between the initial and final weights. Calibration reduced or eliminated discrepancies between the census counts and the corresponding NHS estimates for the 58 constraints at the WA level (see Appendix C). Some discrepancies remain, however, since constraints are sometimes discarded (see Section 7.1). The census count/NHS estimate discrepancy is defined as

$$\text{Census count/NHS estimate discrepancy} = \frac{\text{NHS estimate} - \text{Census count}}{\text{Census count}}$$

The numerator in the above expression (NHS estimate – census count) is referred to as the 'census count/NHS estimate difference'. By dividing this value by the census count, the census count/NHS estimate difference 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 7.3.1 shows the 2011 Canada-level census count/NHS estimate differences for the 58 WA-level constraints for both the initial weights and the final weights. The characteristic FEMALE is not one of the 58 WA-level constraints listed in Appendix C because it can be calculated by the difference of total persons and males. However, it has been included in this table for interest purposes. It can be seen that the sum of the final weight difference for males and females equals 0.It should also be mentioned that the census count for the total persons characteristic (TOTPERS) is less than the published figure of the 2011 Census (33,476,688). There are two reasons for this difference. First, the weighting process only utilized private households. Collective dwellings were not in-scope for the NHS and were excluded from the NHS. However, they were used for the census counts. Second, the cases mentioned in Section 4.2 (the five census subdivisions (CSDs) corresponding to five Indian reserves that were excluded from the universe because of a very low response rate in the NHS) were not part of the weighting process but were part of the published figure.

It is not enough to simply observe the difference between the NHS estimate and the census count. It is better to consider the difference relative to the size of the census count. Therefore, Table 7.3.1 shows discrepancies based on initial weights and final weights. Most cases in Table 7.3.1 had a discrepancy between -1.00% and 1.00%.

The census count/NHS estimate difference for initial weights tended to be greater than the census count/NHS estimate difference for the final weights. This shows the importance of the calibration process. As mentioned in Section 7.1, a difference between the census count and NHS estimate could occur in a WA for a characteristic if its constraint is dropped during calibration. In other words, the process did not control on any dropped constraint for a given WA. If the constraint is dropped in many WAs, these differences could partially cancel, or they could add up to create a large difference at the Canada level. Total persons (TOTPERS) and total households (TOTHHLDS) were the only mandatory constraints for which census counts/NHS estimates agreement had to be guaranteed for all WAs, so the final weight difference and discrepancy for these were 0. However, all other constraints had to be dropped for some WAs.

Section 7.1 pointed out the constraints that were dropped frequently and where high differences or discrepancies might lie. The effect of dropping those constraints can be seen in Table 7.3.1. The person constraints Total persons aged 15 and over, Single, Married, Number of children, Couple, being in an Economic family, and being in a family with a child were almost always dropped. Of these, the Married and Couple constraints had quite large differences (55,107 and 54,021 respectively). However, because the census counts were so high, the discrepancies for these constraints were not too high (0.43% and 0.34% respectively). On the other hand, the number of separated people had a discrepancy of -3.35%. This relatively small constraint was dropped 4,192 times and created an underestimate of -27,154.

The largest differences and discrepancies were typically found in some of the language variables. Since many languages are not found in high numbers in many parts of the country, these variables were often dropped during calibration, which led to some differences in the census counts and NHS estimates. Furthermore, because many of them have relatively low census counts, the differences are magnified, resulting in greater discrepancies. The largest difference was in the EN_EN_EN constraint where people reported speaking English as their mother tongue, home language, and official language. The overestimate of 166,801 was by far the largest difference for any variable. However, because this was the most common language scenario, the discrepancy was 0.99%, meaning that the characteristic was overestimated by about 1%. The largest discrepancy belonged to the FR_EN_BI characteristic in which people reported having French as the mother tongue, English as the home language, and able to speak both English and French. This constraint was frequently dropped for having at least one respondent in the WA, but less than 30 respondents. As a result, this characteristic, which had a census count around 351,000, had an overestimate of nearly 41,000 and a resulting discrepancy of 11.68%. Other language constraints with high discrepancies were African languages (8.49%) and Filipino languages (5.53%).

Most other person-level characteristics had lower discrepancies, particularly the age and gender characteristics which all had estimates within 0.2% of their census count. Only the group aged 75 and over with a discrepancy of 0.28% exceed 0.2%. This was also the age group with the greatest census count/NHS estimate difference (-5,365).

Household variables had mixed results. Households of size 1, 2, and 6 or more were all underestimated. The households with at least 6 people had the greatest underestimate (-14,493), which resulted in the large discrepancy of -3.69%. On the other hand, the household of size 4 had the largest difference (24,124) and a resulting discrepancy of 1.27%.

The closest estimates for a non-mandatory constraint belonged to the Age 5 to 9 group and the Korean language group with differences of 69 and 72 respectively. The small Korean language group had a discrepancy of 0.05%. The smallest discrepancies belonged to the Single detached dwellings (0.00%), Age 5 to 9 group (0.00%), the Age 50 to 54 group (0.01%), and the economic family constraint (-0.02%). Note that these discrepancies have been rounded to the nearest 0.01.

Table 7.3.1 2011 Census count and NHS estimate differences and discrepancies (%) for Canada

	Census	Initial weights, NHS		Final weights, NHS		
Characteristic	Counts	Estimates	Difference	Estimates	Difference	Discrepancy (%)
TOTPERS	32,852,323	32,570,247	-282,076	32,852,323	0	0.00
TPERGE15	27,266,716	26,792,223	-474,493	27,259,524	-7,192	-0.03
MALE	16,151,561	15,943,350	-208,211	16,163,113	11,552	0.07
FEMALE*	16,700,762	16,626,898	-73,864	16,689,210	-11,552	-0.07
MALEGE15	13,291,793	12,983,638	-308,155	13,295,352	3,559	0.03
AGE4	1,869,871	1,938,216	68,345	1,873,691	3,820	0.20
AGE9	1,803,353	1,875,881	72,528	1,803,422	69	0.00
AGE14	1,912,383	1,963,928	51,545	1,915,687	3,304	0.17
AGE19	2,161,432	2,109,902	-51,530	2,160,275	-1,157	-0.05
AGE24	2,167,042	2,017,460	-149,582	2,163,793	-3,249	-0.15
AGE29	2,150,627	2,054,640	-95,987	2,149,302	-1,325	-0.06
AGE34	2,145,318	2,115,595	-29,723	2,144,647	-671	-0.03
AGE39	2,156,499	2,149,240	-7,259	2,155,348	-1,151	-0.05
AGE44	2,305,179	2,272,305	-32,874	2,306,085	906	0.04
AGE49	2,651,229	2,585,140	-66,089	2,655,581	4,352	0.16
AGE54	2,633,806	2,567,868	-65,938	2,633,979	173	0.01
AGE59	2,316,094	2,267,609	-48,485	2,314,650	-1,444	-0.06
AGE64	2,027,821	2,024,784	-3,037	2,024,329	-3,492	-0.17
AGE74	2,619,303	2,671,238	51,935	2,624,535	5,232	0.20
AGE75PL	1,932,366	1,956,441	24,075	1,927,001	-5,365	-0.28
SINGLE	15,460,197	15,203,908	-256,289	15,449,727	-10,470	-0.07
MARRIED	12,823,743	12,994,368	170,625	12,878,850	55,107	0.43
SEPARATED	810,778	761,604	-49,174	783,624	-27,154	-3.35
DIVORCED	2,323,491	2,193,851	-129,640	2,319,013	-4,478	-0.19
WIDOWED	1,434,114	1,416,516	-17,598	1,421,110	-13,004	-0.91
COMLAW	3,135,282	3,002,520	-132,762	3,126,473	-8,809	-0.28
COUPLE	15,722,178	15,786,489	64,311	15,776,199	54,021	0.34
CHILD	9,969,150	9,972,358	3,208	9,950,695	-18,455	-0.19
LONEPAR	1,527,427	1,487,512	-39,915	1,499,376	-28,051	-1.84
INEFAM	27,954,234	27,893,392	-60,842	27,948,792	-5,442	-0.02
NOINEFAM	4,898,089	4,676,855	-221,234	4,903,531	5,442	0.11
NOTINFAM	5,633,568	5,323,889	-309,679	5,626,052	-7,516	-0.13
EN_EN_BI	1,613,378	1,584,730	-28,648	1,610,801	-2,577	-0.16
EN_EN_EN	16,806,378	16,990,679	184,301	16,973,179	166,801	0.99
FR_EN_BI	351,034	388,803	37,769	392,021	40,987	11.68
FR_FR_BI	2,646,831	2,731,130	84,299	2,666,821	19,990	0.76
FR_FR_FR	3,773,941	3,731,106	-42,835	3,806,924	32,983	0.87
LAN_ABOR	196,928	186,974	-9,954	191,850	-5,078	-2.58
LAN_AFR	76,951	77,959	1,008	83,485	6,534	8.49
LAN_ARAB	327,059	325,422	-1,637	323,124	-3,935	-1.20
LAN_CHIN	1,067,628	1,048,989	-18,639	1,066,952	-676	-0.06
LAN_GER	542,498	511,944	-30,554	536,789	-5,709	-1.05
LAN_INDO	877,645	896,995	19,350	875,699	-1,946	-0.22
LAN_IRAN	191,909	189,444	-2,465	192,127	218	0.11
LAN_KOR	137,141	131,245	-5,896	137,213	72	0.05

Table 7.3.1 2011 Census count and NHS estimate differences and discrepancies (%) for Canada (continued)

	Census	Initial weights, NHS		Final weights, NHS		
Characteristic	**Counts**	**Estimates**	**Difference**	**Estimates**	**Difference**	**Discrepancy (%)**
LAN_PILI	111,744	106,386	-5,358	117,923	6,179	5.53
LAN_ROM	1,113,783	1,112,691	-1,092	1,096,133	-17,650	-1.58
LAN_SLAV	683,708	665,284	-18,424	668,188	-15,520	-2.27
CHILDFAM	5,636,164	5,612,297	-23,867	5,633,721	-2,443	-0.04
NOCLDFAM	3,752,352	3,768,460	16,108	3,753,755	1,403	0.04
TOTHHLD	13,319,251	13,283,671	-35,580	13,319,251	0	0.00
HHSIZE1	3,673,018	3,696,360	23,342	3,662,801	-10,217	-0.28
HHSIZE2	4,544,542	4,557,124	12,582	4,536,175	-8,367	-0.18
HHSIZE3	2,081,661	2,057,036	-24,625	2,086,205	4,544	0.22
HHSIZE4	1,903,101	1,885,895	-17,206	1,927,225	24,124	1.27
HHSIZE5	724,278	713,786	-10,492	728,688	4,410	0.61
HHSIZEG6	392,651	373,470	-19,181	378,158	-14,493	-3.69
SNGLDET	7,327,904	7,341,850	13,946	7,327,785	-119	0.00
APTLT5	2,431,005	2,392,925	-38,080	2,423,005	-8,000	-0.33

* The characteristic FEMALE is not one of the 58 WA-level constraints listed in Appendix C.

Sources: Statistics Canada, 2011 Census and 2011 National Household Survey.

8. Variance

The error in an estimate is the difference between the estimate and the actual value of what is being estimated. Sampling error and household non-response error are two sources of error in the NHS. Sampling error stems from the fact that the estimates are based on observations from a sample and not from the entire population. Household non-response error occurs when households selected in the sample do not respond to the survey. Estimation methods also have an impact on the error. Some methods are more precise than others in measuring a particular characteristic of the population.

The error has two components: a random component, the variance; and a systematic component, the bias. The variance measures how much the estimate varies from the average that would result from hypothetical repetitions of the survey process. The variance can be estimated using data from respondents in the sample. The 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. Non-response bias is covered in Chapter 5. The sampling and estimation methods used in the NHS generate negligible bias. This chapter deals with sampling variance and variance due to household non-response.

To estimate the variance, the first step is to derive estimators using mathematical formulas. Those formulas have to take into account the various sampling and weighting steps described in previous chapters. The formulas specify how data from respondents are used to produce variance estimates. To perform the derivations, certain assumptions and simplifications had to be made. First, it was assumed that the sample and subsample selected by systematic stratified sampling were selected by stratified simple random sampling without replacement. This assumption is necessary when systematic sampling is performed. The assumption is reasonable because the strata are small and their population can therefore be treated as homogeneous. The second assumption concerns the mechanism that generates household non-response. It was assumed to be equivalent to a Poisson sampling process in which a household's probability of response/selection is equal to the ratio of its weight assigned before non-response adjustment weight transfer to the weight assigned to it after the transfer. For example, a respondent household whose weight doubled following the transfer would have an estimated response probability of 50%. It was also assumed that this probability was known, not estimated. Making these assumptions about the non-response mechanisms means that the non-response bias in the variance estimate is ignored. The last simplification is that surprise respondents were excluded from the variance estimation process. This makes the process much easier and has little impact on the estimate since surprise respondents make up a small fraction of all respondents. The technical details of the variance estimation process are provided in Verret (2013).

Since the variance estimators based on the formulas described above tend to underestimate the actual variance, the final step was to correct them using simulations. To that end, an artificial population was first generated using sample data. Multiple samples were then selected independently (i.e., replicates), and the weighting methods were applied to each one. The calculated variance estimators were compared with the actual variability from replicate to replicate. This provides an upward adjustment that is applied to all the variance estimates calculated with the formulas.

The estimated variance can be used to construct several types of measures of an estimate's variability. For example, it can be used to calculate standard errors or coefficients of variation (CVs). The standard error is equal to the square root of the variance. The CV is the ratio of the estimate's standard error to the estimate itself. The smaller these measures of variability are, the more precise the estimate. The CV is a particularly interesting measure of variability in that it does not depend on the unit of measure of the characteristic being studied and can be expressed as a percentage. It is important to carefully distinguish between these measures of variability and other measures of quality that are not, strictly speaking, measures of variability. Examples of such measures are the bias indicators, response rate and global non-response rate of the NHS. 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 due to household non-response and item non-response. Table 8.1 shows the CVs of various estimates at the Canada, provincial and territorial levels. The contents of this table can be consulted or downloaded at the following address: http://www12.statcan.gc.ca/nhs-enm/2011/ref/cv/index.cfm?Lang=E.

The sampling and estimation methods used in the 2011 NHS are different from the ones used for the 2006 Census long questionnaire. The magnitude and effect of non-response are very different in 2011. All that has an impact on the variability of the estimates. A methodological note comparing the CVs of the 2011 estimates with the CVs of the 2006 estimates for some variables is available at http://www12.statcan.gc.ca/nhs-enm/2011/ref/cv/cvnote.cfm?Lang=E.

Table 8.1 Coefficients of variation (%) of the 2011 National Household Survey – Canada, provinces and territories

Characteristic	Canada	Newfoundland and Labrador	Prince Edward Island	Nova Scotia	New Brunswick	Quebec	Ontario
Population characteristic							
Immigration and citizenship							
Total population in private households by citizenship	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Canadian citizens	0.01	0.06	0.18	0.06	0.06	0.02	0.03
Not Canadian citizens	0.23	6.06	5.87	2.40	3.01	0.50	0.35
Total population in private households by immigrant status	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Non-immigrants	0.02	0.07	0.24	0.09	0.08	0.03	0.04
Immigrants	0.08	3.55	3.88	1.41	1.87	0.21	0.10
Non-permanent residents	0.61	9.99	16.29	4.97	5.98	1.28	1.02
Total immigrant population in private households by place of birth	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Americas	0.29	6.74	7.96	2.94	3.11	0.60	0.41
Europe	0.17	4.79	6.15	2.08	2.98	0.46	0.22
Africa	0.50	16.08	41.14	7.51	9.04	0.72	0.83
Asia	0.13	7.95	6.87	3.07	4.43	0.50	0.17
Oceania and other	1.32	28.55	55.64	12.39	15.27	6.86	2.51
Visible minority							
Total population in private households by visible minority	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Visible minority population	0.10	5.27	6.21	1.76	2.90	0.30	0.14
Not a visible minority	0.02	0.07	0.20	0.10	0.07	0.04	0.05
Aboriginal peoples							
Total population in private households by Aboriginal identity	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Aboriginal identity	0.26	1.61	8.37	1.71	2.10	0.72	0.60
Non-Aboriginal identity	0.01	0.12	0.14	0.07	0.07	0.01	0.02
Mobility							
Total – Mobility status 5 years ago	0.00	0.03	0.04	0.02	0.02	0.01	0.00
Non-movers	0.05	0.35	0.72	0.27	0.29	0.09	0.08
Movers	0.08	0.86	1.52	0.54	0.60	0.16	0.14
Education							
Total population aged 15 years and over by highest certificate, diploma or degree	0.01	0.05	0.10	0.04	0.03	0.01	0.01
No certificate, diploma or degree	0.10	0.72	1.61	0.60	0.61	0.19	0.17
High school diploma or equivalent	0.09	0.86	1.44	0.59	0.59	0.20	0.14
Postsecondary certificate, diploma or degree	0.04	0.45	0.80	0.29	0.35	0.08	0.07
University certificate, diploma or degree at bachelor level or above	0.09	1.11	1.79	0.62	0.77	0.19	0.14
Labour							
Total population aged 15 years and over by language used most often at work	0.03	0.29	0.41	0.20	0.20	0.06	0.05
English, single response	0.04	0.30	0.44	0.20	0.27	0.30	0.05
French, single response	0.08	8.67	7.41	2.63	0.66	0.08	0.75
Non-official languages, single response	0.51	6.10	15.65	7.38	7.02	1.20	0.83

Table 8.1 Coefficients of variation (%) of the 2011 National Household Survey – Canada, provinces and territories (continued)

Characteristic	Manitoba	Saskat-chewan	Alberta	British Columbia	Yukon	Northwest Territories	Nunavut
Population characteristic							
Immigration and citizenship							
Total population in private households by citizenship	0.00	0.00	0.00	0.00	0.00	0.02	0.00
Canadian citizens	0.08	0.07	0.05	0.05	0.34	0.16	0.02
Not Canadian citizens	1.20	1.84	0.68	0.52	6.26	5.91	3.01
Total population in private households by immigrant status	0.00	0.00	0.00	0.00	0.00	0.02	0.00
Non-immigrants	0.11	0.09	0.07	0.07	0.50	0.22	0.04
Immigrants	0.56	1.19	0.29	0.18	3.95	2.94	1.74
Non-permanent residents	4.10	4.76	1.58	1.37	21.79	12.52	6.22
Total immigrant population in private households by place of birth	0.00	0.00	0.00	0.00	0.00	0.02	0.00
Americas	1.81	3.54	1.08	0.88	9.58	8.62	3.71
Europe	1.20	2.19	0.64	0.44	5.92	5.72	2.77
Africa	3.61	5.97	1.65	1.73	28.75	13.57	5.28
Asia	0.84	1.87	0.46	0.23	9.38	5.51	3.22
Oceania and other	10.08	11.08	3.56	1.83	22.02	24.68	6.46
Visible minority							
Total population in private households by visible minority	0.00	0.00	0.00	0.00	0.00	0.02	0.00
Visible minority population	0.69	1.39	0.33	0.19	6.62	3.54	2.19
Not a visible minority	0.10	0.09	0.07	0.07	0.43	0.25	0.04
Aboriginal peoples							
Total population in private households by Aboriginal identity	0.00	0.00	0.00	0.00	0.00	0.02	0.00
Aboriginal identity	0.62	0.71	0.72	0.64	2.72	0.67	0.11
Non-Aboriginal identity	0.12	0.13	0.05	0.04	0.82	0.72	0.67
Mobility							
Total – Mobility status 5 years ago	0.02	0.02	0.01	0.01	0.00	0.06	0.04
Non-movers	0.26	0.29	0.17	0.15	1.56	0.89	0.47
Movers	0.42	0.47	0.21	0.20	1.89	0.94	0.62
Education							
Total population aged 15 years and over by highest certificate, diploma or degree	0.03	0.04	0.02	0.02	0.00	0.09	0.06
No certificate, diploma or degree	0.46	0.55	0.33	0.30	2.91	0.86	0.32
High school diploma or equivalent	0.44	0.52	0.27	0.23	2.71	1.49	0.93
Postsecondary certificate, diploma or degree	0.27	0.32	0.14	0.12	1.23	0.70	0.50
University certificate, diploma or degree at bachelor level or above	0.54	0.70	0.29	0.24	2.75	1.73	0.99
Labour							
Total population aged 15 years and over by language used most often at work	0.14	0.16	0.08	0.08	0.51	0.34	0.25
English, single response	0.15	0.17	0.08	0.09	0.56	0.36	0.35
French, single response	3.04	6.39	3.39	3.25	17.74	13.75	4.66
Non-official languages, single response	2.78	4.28	2.14	0.94	21.23	3.78	0.83

Table 8.1 Coefficients of variation (%) of the 2011 National Household Survey – Canada, provinces and territories (continued)

Characteristic	Canada	Newfound-land and Labrador	Prince Edward Island	Nova Scotia	New Brunswick	Quebec	Ontario
Total population aged 15 years and over by labour force status	0.01	0.05	0.10	0.04	0.03	0.01	0.01
In the labour force	0.03	0.33	0.47	0.22	0.23	0.06	0.05
Employed	0.04	0.41	0.62	0.26	0.27	0.07	0.06
Unemployed	0.22	1.47	2.87	1.25	1.32	0.45	0.34
Not in the labour force	0.06	0.48	1.03	0.37	0.39	0.12	0.10
Housing							
Total number of occupied private dwellings by condition of dwelling	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Only regular maintenance or minor repairs needed	0.02	0.18	0.33	0.15	0.16	0.04	0.03
Major repairs needed	0.25	2.09	4.07	1.37	1.50	0.49	0.45
Total number of occupied private dwellings by number of rooms	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Average number of rooms per dwelling	0.03	0.24	0.45	0.16	0.18	0.04	0.04
Total number of private households by tenure	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Owner	0.03	0.26	0.47	0.19	0.19	0.07	0.05
Renter	0.07	0.90	1.31	0.47	0.60	0.11	0.13
Band housing	0.12	1.07	3.63	0.30	0.65	0.55	0.56
Shelter costs							
Total number of owner and tenant households with household total income greater than zero, in non-farm, non-reserve private dwellings by shelter-cost-to-income ratio	0.01	0.02	0.12	0.04	0.04	0.01	0.01
Spending less than 30% of household total income on shelter costs	0.04	0.29	0.61	0.23	0.23	0.07	0.07
Spending 30% or more of household total income on shelter costs	0.12	1.32	2.46	0.79	0.99	0.24	0.19
Spending 30% to less than 100% of household total income on shelter costs	0.14	1.50	2.78	0.93	1.15	0.28	0.22
Income of individuals							
Total income in 2010 of population aged 15 years and over	0.01	0.05	0.10	0.04	0.03	0.01	0.01
Average income ($)	0.09	0.50	0.98	0.44	0.41	0.15	0.14
After-tax income in 2010 of population aged 15 years and over	0.01	0.05	0.10	0.04	0.03	0.01	0.01
Average after-tax income ($)	0.07	0.42	0.84	0.33	0.34	0.13	0.11
Income of families							
Family income in 2010 of economic families	0.01	0.09	0.20	0.07	0.07	0.02	0.02
Average family income ($)	0.10	0.55	1.11	0.51	0.46	0.18	0.15
Average after-tax family income ($)	0.08	0.46	0.95	0.38	0.38	0.15	0.13
Average family size	0.01	0.09	0.20	0.07	0.07	0.02	0.02

Table 8.1 Coefficients of variation (%) of the 2011 National Household Survey – Canada, provinces and territories (continued)

Characteristic	Manitoba	Saskat-chewan	Alberta	British Columbia	Yukon	Northwest Territories	Nunavut
Total population aged 15 years and over by labour force status	0.03	0.04	0.02	0.02	0.00	0.09	0.06
In the labour force	0.16	0.19	0.09	0.09	0.74	0.37	0.28
Employed	0.18	0.21	0.10	0.10	0.91	0.46	0.33
Unemployed	1.27	1.48	0.74	0.59	4.80	1.88	1.14
Not in the labour force	0.33	0.42	0.24	0.16	2.50	1.13	0.50
Housing							
Total number of occupied private dwellings by condition of dwelling	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Only regular maintenance or minor repairs needed	0.13	0.15	0.06	0.05	0.84	0.35	0.34
Major repairs needed	1.05	1.24	0.83	0.70	4.34	1.68	0.90
Total number of occupied private dwellings by number of rooms	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Average number of rooms per dwelling	0.13	0.17	0.08	0.07	0.91	0.39	0.19
Total number of private households by tenure	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Owner	0.15	0.19	0.11	0.10	1.05	0.66	0.92
Renter	0.38	0.57	0.30	0.24	2.37	0.72	0.24
Band housing	0.16	0.21	0.16	0.65	2.49	3.78	…
Shelter costs							
Total number of owner and tenant households with household total income greater than zero, in non-farm, non-reserve private dwellings by shelter-cost-to-income ratio	0.06	0.11	0.04	0.02	0.16	0.06	0.01
Spending less than 30% of household total income on shelter costs	0.20	0.25	0.13	0.13	1.00	0.51	0.14
Spending 30% or more of household total income on shelter costs	0.79	0.89	0.40	0.29	4.02	3.05	2.07
Spending 30% to less than 100% of household total income on shelter costs	0.91	1.01	0.47	0.35	4.59	3.44	2.33
Income of individuals							
Total income in 2010 of population aged 15 years and over	0.03	0.04	0.02	0.02	0.00	0.09	0.06
Average income ($)	0.33	0.46	0.33	0.23	1.48	0.65	0.38
After-tax income in 2010 of population aged 15 years and over	0.03	0.04	0.02	0.02	0.00	0.09	0.06
Average after-tax income ($)	0.27	0.39	0.28	0.19	1.30	0.58	0.35
Income of families							
Family income in 2010 of economic families	0.06	0.08	0.04	0.04	0.30	0.21	0.10
Average family income ($)	0.38	0.52	0.35	0.27	1.80	0.70	0.43
Average after-tax family income ($)	0.31	0.45	0.30	0.22	1.58	0.62	0.40
Average family size	0.06	0.08	0.04	0.04	0.35	0.21	0.10

Table 8.1 Coefficients of variation (%) of the 2011 National Household Survey – Canada, provinces and territories (continued)

Characteristic	Canada	Newfound-land and Labrador	Prince Edward Island	Nova Scotia	New Brunswick	Quebec	Ontario
Income of households							
Household income in 2010 of private households	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Average household total income ($)	0.09	0.51	1.01	0.45	0.42	0.15	0.14
Average after-tax household income ($)	0.07	0.43	0.87	0.34	0.35	0.13	0.12

Characteristic	Manitoba	Saskat-chewan	Alberta	British Columbia	Yukon	Northwest Territories	Nunavut
Income of households							
Household income in 2010 of private households	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Average household total income ($)	0.34	0.48	0.33	0.24	1.57	0.66	0.39
Average after-tax household income ($)	0.28	0.40	0.29	0.20	1.38	0.59	0.36

… not applicable

Source: Statistics Canada, 2011 National Household Survey.

9. Conclusion

The introduction of the 2011 National Household Survey (NHS) to replace the census long-form questionnaire resulted in many challenges. As a result of the non-mandatory nature of the NHS, the non-response rates were higher than for previous long-form censuses. The types and patterns of non-response were also different. A more complex approach was thus required for the sampling design, non-response follow-up, weighting and estimation procedures for the NHS.

Changes included an increase in the sample size from one in five to one in three private households. An improved strategy for non-response follow-up was implemented to ensure that certain areas and populations identified to be at risk for higher than usual non-response rates were targeted. Although there are several limitations, bias indicators due to total non-response were produced for a few characteristics. Finally, the complex sample design required for the NHS led to the development of innovative new weighting and variance estimation methodology that are quite different from previous censuses.

Appendix A Glossary

The definitions of census terms, variables and concepts are presented here as they appear in the *2011 Census Dictionary* (Catalogue no. 98-301-X2011001). Users should refer to the *2011 Census Dictionary* for full definitions and additional remarks related to any concepts, such as information on direct and derived variables and their respective universe.

Census division (CD): Census division (CD) is the general term for provincially legislated areas (such as county, *municipalité régionale de comté* and regional district) or their equivalents. Census divisions are intermediate geographic areas between the province/territory level and the municipality (census subdivision).

Census subdivision (CSD): Census subdivision (CSD) is the general term for municipalities (as determined by provincial/territorial legislation) or areas treated as municipal equivalents for statistical purposes (e.g., Indian reserves, Indian settlements and unorganized territories).

Census tract (CT): Census tracts (CTs) are small, relatively stable geographic areas that usually have a population between 2,500 and 8,000 persons. 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): Small geographic units used for the collection of census and NHS data. Collection units cover all the territory of Canada.

Collective dwelling: 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 homes, nursing homes, hospitals, staff residences, communal quarters (military bases), work camps, jails, group homes, and so on. Collective dwellings may be occupied by usual residents or solely by foreign residents and/or by temporarily present persons.

Dissemination area (DA): A dissemination area (DA) is a small, relatively stable geographic unit composed of one or more adjacent dissemination blocks. It is the smallest standard geographic area for which all census data are disseminated. DAs cover all the territory of Canada.

Dwelling: Refers to a set of living quarters in which a person or a group of persons reside or could reside.

Dwelling, private: Refers to a separate set of living quarters with a private entrance either from outside 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 someone else. 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, solar heating panels, etc.).
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).

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 by foreign and/or temporary residents and unoccupied dwellings. Marginal dwellings are classified as occupied by usual residents or by foreign and/or temporary residents. Marginal dwellings that were unoccupied on May 10, 2011, are not counted in the housing stock.

Dwelling, private, occupied by usual residents: 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, 2011. Unless otherwise specified, all data in housing products are for occupied private dwellings, rather than for unoccupied private dwellings or dwellings occupied solely by foreign and/or temporary residents.

Household: Refers to a person or a group of persons (other than foreign residents) who occupy the same dwelling and do not have a usual place of residence elsewhere in Canada. It may consist of a family group (census family) with or without other persons, of two or more families sharing a dwelling, of a group of unrelated persons, or of one person living alone. Household members who are temporarily absent on May 10, 2011 (e.g., temporarily residing elsewhere) are considered as part of their usual household. Every person is a member of one and only one household. Unless otherwise specified, all data in household reports are for private households only.

Households are classified into three groups: private households, collective households and households outside Canada.

Household, private: Refers to a person or a group of persons (other than foreign residents) who occupy a private dwelling and do not have a usual place of residence elsewhere in Canada.

Appendix B The history of sampling in the Canadian censuses

Sampling was first used in the Canadian census in 1941. A housing schedule was completed for every tenth 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, thus allowing 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 tenth enumeration area across Canada was selected and all population schedules in these areas were processed in advance.

Again in 1951, the census of housing was conducted on a sample basis. 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 over 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) that indicated that the effect of the enumerator was a major contribution to the variance of census figures in a canvasser census. Thus the use of self-enumeration was expected to reduce the variance of census figures through reducing the effect of the enumerator, while at the same time giving the respondent more time and privacy in which to answer the census questions—factors which might also 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, while 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 content of the 1976 Census was considerably less than that of 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, it was 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, response burden, and improved timeliness (see Royce, 1983). The one-in-five sampling rate had been maintained for every census from 1981 to 2006.

Information previously collected by the mandatory long-form census questionnaire was collected as part of the National Household Survey in 2011. With this change, every household was required to answer the ten questions that were contained in the census questionnaire.

Appendix C Constraints used in the weighting process

The following is a list of the 60 constraints used to generate household weights for the 2011 NHS. The first 58 are at the WA level only.

No.	Constraint	Level	Description
1	TOTPERS	WA (Person)	Total number of persons
2	TPERGE15	WA (Person)	Total number of persons aged 15 and over
3	MALE	WA (Person)	Males
4	MALEGE15	WA (Person)	Males aged 15 and over
5	AGE4	WA (Person)	Persons aged 0 to 4
6	AGE9	WA (Person)	Persons aged 5 to 9
7	AGE14	WA (Person)	Persons aged 10 to 14
8	AGE19	WA (Person)	Persons aged 15 to 19
9	AGE24	WA (Person)	Persons aged 20 to 24
10	AGE29	WA (Person)	Persons aged 25 to 29
11	AGE34	WA (Person)	Persons aged 30 to 34
12	AGE39	WA (Person)	Persons aged 35 to 39
13	AGE44	WA (Person)	Persons aged 40 to 44
14	AGE49	WA (Person)	Persons aged 45 to 49
15	AGE54	WA (Person)	Persons aged 50 to 54
16	AGE59	WA (Person)	Persons aged 55 to 59
17	AGE64	WA (Person)	Persons aged 60 to 64
18	AGE74	WA (Person)	Persons aged 65 to 74
19	AGE75PL	WA (Person)	Persons aged 75 and over
20	SINGLE	WA (Person)	Single (never married)
21	MARRIED	WA (Person)	Married persons
22	SEPARATED	WA (Person)	Separated persons
23	DIVORCED	WA (Person)	Divorced persons
24	WIDOWED	WA (Person)	Widowed persons
25	COMLAW	WA (Person)	Persons in a common law relationship
26	COUPLE	WA (Person)	Persons that are part of a couple (marriage, partnership)
27	CHILD	WA (Person)	Number of children
28	LONEPAR	WA (Person)	Lone parents
29	INEFAM	WA (Person)	Persons in an economic family
30	NOINEFAM	WA (Person)	Persons in household not part of economic family
31	NOTINFAM	WA (Person)	Persons in household not part of census family
32	EN_EN_BI	WA (Person)	MT = English, HL = English, OL = Bilingual *
33	EN_EN_EN	WA (Person)	MT = English, HL = English, OL = English *
34	FR_EN_BI	WA (Person)	MT = French, HL = English, OL = Bilingual *
35	FR_FR_BI	WA (Person)	MT = French, HL = French, OL = Bilingual *
36	FR_FR_FR	WA (Person)	MT = French, HL = French, OL = French *
37	LAN_ABOR	WA (Person)	Persons whose mother tongue is an Aboriginal language
38	LAN_AFR	WA (Person)	Persons whose mother tongue is an African language

Appendix C Constraints used in the weighting process (continued)

No.	Constraint	Level	Description
39	LAN_ARAB	WA (Person)	Persons whose mother tongue is Arabic
40	LAN_CHIN	WA (Person)	Persons whose mother tongue is a Chinese language
41	LAN_GER	WA (Person)	Persons whose mother tongue is a Germanic language
42	LAN_INDO	WA (Person)	Persons whose mother tongue is an Indian language
43	LAN_IRAN	WA (Person)	Persons whose mother tongue is an Iranian language
44	LAN_KOR	WA (Person)	Persons whose mother tongue is a Korean language
45	LAN_PILI	WA (Person)	Persons whose mother tongue is Pilipino**
46	LAN_ROM	WA (Person)	Persons whose mother tongue is a Romanic language
47	LAN_SLAV	WA (Person)	Persons whose mother tongue is a Slavic language
48	CHILDFAM	WA (Census Family)	Census families with children
49	NOCLDFAM	WA (Census Family)	Census families without children
50	TOTHHLD	WA (Household)	Total number of households
51	HHSIZE1	WA (Household)	Number of households with 1 person
52	HHSIZE2	WA (Household)	Number of households with 2 people
53	HHSIZE3	WA (Household)	Number of households with 3 people
54	HHSIZE4	WA (Household)	Number of households with 4 people
55	HHSIZE5	WA (Household)	Number of households with 5 people
56	HHSIZEG6	WA (Household)	Number of households with 6 or more people
57	SNGLDET	WA (Household)	Number of households in single detached dwellings
58	APTLT5	WA (Household)	Number of households in apartments with less than 5 floors
59	HHDACT	DA	Number of households in each dissemination area of the WA
60	PPDACT	DA	Number of persons in each dissemination area of the WA

Notes:
* MT is Mother tongue, HL is Home language, and OL is Official language. Bilingual refers to the ability to speak English and French.
** Tagalog was not included with the Pilipino constraint.

Bibliography

Chepita, R. 2012. *Non-Response Follow-up Strategy and Sampling Design for the 2011 National Household Survey*. Internal report. Statistics Canada. Ottawa. Ontario.

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. December. p. 1016–1041.

Hansen, M.H. and W.N. Hurwitz. 1946. 'The Problem of Non-Response in Sample Surveys'. *Journal of the American Statistical Association*. Vol. 41. p. 517–529.

Hansen, M.H., W.N. Hurwitz and M.A. Bershad. 1959. 'Measurement Errors in Censuses and Surveys'. *Bulletin of the International Statistical Institute*. Vol. 38. p. 359–374.

Kish, L. 1965. *Survey Sampling*. New York. Wiley.

Rodrigue, J.-F., P. Mathieu, J.-P. Morin and J. Taylor. 2012. La méthodologie de collecte du Recensement et de l'Enquête nationale auprès des ménages de 2011 au Canada. *Journées de méthodologie statistique*. Paris, France.

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

Statistics Canada. 2009. 2006 Census Technical Report: Sampling and Weighting. Statistics Canada Catalogue no. 92-568-X. Ottawa, Ontario.

——. 2012. Census Dictionary. Statistics Canada Catalogue no. 98-301-X2011001. Ottawa, Ontario.

——. 2013. *National Household Survey User Guide*. Statistics Canada Catalogue no. 99-001-X2011001. Ottawa, Ontario.

——. 2015. *Census Technical Report on Coverage*. Statistics Canada Catalogue no. 98-303-X201101. Ottawa, Ontario.

——. 2014. CANCEIS, version 5.2. Basic User Guide. Ottawa, Ontario.

Verret, F. 2013. The estimation methodology of the 2011 National Household Survey. In *JSM Proceedings.* Survey Research Methods Section. Montréal, Quebec, Canada. American Statistical Association.