Catalogue No. 92-394-XIE

Coverage

2001 Census Technical Report

Statistics Canada Statistique Canada

Canada


ELECTRONIC PUBLICATIONS AVAILABLE AT
www.statcan.ca

2001 Census Technical Report

Coverage

Introduction

The 2001 Census required the participation of the entire population of Canada, some 30 million people distributed over a territory of 9 million square kilometres. Although there are high quality standards governing the collection and processing of the data, it is not possible to eliminate all errors. In order to help users assess the usefulness of census data for their purposes, the 2001 Census Technical Reports detail the conceptual framework and definitions used in conducting the census, as well as the data collection and processing procedures employed. Also, the principal sources of error, including where possible the size of these errors, are also described, as are any unusual circumstances which might limit the usefulness or interpretation of census data. With this information, users can determine the risks involved in basing conclusions or decisions on census data.

This *2001 Census Technical Report* deals with coverage errors which occurred when persons or dwellings were missed by the 2001 Census or enumerated in error. Coverage errors are one of the most important types of error, since they affect not only the accuracy of the counts of the various census universes, but also the accuracy of all of the census data describing the characteristics of these universes. **Users of census data should be aware that the presence of coverage error in the 2001 Census means that tabulations may present the results of a less than complete enumeration. This is particularly so for groups with a significant coverage error. Users are directed to Chapters 5, 7 and 11 to obtain estimates of population and household coverage errors for a variety of demographic and geographic levels and groupings.**

Chapters 1 to 3 describe the data collection and data processing procedures in which coverage errors may occur as well as the steps taken to control such errors. Chapter 4 defines the census universes that the census attempts to cover and serves as the framework for the Coverage Error Measurement Program. Chapter 5 presents some net undercoverage estimates from the 2001 Census. Chapter 6 gives an overview of the Coverage Error Measurement Program for the 2001 Census. Chapters 7 through 10 describe the methodology and results of each coverage study. Chapter 11 shows how results of the four coverage studies are combined to form estimates of coverage error and gives further estimates. Chapter 12 provides a historical perspective and Chapter 13 presents an evaluation of the largest coverage study, the Reverse Record Check. Finally, Chapter 14 is an overview of the coverage issues related to Indian reserves. Please note that Appendix A lists all of the abbreviations used in this report.

This report has been prepared by Mark Armstrong, Peter Dick, Heather Farr, Gildas Kleim, Josée Morel, Heather Richards, Karen Switzer, Alain Théberge and Christian Thibault, members of the Social Survey Methods Division. Jean Dumais, David Dolson and Jocelyn Tourigny contributed valuable comments on earlier drafts that improved the content and readability of the final report. The support of members of the Census Operations Division, the Demography Division, and the Social Survey Methods Division is noted with appreciation.

Users will find additional information on census concepts, variables and geography in the *2001 Census Dictionary* (Catalogue No. 92-378-XIE), and an overview of the complete census process in the *2001 Census Handbook* (Catalogue No. 92-379-XIE).

1. Census Data Collection

1.1 General

The data collection stage of the 2001 Census process ensures that each of the 11.8 million households in Canada is enumerated on Census Day (Tuesday, May 15, 2001). The census enumerates the entire Canadian population, which consists of Canadian citizens (by birth and by naturalization), landed immigrants, and non-permanent residents. Non-permanent residents are persons living in Canada who have a Minister's permit, student or employment authorization, or who are claiming refugee status, and family members living with them.

The census also counts Canadian citizens and landed immigrants who are temporarily outside the country on Census Day, including federal and provincial government employees working outside Canada, Canadian embassy staff posted to other countries, members of the Canadian Armed Forces stationed abroad, and all Canadian crew members of merchant vessels. Because the census enumerates people where they usually or typically reside rather than where they physically happen to be on Census Day, the Census of Canada is considered a *de jure* census. This means that people outside the country on Census Day were enumerated if their usual or normal place of residence was back in Canada. Some countries conduct a *de facto* census. This type of census is based on where persons actually happen to be on Census Day and not necessarily where they live.

The Census of Canada uses different forms and questionnaires to collect data. The following forms are referred to in this report.

A Form 1 is called a Visitation Record (VR). The VR is used to list every occupied and unoccupied private dwelling, collective dwelling, agricultural operation and agricultural operator in the enumeration area. The VR serves as an address listing for field operations and control purposes for census collection.

The basic short questionnaire is called the 2A. The 2A questionnaire has ten questions and is distributed to every four in five households. The 2B is a longer questionnaire that collects the same information as the 2A plus additional information on a variety of topics. The 2B questionnaire is distributed to every one in five households. Each household that receives a 2A or 2B census questionnaire is asked to enumerate and provide information on all household members who fall into the census population.

A Form 3 (A and B) is used to enumerate persons in a collective dwelling (each person in the collective dwelling would complete a separate Form 3). It can also be used to enumerate usual residents in a private household who prefer to be enumerated on their own census questionnaire rather than be included on a 2A or 2B questionnaire. A Form 4 is completed by census staff in situations where households were absent or dwellings were unoccupied on Census Day.

1.2 Collection Methods

To ensure the best possible collection coverage, Canada is divided into small geographic areas called enumeration areas (EAs). For collection purposes, each EA is under the responsibility of a census representative (CR). CRs are involved in mapping, listing, distribution and verification activities in their assigned EAs and they ensure that all questionnaires are returned to the processing centres. The number of households in an EA ranges from 175 in rural areas to 600 in urban areas. In the 2001 Census, there were 42,851 enumeration areas in Canada. CRs work under the supervision of field census commissioners (CCs). The 2,917 CCs in 2001 were responsible for hiring CRs and for the planning and management of field collection activities in their designated area.

In 2001, approximately 98% of households were self-enumerated. Self-enumeration requires that a CR drop off a census questionnaire at each household during the two weeks before Census Day. An adult, or

any other responsible member of the household, is asked to complete the questionnaire for all members of the household, and then return the questionnaire by mail in a pre-addressed envelope.

Approximately 2% of households were enumerated in the 2001 Census using the canvasser enumeration method. In this case, a CR visits the household and completes a questionnaire for the household by way of an interview. This method is normally used in remote and northern areas of the country, and on most Indian reserves. The canvasser enumeration method is also used in certain urban areas where it is considered highly likely that respondents would not return a questionnaire.

CRs and CCs are involved in a number of field-related collection activities. These include contacting a household to resolve problems that typically relate to the completeness or consistency of the information provided. They also deal with situations where no questionnaire is returned.

During the field collection operations, the CRs delivered a questionnaire to each dwelling within their EA, and wrote the person's name (if possible) and the address in their Visitation Records (VRs). At the same time, they copied down the unique identifiers that would later be captured and used to assign each household and dwelling to the correct geographical area. As well, they identified the block number for the dwelling from their EA map and copied the number into the VR and onto the questionnaire. These block numbers were later data-captured so that all the dwellings in Canada could be identified as belonging to a particular block.

2. Census Data Processing

2.1 Introduction

This part of the census process involved the processing of all the completed questionnaires. This encompasses everything from the key entry of the questionnaire data through to the creation of an accurate and complete retrieval database. Considered here are the steps of manual and automated data capture, questionnaire imaging, editing, error correction, coding, imputation and weighting. The final database was transferred to the Data Quality Measurement Project to determine the overall quality of the data, and to the Census Dissemination Project for the production and marketing of the 2001 Census products and services. In the remainder of this chapter, each data processing operation will be summarized.

An important innovation for the 2001 Census was to create an image retrieval system giving access to the images (pictures) of all the census questionnaires and Visitation Records (see Section 2.3). This would make it possible during subsequent processes to access original census questionnaires and forms without having to manually handle thousands of boxes and paper documents, as was required in past censuses.

2.2 Regional Processing

The Regional Processing team was responsible for the data capture of the questionnaire information into a machine-readable format for subsequent processing. This team was also responsible for the manual research and coding of the industry and occupation responses from 2B questionnaires. Given the number of census questionnaires and quantity of information to be captured (representing over four billion keystrokes), Regional Processing, since the 1981 Census, has been contracting this work out to Revenue Canada, now called the Canada Customs and Revenue Agency (CCRA). CCRA has used their network of systems, resources and staff to key and code census data. By using the staff and infrastructure already in place at CCRA, the census realized cost savings. Census data quality also benefits from the experience that CCRA has in processing past census questionnaires. For the 2001 Census, approximately 2,800 CCRA employees were sworn to secrecy under the Statistics Act to perform the census work. By this arrangement, CCRA employees work under the same rules and regulations as those which apply to the employees of Statistics Canada.

When the collection activities for a specific enumeration area (EA) were completed, the questionnaires, along with maps and Visitation Records, were shipped in EA boxes from the field collection units to one of eight designated CCRA tax centres across Canada.

The first processing step was to prepare completed questionnaires for data capture. This traditionally included the manual assignment of codes to the written answers provided by the respondents. For 2001, most of the written responses were converted to codes using automated systems (see Section 2.5). The only written responses that had to be manually coded for the 2001 Census were the questions on industry and occupation contained on the 2B questionnaires. Research into the automation of the coding of these questions has begun, and it is expected that an automated system will be operational for the 2006 Census.

The industry responses were coded at CCRA according to the North American Industry Classification System (NAICS), which was introduced as a standard within Statistics Canada a few years ago. NAICS is designed to provide a common framework for Canada, the United States and Mexico, which will enable the production of industry statistics under the North American Free Trade Agreement (NAFTA). This meant a change for industry coding from the last census where the type of industry was coded using the 1980 Standard Industrial Classification (SIC). In order to allow longitudinal comparisons, the 2001 industry question on the 2B questionnaire was also coded using the 1980 SIC during the Automated Coding phase (see Section 2.5).

Once the questionnaires were received and registered at one of the CCRA tax centres, and the industry and occupation codes assigned, the next step was to sort, label and batch the questionnaires in preparation for data capture. The labels affixed to each questionnaire contained a unique sequence number that was used to control the movement of the questionnaire throughout the CCRA operations. For the first time, the label also included a bar code to facilitate the scanning of the questionnaire in the imaging operation (see Section 2.3).

Data capture was then performed by traditional manual keying. Verification of the accuracy of the data capture operation was done by selecting a sample of questionnaires that were already key-entered and recapturing the data from the questionnaires in this sample. Quality control statistics were produced by comparing the two sets of captured data. As expected, the keying of data from the census questionnaires introduces some error. Errors occur for a variety of reasons, including inaccurate keying, poorly written or indicated responses on the questionnaires, and missed responses during key entry. The key verification process reduces keying error to a minimum.

As the data were keyed, they were transmitted in real time over dedicated communication lines to the CCRA computer in Ottawa. Within 24 hours, the data were then transferred to tape cartridges and transported by bonded carrier to Statistics Canada, where they were loaded into the mainframe computer. Questionnaires were reassembled into their EA boxes for shipment to the Statistics Canada 2001 processing site in Ottawa. After all the data were keyed, transferred to Statistics Canada and confirmed as being fully received by the Agency, no census data remained with the CCRA.

2.3 Imaging

In previous censuses, the remaining processing steps that required access to the questionnaires and Visitation Records (VRs) used the paper documents. For 2001, the need to handle the paper was eliminated by imaging (scanning) all the questionnaires and VRs as soon as they arrived at the 2001 processing site from the Canada Customs and Revenue Agency (CCRA) centres. Subsequent operations then had access to the questionnaires and VR images using an image retrieval system. This minimized the need to manage the original paper documents.

As the enumeration area (EA) boxes arrived at the 2001 processing site, they were registered. The documents were then prepared for imaging.

The 13 million documents (mainly questionnaires) were imaged using 15 high-volume scanners running five days a week, two shifts per day. The geographic identifier required to identify each document image was automatically assigned using the bar code on the label affixed during the data capture operations at CCRA (see Section 2.2). Quality control was performed to ensure that each document contained the correct number of pages, and that the number of questionnaires by form type was correct for each EA. A resolution operation resolved any difficulties that arose. Images were written to optical platters for subsequent access and archiving. They were also kept in magnetic storage for immediate access by the Interactive Verification activities.

2.4 Interactive Verification

The main objective of Interactive Verification was to identify and correct errors in the data, for which proper resolution required reference to the images of the questionnaires and/or Visitation Records. A detailed set of edit rules was applied to the captured data to identify possible errors, such as households with missing or duplicate persons, incorrect enumeration of foreign or temporary residents, questionnaires assigned to the wrong household, or misclassification of dwellings as occupied or unoccupied. A thorough review of the information on all relevant census forms was conducted to determine the appropriate corrective action for each edit failure. In some cases, this required adding and/or deleting persons or dwellings.

As the census data arrived on cartridges from the Canada Customs and Revenue Agency, they were loaded into Statistics Canada's computers in preparation for the Interactive Verification activities. A series of automated "structural" edits were performed, mainly to verify the information filled out by the Census Representative (CR) on the front cover of the questionnaire. These edits included, among other things, matching questionnaire and household types, cross-checking the number of questionnaires and people enumerated, and verifying that the geographic identifiers were unique. Some edits were also performed on the income information on the 2B questionnaire, so that anomalies could be examined by income subject-matter specialists.

All edits were done by enumeration area (EA). Errors were flagged, and then corrected by referring to the images of the questionnaires and Visitation Record (VR) for that EA. The corrections were made to the electronic data using an interactive PC-based system. Some of the corrections were also electronically noted on the questionnaire images or on the VRs.

Once the EA editing work was completed, automated and manual processes were then used to verify the geographical identifiers that the CR had copied from the EA map onto the questionnaire and VR.

Interactive Verification also performed some special processing to ensure that Canadians living outside Canada on Census Day (people aboard coast guard and Canadian Armed Forces vessels, Canadian-registered merchant vessels, and diplomatic and military personnel) were enumerated properly.

As a final step in the Interactive Verification process, the data were reformatted and forwarded on for the final processing steps. These were the Automated Coding and Edit and Imputation phases.

2.5 Automated and Interactive Coding

Automated coding is the process of matching the write-in responses that were data-captured from the 2B questionnaires during Regional Processing (see Section 2.2) to entries in an automated reference file/classification structure containing a series of words or phrases and corresponding numerical codes. Although a large percentage of write-in responses can be coded in a purely automated manner, a number of responses always remain unmatched. Specially trained coding persons and subject-matter specialists reviewed all unmatched responses. Using the PC-based interactive coding systems and by examining responses to other questions on the questionnaire, sometimes relating to other members of the household, they assigned the appropriate numerical code. Automated coding was applied to write-in responses for the following questions on the 2B questionnaire:

and

- relationship to Person 1;
- language spoken at home;
- non-official languages;
- first language learned in childhood (mother tongue);
- language of work;
- place of birth;
- place of birth of parents;
- citizenship;
- ethnic origin (ancestry);
- population group;
- Indian Band/First Nation;
- place of residence 1 year ago;

- place of residence 5 years ago;
- major field of study;
- religion;
- place of work; and
- industry according to the North American Industry Classification System (and later to the 1980 Standard Industrial Classification).

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

2.6 Edit and Imputation

The data collected in any survey or census contains omissions and inconsistencies. These errors can be the result of respondents answering the questions incorrectly or incompletely, or they can be due to errors generated during processing. For example, a respondent may be reluctant to answer a question, may fail to remember the right answer or may misunderstand the question. Census staff may code responses incorrectly or may make other mistakes during processing.

One of the first tasks of the Edit and Imputation project is to ensure that all dwellings classified as "occupied" have a household size. For those occupied dwellings for which a regular questionnaire (a Form 2A or 2B) was not completed, and for which only the dwelling non-response questionnaire (a Form 4) was received, the first job in Edit and Imputation was to ensure that the dwelling had a valid household size. For those dwellings where the household size was "unknown", the procedure was to impute the household size of the nearest neighbour. However, for 2001, a new procedure was introduced to estimate the household size in these Forms 4 dwellings. See Chapter 7 concerning the Dwelling Classification Study for more details.

The final clean-up of the data was done in Edit and Imputation and was, for the most part, fully automated. It applied a series of detailed edit rules that identified any missing or inconsistent responses. These missing or inconsistent responses were corrected most of the time by changing the values of as few variables as possible through imputation. Imputation invoked either "deterministic" or "minimum-change hot-deck" methods. For deterministic imputation, errors were corrected by inferring the appropriate response value from responses to other questions. For minimum-change hot-deck imputation, a record with a number of characteristics in common with the record in error was selected. Data from this "donor" record were borrowed and used to change the minimum number of variables necessary to resolve all the edit failures.

Two different automated systems were used to carry out this processing.

The **N**earest-neighbour **I**mputation **M**ethod (NIM), developed for the 1996 Census for performing Edit and Imputation for basic demographic characteristics such as age, sex, marital status, common-law status and relationship to Person 1, was expanded for 2001 and implemented in a system called CANCEIS (**CAN**adian **C**ensus **E**dit and **I**mputation **S**ystem) to include Edit and Imputation for such variables as industry, place of work, mode of transportation and mobility. As in 1996, CANCEIS continued to allow more extensive and exact edits to be applied to the response data, while preserving responses through minimum-change hot-deck imputation.

SPIDER (**S**ystem for **P**rocessing **I**nstructions from **D**irectly **E**ntered **R**equirements) was used to process the remaining census variables, such as mother tongue, dwelling and income. This tool translated subject-matter requirements, identified through decision logic tables, into computer-executable modules. SPIDER performed both deterministic and hot-deck imputation.

2.7 Coverage Adjustments for Unoccupied and Non-response Dwellings

The Dwelling Classification Study (DCS) takes a sample of dwellings reported as being either unoccupied or occupied during the collection process. Later, DCS interviewers return to these dwellings to determine if, on Census Day, they were occupied, unoccupied or should not have been listed because they did not meet the census definition of a dwelling.

If a dwelling was occupied, one of two separate adjustments is made to the census database. If the dwelling was listed as unoccupied in the census, then a technique called "random additions" was applied to add households and persons to the census database. In the 2001 Census, 111,628 households and 222,720 persons were added to the database to account for the estimated number of persons living in "unoccupied" dwellings. The second adjustment was concerned with occupied dwellings for which a completed census questionaire was not received, i.e. non-response dwellings, and consisted in adjusting all such dwellings by creating a new household size for them on the census database. A total of 143,681 households with 317,587 persons were added to the census database through this adjustment.

The DCS is the only coverage study that results in the addition of households and persons to the census database. The other three coverage studies presented in this report do not create any changes to the census database. The results of these studies impact in the calculation of official population counts produced at Statistics Canada.

More detail on the scope and methodology of the DCS is given in Chapter 7.

2.8 Weighting

Data on age, sex, marital status, common-law status, mother tongue and relationship to Person 1 were collected from almost all Canadians. However, the bulk of the data gathered in the census came from the one-in-five, or 20%, sample of households which received a 2B questionnaire (see Section 1.1). Weighting, applied to the respondent data after Edit and Imputation, was used to adjust the census sample to represent the whole population.

The weighting method produces weights that are used to form estimates from the 20% sample data. For the 2001 Census, weighting employed a methodology known as calibration (or regression) estimation. Calibration estimation started with initial weights of approximately five and then adjusted them by the smallest possible amount needed to ensure closer agreement between the sample estimates (e.g. number of males, number of people aged 15 to 19) and the actual population counts obtained from the 2A questionnaires for age, sex, marital status, common-law status and household size.

3. Data Quality Measurement

3.1 General

Throughout the census-taking process, every effort was made to ensure high-quality results. Rigorous quality standards were set for data collection and processing, and the Public Communications Program assisted in minimizing non-response. A Data Quality Measurement Program was established to provide users with information on the quality and limitations of census data.

Although considerable effort is made to produce data of the highest quality, the resulting data are subject to a certain degree of inaccuracy. To assess the usefulness of census data for their purposes and to understand the risk involved in drawing conclusions or making decisions on the basis of these data, users should be aware of inaccuracies and need to appreciate their origin and composition.

This report focuses on 2001 Census coverage issues and presents data from various postcensal surveys designed to measure undercoverage and overcoverage. Within the 2001 Census Technical Reports Series, users will find additional detailed 2001 Census information in *Sampling and Weighting* (Catalogue No. 92-395-XIE).

In any census, it is impossible to achieve perfect coverage. This is in spite of field follow-up and verification by the census representatives and the census commissioners and in spite of the steps described in Section 3.2. Therefore, coverage studies are undertaken to measure the extent of coverage errors that occur when dwellings or individuals are missed, incorrectly included or counted more than once. Considered in this report are the Dwelling Classification Study (Chapter 7), the Reverse Record Check (Chapter 8), the Automated Match Study (Chapter 9) and the Collective Dwelling Study (Chapter 10).

3.2 Sources of Coverage Error and Their Control

In most cases, coverage errors occurred during the field collection stage. For example, undercoverage of persons and households occurred when occupied dwellings were missed completely or when they were misclassified as "unoccupied". Population undercoverage also occurred when a person was missed within an enumerated household. Overcoverage can occur when there is uncertainty about a person's usual place of residence. One example of this is a university student who is enumerated at both the home of a parent and at the university residence.

Coverage errors can also be introduced during the processing stage when records for persons or households are erroneously cancelled, lost, or artificially created.

These potential sources of errors were recognized during the planning of the 2001 Census, and the following control measures were taken to minimize them:

(a) careful definition and mapping of enumeration area (EA) boundaries to ensure no areas were left out or included twice;

(b) instructions in the Census Representative's (CR's) manual on how to canvass his/her EA so as to minimize the risk of missing dwellings;

(c) creation of an Address Register from sources independent of the census and the use of this list by the CR to check if any dwellings were missed;

(d) pre-identification of collective dwellings that were to be checked out by field staff to ensure that, if occupied, they were covered in the census;

(e) special procedures to enumerate persons who have difficulty responding (e.g. difficulty in English and French, or literacy problems) and who are located in special core areas of major cities;

(f) special procedures to enumerate the population on Indian reserves and Indian settlements;

(g) publicity messages to inform Canadians about the census, including what to do if they did not receive a questionnaire;

(h) instructions on "Whom to Include" on the census questionnaire to assist respondents in deciding whom to include; and

(i) questions on the census questionnaire asking if there were any persons the respondent was not sure whether or not to list, and a follow-up to assist the respondent in these cases.

These procedures, along with appropriate training, supervisory checks, and quality control systems during census collection and processing, helped reduce the number of coverage errors. However, not all errors can be eliminated, hence the need to evaluate the level of coverage errors.

In the 2001 Census, there is a specific coverage issue which users of census data should be aware of. On some Indian reserves and Indian settlements, enumeration was interrupted before it could be completed. Moreover, some Indian reserves and Indian settlements were enumerated too late to be included, or the quality of the collected data was considered inadequate. A similar problem occurred in the 1996, 1991 and 1986 censuses and, to a lesser extent, in the 1981 Census. These Indian reserves and Indian settlements (a total of 30 in 2001) are called "incompletely enumerated Indian reserves and Indian settlements". Data are not available for these areas and are not included in 2001 Census products. Though these Indian reserves and settlements are not included in the 2001 Census products, population estimates are produced for them based on a statistical model. Chapter 14 gives an overview of this coverage issue. More information relating to Indian reserves and settlements can be found in Dick and So (2004).

3.3 Coverage Errors – Definitions

Coverage errors may be defined as errors that affect the accuracy of the counts of the various census universes (see Chapter 4). There are two types of coverage errors: undercoverage and overcoverage. Undercoverage occurs when a unit that is part of a census universe is completely missed by the census. Overcoverage, on the other hand, may occur in two ways. First, and most common, is the situation where a unit that is part of a census universe is enumerated more than once. Secondly, a unit outside the census universe (e.g. a foreign resident, a fictitious person, or an unoccupied marginal dwelling) may be erroneously enumerated. A geographic error alone does not constitute a coverage error. That is, a person who is enumerated in the wrong geographic area does not constitute overcoverage for the area in which the person was enumerated and undercoverage for the area in which the person should have been enumerated.

Undercoverage of a household is defined as the situation where all persons in the household are missed. Situations where some but not all of the persons in the household are missed are not considered as household undercoverage, even though they cause an error in the characteristics of the household, such as its size or composition. Similarly, overcoverage of a household is defined as the situation where all members of the household are enumerated more than once.

Undercoverage, overcoverage, and the net of the two (net undercoverage) are most usefully expressed as rates of the population that should have been enumerated in the census. Mathematically, let T represent the total or "true" number of units in the universe in question, and let C be the published census count for this universe. Since the concept is the same for persons, households, dwellings or families, we will use the general term "units". The error resulting from the use of C instead of T is then:

$$N = T - C$$

This error, denoted as N, is called the **net coverage error**.

Let U denote the total undercoverage; that is, the total number of units in the universe in question that were missed in the census, and let E denote the total number of units in the universe that were enumerated at least once. Then

$$T = U + E$$

and therefore

$$\begin{aligned} N &= (U + E) - C \\ &= U - (C - E) \\ &= U - O \end{aligned}$$

where the term $O = C - E$ is defined as overcoverage. This error results not only from counting units more than once, but also from counting units that are outside the universe in question.

Expressed as a proportion of the total number of units in the universe in question, the rates of coverage error are:

(a) **undercoverage rate:** $R_U = U/T;$
(b) **overcoverage rate:** $R_O = O/T;$
(c) **net undercoverage rate:** $R_N = N/T;$

so that

$$R_N = (U - O)/T = R_U - R_O$$

A positive net undercoverage rate indicates that undercoverage is larger than overcoverage, whereas a negative net undercoverage rate indicates that undercoverage is smaller than overcoverage. In most cases, undercoverage is larger than overcoverage. Thus, most net undercoverage numbers and rates will be positive.

4. Census Universes

4.1 Introduction

The 2001 Census involved the enumeration of the following five universes:

1. the population universe;
2. the dwelling universe;
3. the household universe;
4. the census family universe; and
5. the economic family universe.

Although a description of each of these universes is given below, the scope of the 2001 Coverage Error Measurement Program addressed only the population and dwelling universes. Readers can refer to the *2001 Census Dictionary* (Catalogue No. 92-378-XIE) which contains more detail on the variables associated with each universe.

4.2 Population Universe

The following groups of persons were included in the population universe of the 2001 Census:

- Canadian citizens and landed immigrants with a usual place of residence in Canada;
- Canadian citizens and landed immigrants who are outside Canada as employees of the Canadian government (federal, provincial and territorial) or members of the Canadian Armed Forces, and their families;
- Canadian citizens and landed immigrants at sea or in port aboard merchant vessels under Canadian registry; and
- non-permanent residents:
    - persons in Canada claiming refugee status;
    - persons in Canada who hold a student authorization (foreign students, student visa holders);
    - persons in Canada who hold an employment authorization (foreign workers, work permit holders);
    - persons in Canada who hold a Minister's permit (including extensions); and
    - all non-Canadian-born dependants of persons claiming refugee status or of persons holding student authorizations, employment authorizations, or Minister's permits.

The following groups of persons, known collectively as foreign residents, were not included in the population universe of the 2001 Census:

- government representatives of another country attached to the embassy, high commission, or other diplomatic body of that country in Canada, and their families;
- members of the Armed Forces of another country who are stationed in Canada and their families; and
- residents of another country visiting in Canada temporarily.

Since the 1991 Census, non-permanent residents are included in the population universe. Previously, non-permanent residents were considered to be foreign residents and were not included in the population

universe. Users should bear this in mind when comparing data from 1991, 1996 and 2001 to data from previous censuses.

The above definitions indicate which persons should be included in the census, but not where these persons should be enumerated. As mentioned, the Canadian census uses the *de jure* method of enumeration, whereby persons are to be enumerated at their usual place of residence, even if they are temporarily away at the time of the census. Persons away from their usual place of residence and residing elsewhere in Canada are to be enumerated at their usual place of residence and are considered "temporary residents" at the other location ("temporary residents" should not be confused with "non-permanent residents", which refers to the legal status of the person while in Canada). Persons without a usual place of residence are to be enumerated wherever they happen to be on Census Day. Some countries use the *de facto* method of enumeration whereby all persons are to be enumerated wherever they are on Census Day, regardless of their usual place of residence.

4.3 Dwelling Universe

A dwelling is defined as a set of living quarters in which a person or group of persons resides or could reside. Only dwellings in Canada are included. There are two types of dwellings:

- A private dwelling is 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 which can be used without passing through the living quarters of someone else.
- A collective dwelling is a dwelling of a commercial, institutional or communal nature. It may be identified by a sign on the premises or by a census representative speaking with the person in charge or with a resident or a neighbour, etc. Included are rooming or lodging houses, hotels, motels, tourist homes, nursing homes, hospitals, staff residences, communal quarters of military bases, work camps, jails, missions, group homes, and so on.

These two main types of dwellings are subject to more detailed classifications:

- Private dwellings can be regular private dwellings, marginal dwellings, or dwellings under construction. Regular private dwellings are further classified into three groups: dwellings occupied by usual residents, dwellings occupied solely by foreign or temporary residents, and unoccupied dwellings. Marginal dwellings and dwellings under construction are classified as occupied by usual residents or occupied solely by foreign or temporary residents. Marginal dwellings and dwellings under construction that were unoccupied on Census Day are not included in the dwelling universe.
- Collective dwellings are classified into dwellings occupied by usual residents, dwellings occupied solely by foreign or temporary residents, and unoccupied collective dwellings. In the case of unoccupied collective dwellings, data were collected but are not included in census products.

In summary, the dwelling universe includes:

- regular private dwellings occupied by usual residents;
- regular private dwellings occupied solely by foreign or temporary residents;
- regular private dwellings that are unoccupied;
- marginal dwellings or dwellings under construction, provided they were occupied on Census Day;
- collective dwellings occupied by usual residents; and
- collective dwellings occupied solely by foreign or temporary residents.

The dwelling universe does not include:

- marginal dwellings or dwellings under construction that were unoccupied on Census Day;
- collective dwellings that were unoccupied on Census Day; and
- dwellings outside Canada.

4.4 Household Universe

The term "household" refers to a person or group of persons, other than foreign or temporary 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 non-family 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 Census Day are considered to be members of the household at their usual place of residence. For census purposes, every person is a member of one and only one household.

Households are classified into three types depending on the type of dwelling in which they reside: private households, collective households, and households outside Canada. Note that households outside Canada are not associated with a dwelling, since dwellings outside Canada are not included in the dwelling universe. Most published census data on households refer to private households only.

4.5 Census Family Universe

The term "census family" refers to:

- a married couple with or without children of either or both spouses living in the same dwelling;
- a couple living common-law with or without children of either or both partners; or
- a lone parent of any marital status, with at least one child living in the same dwelling.

Census families are reported only for the population in private households and households outside Canada.

4.6 Economic Family Universe

An economic family is defined as a group of two or more persons who live in the same dwelling and are related to each other by blood, marriage, common-law, or adoption. Economic families are reported only for the population in private households.

4.7 Relationship of Universes

Table 4.7 gives a summary of the three basic universes according to the location of the dwelling (in Canada or outside Canada), the classification of the dwelling (private or collective), and its occupancy status. For each group in the table, an indication is given as to whether it is included or excluded from each universe. For example, Canadian government employees living outside Canada with their families are included in both the population and household universes but excluded from the dwelling universe. Categories flagged with an asterisk (*) represent groups for which data are collected but excluded from most census products.

Table 4.7 Population, Household, and Dwelling Universes and their Relationships

	Population	Households	Dwellings
Dwellings in Canada			
1. Private dwellings – Regular			
• occupied by usual residents	I	I	I
• occupied solely by foreign or temporary residents	E	E	I*
• unoccupied	–	–	I*
2. Private dwellings – Marginal or under construction			
• occupied by usual residents	I	I	I
• occupied solely by foreign or temporary residents	E	E	I*
• unoccupied	–	–	E
3. Collective dwellings			
• occupied by usual residents	I	I*	I*
• occupied solely by foreign or temporary residents	E	E	I*
• unoccupied	–	E	E
4. Dwelling Classification Study random additions	I	I	I
Dwellings outside Canada			
5. Canadian citizens and landed immigrants who are abroad, either on a military base or attached to a diplomatic mission	I	I*	E
6. Canadian citizens and landed immigrants at sea or in port aboard merchant vessels under Canadian registry	I	I*	E

I = Included E = excluded – = not applicable

* Data were collected but are not included in most products.

5. 2001 Census Net Undercoverage Estimates

The 2001 Census population net undercoverage rate was 2.99%. This means that, on a net basis, the census missed 2.99% of the persons (924,429 persons) that it should have enumerated. The gross population undercoverage rate of 3.95% was offset by the population overcoverage rate of 0.96%. The latter rate indicates that 0.96% of the persons actually enumerated by the census were enumerated in error whereas the former rate indicates that 3.95% of the census target population was not enumerated.

This chapter presents estimates of net undercoverage for a variety of census characteristics. Table 5.1 shows, for each characteristic, the estimated net undercoverage count and standard error as well as the estimated net undercoverage rate and standard error. Estimates presented are for persons.

Table 5.1 Estimated 2001 Census Population Net Undercoverage and Standard Errors, Various Characteristics, 2001 Reverse Record Check

Characteristics	Net number of persons missed		Population net undercoverage rate	
	Estimated number	Standard error	Estimated rate (%)	Standard error (%)
Canada	**924,429**	**44,749**	**2.99**	**0.14**
Newfoundland and Labrador	9,401	1,782	1.80	0.33
Prince Edward Island	1,325	775	0.97	0.56
Nova Scotia	24,521	4,170	2.63	0.44
New Brunswick	20,095	3,555	2.68	0.46
Quebec	140,232	21,033	1.90	0.28
Ontario	436,349	33,472	3.68	0.27
Manitoba	30,903	5,423	2.69	0.46
Saskatchewan	21,231	4,333	2.12	0.42
Alberta	69,857	11,308	2.29	0.36
British Columbia	164,542	15,598	4.04	0.37
Yukon	1,423	372	4.73	1.18
Northwest Territories	3,295	362	8.11	0.82
Nunavut	1,256	411	4.49	1.40
Census metropolitan areas	**626,971**	**38,179**	**3.15**	**0.19**
St. John's	4,541	1,232	2.56	0.68
Halifax	8,097	2,736	2.20	0.73
Saint John	5,522	1,455	4.31	1.09
Chicoutimi–Jonquière[1]	610	2,610	0.39	1.67
Québec	14,598	6,930	2.09	0.97
Sherbrooke	580	2,560	0.38	1.65

Characteristics	Net number of persons missed		Population net undercoverage rate	
	Estimated number	Standard error	Estimated rate (%)	Standard error (%)
Trois-Rivières	443	2,113	0.32	1.53
Montréal	67,291	15,150	1.93	0.43
Ottawa–Hull[2]	44,633	11,258	4.03	0.97
Kingston	-3,993[3]	2,920	-2.79[3]	2.10
London	14,555	6,189	3.26	1.29
Oshawa	2,452	2,897	0.82	0.96
Toronto	255,368	24,343	5.17	0.47
Hamilton	11,519	5,561	1.71	0.81
St. Catharines–Niagara	24,042	6,526	5.99	1.53
Kitchener	9,746	5,581	2.30	1.29
Windsor	2,180	4,159	0.70	1.33
Greater Sudbury	7,802	4,712	4.77	2.75
Thunder Bay	1,819	1,965	1.47	1.56
Winnipeg	10,830	4,015	1.59	0.58
Regina	2,426	1,503	1.24	0.76
Saskatoon	2,768	2,255	1.21	0.97
Calgary	20,044	6,974	2.06	0.70
Edmonton	18,072	6,283	1.89	0.64
Abbotsford	8,481	2,858	5.44	1.73
Vancouver	79,757	11,884	3.86	0.55
Victoria	12,784	3,635	3.94	1.08
Age and sex				
Both sexes	**924,429**	**44,749**	**2.99**	**0.14**
0–4 years	60,665	13,508	3.45	0.74
5–14 years	56,502	16,992	1.38	0.41
15–19 years	52,945	12,913	2.51	0.60
20–24 years	151,443	16,120	7.19	0.71
25–34 years	307,608	17,288	7.15	0.37
35–44 years	187,782	18,640	3.55	0.34
45–54 years	63,011	13,185	1.41	0.29
55–64 years	24,280	11,044	0.84	0.38
65 years and over	20,194	15,833	0.52	0.40

Characteristics	Net number of persons missed		Population net undercoverage rate	
	Estimated number	Standard error	Estimated rate (%)	Standard error (%)
Males	609,231	32,634	3.98	0.20
0–4 years	23,854	8,201	2.67	0.89
5–14 years	16,268	11,306	0.78	0.54
15–19 years	44,224	9,877	4.03	0.86
20–24 years	100,013	12,308	9.24	1.03
25–34 years	209,784	13,773	9.64	0.57
35–44 years	139,629	14,397	5.26	0.51
45–54 years	48,187	10,029	2.16	0.44
55–64 years	14,453	7,837	1.01	0.54
65 years and over	12,820	9,726	0.77	0.58
Females	**315,199**	**31,393**	**2.02**	**0.20**
0–4 years	36,811	10,889	4.26	1.21
5–14 years	40,234	12,730	2.01	0.62
15–19 years	8,722	8,348	0.86	0.82
20–24 years	51,430	10,484	5.02	0.97
25–34 years	97,823	10,574	4.60	0.47
35–44 years	48,153	11,903	1.83	0.44
45–54 years	14,824	8,559	0.66	0.38
55–64 years	9,826	7,775	0.67	0.53
65 years and over	7,375	12,487	0.33	0.56
Legal marital status and sex				
Both sexes	**924,429**	**44,750**	**2.99**	**0.14**
Married or separated	176,984	24,727	1.37	0.19
Divorced	64,975	10,811	3.38	0.54
Widowed	30,792	9,361	1.96	0.58
Never married	635,451	35,162	4.38	0.23
Unknown[4]	16,227	3,503	.	.
Males	**609,231**	**32,634**	**3.98**	**0.20**

Characteristics	Net number of persons missed		Population net undercoverage rate	
	Estimated number	Standard error	Estimated rate (%)	Standard error (%)
Married or separated	117,052	17,816	1.81	0.27
Divorced	56,126	8,567	6.49	0.93
Widowed	5,048	3,589	1.75	1.22
Never married	420,671	26,023	5.46	0.32
Unknown[4]	10,334	2,697	.	.
Females	**315,199**	**31,393**	**2.02**	**0.20**
Married or separated	59,932	17,229	0.93	0.26
Divorced	8,849	6,606	0.84	0.62
Widowed	25,745	8,649	2.01	0.66
Never married	214,781	24,086	3.16	0.34
Unknown[4]	5,893	2,254	.	.
Marital status and sex				
Both sexes	**924,429**	**44,750**	**2.99**	**0.14**
Married, separated or common-law	236,196	26,864	1.56	0.17
Divorced	63,669	10,175	4.67	0.71
Widowed	31,725	9,352	2.10	0.61
Never married	576,613	33,832	4.47	0.25
Unknown[4]	16,227	3,503	.	.
Males	**609,231**	**32,634**	**3.98**	**0.20**
Married, separated or common-law	154,771	19,500	2.04	0.25
Divorced	51,996	8,037	8.96	1.26
Widowed	5,518	3,589	2.08	1.33
Never married	386,613	25,032	5.61	0.34
Unknown[4]	10,334	2,697	.	.
Females	**315,199**	**31,393**	**2.02**	**0.20**
Married, separated or common-law	81,424	18,597	1.08	0.24
Divorced	11,672	6,254	1.49	0.79
Widowed	26,207	8,640	2.11	0.68

Characteristics	Net number of persons missed		Population net undercoverage rate	
	Estimated number	Standard error	Estimated rate (%)	Standard error (%)
Never married	190,002	23,165	3.16	0.37
Unknown[4]	5,893	2,254	.	.
Common-law status				
Both sexes	**924,429**	**44,750**	**2.99**	**0.14**
In common-law union	62,481	11,132	2.66	0.46
Unknown[4]	15,264	3,412	.	.
Males	**609,231**	**32,634**	**3.98**	**0.20**
In common-law union	39,172	8,220	3.31	0.67
Unknown[4]	9,371	2,578	.	.
Females	**315,199**	**31,393**	**2.02**	**0.20**
In common-law union	23,309	7,496	2.01	0.63
Unknown[4]	5,893	2,254	.	.
Mother tongue				
Total	**924,429**	**44,750**	**2.99**	**0.14**
English	429,538	32,162	2.36	0.17
French	124,431	19,837	1.84	0.29
Other than English or French	361,178	24,043	7.70	0.47
English and French	3,024	4,121	0.97	1.31
English and Other	5,463	5,936	0.71	0.77
French and Other	-3,549[3]	186	-3.60[3]	0.20
English, French and Other	-1,984[3]	100	-2.69[3]	0.14
Unknown[4]	6,328	3,481	.	.

. not available

1. Now known as Saguenay.

2. Now known as Ottawa–Gatineau.

3. Negative estimated numbers and rates indicate a net overcoverage.

4. Estimated rates and standard errors are not produced for the category "Unknown".

5.1 Population Net Undercoverage

Population net undercoverage was higher in the West.

Among the ten provinces, population net undercoverage rates were highest in British Columbia (4.04%), followed by Ontario (3.68%) and Manitoba (2.69%). At the Canada level, the highest rate was found in the Northwest Territories (8.11%). Prince Edward Island (0.97%), Newfoundland and Labrador (1.80%) and Quebec (1.90%) had the lowest rates.

Population net undercoverage was highest for young adults aged 20 to 34.

The high population net undercoverage rate for young adults aged 20 to 34 is a reflection of the higher residence mobility in this age group. The net undercoverage rate for persons aged 20 to 24 was 7.19% and for persons between the ages of 25 and 34, it was only slightly lower (7.15%). Net undercoverage was much higher than the national rate for these age groups for both men and women. Males aged 25 to 34 had the highest net undercoverage rate at 9.64%, followed by males in the 20-to-24 age group, at 9.24%. For all age groups combined, net undercoverage was higher for men (3.98%) than for women (2.02%). The higher rate for men held for all age groups beyond age 14.

Population net undercoverage was highest for divorced persons and never-married persons.

The overall difference of 1.96 percentage points between males and females was largely explained by never-married persons and divorced persons. For divorced males, the rate was 6.49%, compared to 0.84% for divorced females. For never-married males, the net undercoverage rate was 5.46%, compared to 3.16% for females in the same group. Using marital status, which differs from legal marital status in that common-law unions are included in the married category, the highest rates of net undercoverage occurred among divorced (4.67%) and never-married persons (4.47%).

Population net undercoverage was slightly lower for persons living common-law.

The net undercoverage for males living common-law was 3.31%, compared to 3.98% for all males. For females living common-law, the net undercoverage rate was 2.01%, compared to 2.02% for all females.

Population net undercoverage was highest for persons who had a mother tongue "Other than French or English".

There were differences in the rates of net undercoverage between persons who reported English as a mother tongue (2.36%) and those who reported French as a mother tongue (1.84%). Among those who reported a non-official language as their mother tongue however, the rate was substantially higher (7.70%).

6. Coverage Error Measurement Program

6.1 Scope and Objectives

The 2001 Coverage Error Measurement Program focused on the population and the dwelling universes. The following components of coverage error are measured:

- classification errors involving unoccupied private dwellings and census non-response dwellings; and
- undercoverage and overcoverage of the population.

The 2001 Coverage Error Measurement Program consisted of four studies:

- Dwelling Classification Study (DCS)
- Reverse Record Check (RRC)
- Automated Match Study (AMS)
- Collective Dwelling Study (CDS)

The DCS produced estimates of undercoverage arising from the incorrect classification of dwellings as unoccupied and from census non-response. The RRC was designed to measure undercoverage from all sources, including the undercoverage measured by the DCS. Overcoverage was measured by the RRC, the AMS, and the CDS. The AMS focused on persons enumerated in more than one household within the same region (Atlantic, Quebec, Ontario and Western Canada, including British Columbia and the three territories). The CDS estimated overcoverage resulting from persons enumerated as usual residents in non-institutional collective dwellings who were also enumerated at a private dwelling. The RRC was designed to measure overcoverage from all sources. If overcoverage is detected in the AMS or the CDS, then it is removed from the RRC. This ensures that no multiple counting of overcoverage occurs between the three overcoverage studies.

Table 6.1 The 2001 Census Coverage Error Measurement Program

Study	Sample size	Objective
Dwelling Classification Study	1,399 enumeration areas	Measures undercoverage from occupied dwellings misclassified as unoccupied and census non-response.
Reverse Record Check	60,653 persons	Measures undercoverage from all sources and overcoverage not included in the Automated Match Study or the Collective Dwelling Study.
Automated Match Study	17,275 pairs of households	Measures overcoverage from persons enumerated in more than one household in the same region.
Collective Dwelling Study	4,500 usual residents	Measures overcoverage from persons enumerated in a non institutional collective dwelling and a private dwelling.

The data collected from these four studies were used in the following ways:

- Estimates from the DCS were included in the final census counts to account for this specific source of undercoverage.
- Estimates from the RRC, the AMS, and the CDS are not included in the census counts. They are included in the base population for the Population Estimates Program of Statistics Canada.
- Information on the causes and characteristics of coverage errors is used in the planning of the next census to identify areas or sub-groups of the population where the level of coverage error is particularly high.
- Supplementary information collected by the studies was used to evaluate the quality of selected census questions. The RRC, for example, provides the means of linking an individual's answers over two consecutive censuses, thereby enabling the calculation of response error for questions such as date of birth, sex, and mother tongue.
- The results serve to inform users about the nature and levels of coverage errors in the census so that they are better informed when drawing conclusions or making decisions based on census results.

The methodology and results of each of the coverage measurement studies are presented in the following chapters.

6.2 Improvements

The 1996 coverage studies were changed as described below to improve their quality and provide a more accurate measurement of coverage error in 2001.

- The Vacancy Check (VC) was replaced by the Dwelling Classification Study (DCS). This new study is used to correct census counts prior to publication, for dwellings wrongly classified as unoccupied and dwellings identified as occupied for which the census was unable to find a contact, by collecting information on the characteristics of the occupants of those dwellings (also referred to as non-response dwellings). The DCS is more comprehensive and more precise than the old VC.

- Data for the Reverse Record Check (RRC) were collected by computer-assisted telephone interviewing (CATI). Unlike the paper questionnaires used in the past, CATI has the capability to do automated checks and to correct data instantly. As a result, CATI has helped improve data quality. It has also substantially reduced the delays in transferring data to head office and in running additional checks on the data.

- In the case of the Reverse Record Check (RRC), the system that processes the addresses of the selected persons was automated. This system determines whether the persons included in the sample were enumerated at the collected addresses or not. In order to carry out this processing, the clerks had access to all the information they needed at their workstations (e.g., information about the people and the addresses, electronic telephone books, maps). Because the system was more efficient, it was possible to process more addresses and significantly reduce the number still unresolved at the end of processing. These positive results are also partly due to the use of more sophisticated research tools, such as electronic telephone books and maps.

- The 2001 RRC produced estimates for all coverage errors, including those already corrected in published census counts. Adjustments made to the census counts before publication (based on the results of the DCS) were eliminated, and the results of the RRC were used for all cases of undercoverage.

- The institutional component of the Collective Dwelling Study (an overcoverage study) was dropped, and overcoverage estimates for this group (historically very small) were made as part of the RRC.

7. Dwelling Classification Study

7.1 Introduction

One of the potential sources of error in the census is the misclassification of dwellings. This can occur when a dwelling that is truly occupied is classified as "unoccupied", or when a unoccupied dwelling is misclassified as "occupied". If an occupied dwelling is incorrectly classified as unoccupied, undercoverage of both persons and occupied dwellings occurs in the census counts. If an unoccupied dwelling is misclassified as occupied, then an overcount of persons and occupied dwellings occurs. This misclassification of an unoccupied dwelling as occupied can occur when a dwelling was identified as occupied by a census representative (CR) but did not return a census questionnaire. This type of dwelling is referred to as a non-response dwelling.

The erroneous inclusion of marginal dwellings or dwellings under construction in the unoccupied dwellings classification also results in overcoverage, this time of the housing stock.

The purpose of the Dwelling Classification Study (DCS) is to study these types of classification error. The DCS is an extension of the Vacancy Check which was used in previous censuses to re-examine dwellings which were classified as unoccupied by the CR.

The uses of the information collected by the 2001 DCS are as follows:

- to estimate the number of unoccupied dwellings that were outside the housing universe. This includes marginal dwellings, dwellings under construction, and buildings that were not actually dwellings (gas stations, doctor's offices, etc.);
- to estimate the number of occupied dwellings that were misclassified as unoccupied during the census, and estimate the number of households and persons missed as a result of this misclassification;
- **to adjust the census data for households and persons to correct this misclassification;**
- to estimate the number of census non-response dwellings that were unoccupied;
- to estimate the number of persons living in non-response dwellings; and
- **to adjust the household size distribution through imputation for the non-response dwellings.**

The last three points are new to the DCS in 2001, when the study scope was expanded in order to examine non-response dwellings as well as unoccupied dwellings.

7.2 Methodology

7.2.1 Stratification and Sample Selection

The population targeted by the Dwelling Classification Study (DCS) was all non-response dwellings and all unoccupied dwellings identified in the census as of May 15, 2001, excluding dwellings in collective enumeration areas (EAs), canvasser EAs and Indian reserve EAs. These areas were excluded from the sampling frame because of cost and operational considerations.

The sample size for the 2001 DCS was set at 1,399 EAs across Canada. The sampling frame included all mail-back EAs with the exception of Indian reserves. Therefore, the rural areas of the Yukon and Northwest Territories, and all of Nunavut, were not part of the sampling frame. Only Whitehorse, Yellowknife, Hay River and Fort Smith were included in the territorial sample.

The initial screening for the sample selection was done in order to select only EAs which had a mail-back method of collection and which were not linked to Indian reserves, Indian settlements, or other types of Indian census subdivisions. EAs were then split into separate urban and rural frames for sample selection. In order to be included in the urban frame, an EA must initially have been part of a census agglomeration (CA) or census metropolitan area (CMA) of 50,000 or more occupied dwellings. If more than 50% of the EAs in a census commissioner district (CCD) located within the selected CA/CMAs were linked to urban areas, then all EAs within that CCD were considered to be urban, and a single-stage sampling method was used to select the sample. Otherwise, all EAs within that CCD were considered to be rural, and a two-stage sampling method was used. All EAs that did not fall into the urban definition became part of the rural frame. The DCS sample was then selected from these two frames.

Urban (Single-stage) Sample

Single-stage sampling was used in geographically compact areas. Although these tended to be mostly the urban areas, there were three separate components to this sampling frame. In the Yukon (Whitehorse only) and Northwest Territories (Yellowknife, Hay River and Fort Smith only), all EAs in the frame were selected for the DCS. In Prince Edward Island, a simple random sample of 45 EAs was selected. The urban sample for all other provinces was selected by stratifying the urban EAs by CA/CMA within each province. Here, each CA/CMA was considered a stratum, and a simple random sample of the required number of EAs was then selected from each. Each stratum sample contained at least five EAs. This gave a sample of 678 urban EAs in all.

Rural (Two-stage) Sample

The rural sample was selected with the use of a two-stage sampling method in order to reduce field costs. The DCS interviewer field costs, especially travel costs, can rise substantially outside urban areas. Based on prior census data (1986, 1991 and 1996), five EAs grouped together were determined to be an appropriate workload for a DCS interviewer. In order to group five EAs close enough to form a relatively tight unit, two-stage sampling was used. In the first stage, the allocated number of rural CCDs was randomly selected for each province. In the second stage, five EAs were randomly selected from each of the selected CCDs. These sampling procedures produced the 721 EAs in the rural sample.

The DCS sample consisted of all unoccupied dwellings and all non-response dwellings in the sampled EAs. As a result, a total of 29,777 unoccupied and 2,749 non-response dwellings were selected to be part of the DCS sample. Table 7.2.1 shows the sample distribution by province and territory.

Table 7.2.1 Sample Size, Canada, Provinces and Territories, 2001 Dwelling Classification Study

	Number of enumeration areas in sample	Number of unoccupied dwellings in sample	Number of non-response dwellings in sample
Canada	**1,399**	**29,777**	**2,749**
Newfoundland and Labrador	80	2,635	58
Prince Edward Island	45	874	1
Nova Scotia	96	2,501	287
New Brunswick	65	1,090	124
Quebec	271	6,700	860
Ontario	278	4,513	444
Manitoba	94	2,991	90
Saskatchewan	126	2,400	91
Alberta	125	2,157	276
British Columbia	161	2,733	455
Yukon	28	330	55
Northwest Territories	30	853	8
Nunavut	0	0	0

7.2.2 Field Interviews

All dwellings that were either classified as unoccupied on Census Day, or were classified as occupied, but for which no census form was returned in the sampled enumeration areas, were to be checked again in late June or early July 2001, to determine the true occupancy status of the dwellings on Census Day. A Dwelling Classification Study questionnaire was designed for this purpose.

The timing of this operation after census enumeration was left to the discretion of each regional office. In order to determine occupancy status and collect other information, census representatives were instructed to contact current occupants, neighbours, landlords, or any other person with some knowledge about the dwelling. Up to three contact attempts were made for each dwelling. If the dwelling was found to have been occupied on Census Day, the number and the names of occupants of the dwelling on Census Day were also obtained.

7.2.3 Processing, Coding, and Editing

All questionnaires were sent to Ottawa for processing after interviews were completed. Once in head office, they went through preliminary processing and were then data captured.

First, any questionnaires not belonging to the sample were eliminated, and in cases where more than one questionnaire was completed for an address, the correct questionnaire was obtained for each dwelling. Some preliminary edits and general grooming were then carried out before the questionnaires were sent for data capture.

Once data capture was completed, the questionnaires were subjected to an extensive set of consistency edits. The failed-edit questionnaires were examined manually in order to resolve any inconsistencies.

For each dwelling which was found to have been occupied on Census Day, the Dwelling Classification Study (DCS) questionnaire was checked to determine whether another address was listed where the household members might have been enumerated. If an alternate address was given, then the Visitation Record and other census questionnaires were checked to see if the household members were indeed enumerated elsewhere. If they were found to be enumerated elsewhere, then they were not added to the person count, even though the dwelling itself was added to the occupied dwelling count.

At this point, the processing of unoccupied dwellings and non-response dwellings was separated. The list of DCS questionnaires completed for each sampled enumeration area was matched against the final listing of unoccupied dwellings in the census in order to create a final file of unoccupied dwellings for further processing. Dwellings for which a DCS questionnaire was received but no listing was found in the census were removed from the study. Dwellings which were listed in the census as unoccupied, but for which no DCS questionnaire was received, were considered to be non-response dwellings for this study. Similarly, the list of DCS questionnaires was matched against a final list of dwellings in the census which did not return a questionnaire. This was done in order to compile a final file of non-response dwellings for further processing.

7.2.4 Non-response, Imputation, and Weighting

The processing for the unoccupied dwellings and the non-response dwellings, while done separately, was identical for the non-response, imputation and weighting processes.

Total non-response (i.e. no information for a particular dwelling) was handled with an adjustment to the weights within each of several subprovincial areas. These subprovincial areas consisted of the three largest census metropolitan areas (Montréal, Toronto and Vancouver) along with the remaining urban and rural parts of each province and territory.

Item non-response for occupancy status, number of usual residents, and dwelling type was addressed by imputation. Occupancy status was imputed first and then used in the imputation of the other variables.

The weights were then adjusted so that their sum would give the known number of unoccupied/non-response dwellings found in the census for each subprovincial area.

The final step of the Dwelling Classification Study (DCS) processing was the actual adjustment of the census databases. This was accomplished for the unoccupied dwellings by first producing a national-level profile of misclassified dwellings for both urban and rural areas, using the type of dwelling and the number of persons missed because of the misclassification. These national profiles were then used to create estimates of the number of misclassified dwellings by number of persons in the household, type of dwelling, and rural/urban parts at the province and territory level. On the basis of these estimates, enumerated households with the same characteristics (number of persons, type of private dwelling) were selected at random, and their weights in the census were increased by one unit. For each household selected, the weight of one unoccupied dwelling from the same EA was set to zero so that the total number of dwellings would not be increased.

The census was adjusted in the following way for the number of persons living in non-response dwellings. If a household size was entered onto the census questionnaire by the Census Representative in the field, it was accepted as being correct. A profile of non-response dwellings that gave the estimated number of dwellings for each household size (from zero to six persons) for each geographic area was generated by the DCS. A second distribution was created from the observed census data, and a misclassification vector (the difference between these two distributions) was created. Non-response households were then randomly selected from those available in each geographic area, and new household sizes were imputed for them until the misclassification vector was equal to zero (i.e. until the distribution of household sizes

for non-response households in the census was equal to the distribution of non-response household sizes from the DCS).

7.3 Results

The main results for those dwellings which were initially found to be unoccupied by the 2001 Census are shown in tables 7.3.1A, 7.3.1C, 7.3.1D, and 7.3.1E. Table 7.3.1A gives the estimated number and rate of dwellings misclassified as unoccupied by urban/rural area, region, province, selected census metropolitan area and by type of private dwelling. Table 7.3.1B gives the same information for the 1996 Vacancy Check Table 7.3.1C shows the number of households and persons added to the 2001 Census counts because of these misclassifications. Table 7.3.1D shows the number of unoccupied dwellings not in the housing stock, using a similar breakdown of areas as the one given in Table 7.3.1A. Table 7.3.1E shows the dwelling classification error in the census due to dwellings which should not have been part of the housing stock being erroneously classified as unoccupied.

The results for non-response dwellings (dwellings for which a census form was not returned but which were considered occupied) are shown in Table 7.3.2. This table gives the number of persons estimated by the DCS (with estimated standard errors) to be living in these dwellings.

7.3.1 Dwellings Listed as Unoccupied in the Census

A. Occupied Dwellings

Table 7.3.1A shows that at the Canada level 12.7% of dwellings classified as unoccupied during the census were actually occupied. This includes dwellings which were occupied solely by foreign or temporary residents, as well as dwellings for which one or more persons was enumerated elsewhere in Canada. This number is up from 7.8% in 1996 (see Table 7.3.1B). The misclassification of dwellings was much more prevalent in urban areas (21.3%) than in rural areas (5.7%). The rate for rural areas is very close to that of 1996, while the rate for urban areas represents a large increase from that of 9.9% in 1996. The increase in urban areas between 1996 and 2001 is largely due to increases in the three largest census metropolitan areas (CMAs) (Montréal, Toronto and Vancouver), particularly in large apartment buildings.

Table 7.3.1A Estimated Number of Occupied Dwellings Misclassified as Unoccupied in the 2001 Census, 2001 Dwelling Classification Study [1]

Region / Private dwelling type	Number of dwellings classified as unoccupied in the 2001 census	Occupied dwellings			
		Estimated total	Standard error	Rate (%)	Standard error (%)
Canada	**904,236**	**114,603**	**7,166**	**12.7**	**0.8**
Urban	405,427	86,370	5,814	21.3	1.4
Rural	498,809	28,233	4,176	5.7	0.8
Atlantic provinces	**112,357**	**6,961**	**693**	**6.2**	**0.6**
Newfoundland and Labrador	35,896	1,511	291	4.2	0.8
Prince Edward Island	5,121	396	100	7.7	2.0
Nova Scotia	42,931	3,411	519	7.9	1.2
New Brunswick	28,409	1,643	341	5.8	1.2

Region / Private dwelling type	Number of dwellings classified as unoccupied in the 2001 census	Occupied dwellings			
		Estimated total	Standard error	Rate (%)	Standard error (%)
Quebec	213,062	31,007	4,577	14.6	2.1
Ontario	316,474	36,647	4,676	11.6	1.5
Prairies	**150,294**	**15,303**	**1,729**	**10.2**	**1.2**
Manitoba	37,592	2,737	434	7.3	1.2
Saskatchewan	47,503	3,113	486	6.6	1.0
Alberta	65,199	9,452	1,602	14.5	2.5
British Columbia	110,755	24,545	2,252	22.2	2.0
Territories	**1,294**	**140**	**27**	**10.8**	**2.1**
Yukon	395	72	10	18.2	2.5
Northwest Territories	899	68	25	7.6	2.8
Selected census metropolitan areas					
Montréal	42,174	8,435	1,737	20.0	4.1
Toronto	40,020	17,557	4,330	43.9	10.8
Vancouver	29,565	11,034	1,614	37.3	5.5
Type of private dwelling					
Single-detached	423,328	46,616	3,554	11.0	0.8
Apartment in a building with five or more storeys	39,365	18,796	4,427	47.7	11.2
Other type of dwelling	230,899	49,191	4,434	21.3	1.9
Not in housing stock	210,644	...	...	...	...

... not applicable

[1] Nunavut is not applicable

Table 7.3.1B Estimated Number of Occupied Dwellings Misclassified as Unoccupied in the 1996 Census, 1996 Vacancy Check

Region / Private dwelling type	Number of dwellings classified as unoccupied in the 1996 census	Occupied dwellings			
		Estimated total	Standard error	Rate (%)	Standard error (%)
Canada	**781,594**	**61,287**	**3,296**	**7.8**	**0.4**
Urban	415,474	41,295	2,570	9.9	0.6
Rural	366,120	19,992	2,021	5.5	0.6
Atlantic provinces	**67,671**	**3,303**	**488**	**4.9**	**0.7**
Newfoundland	23,081	1,001	303	4.3	1.3
Prince Edward Island	3,109	176	40	5.7	1.3
Nova Scotia	25,842	1,305	291	5.1	1.1
New Brunswick	15,639	822	246	5.3	1.6
Quebec	216,838	13,298	1,502	6.1	0.7
Ontario	262,721	26,611	2,457	10.1	0.9
Prairies	**142,773**	**9,187**	**849**	**6.4**	**0.6**
Manitoba	32,598	1,900	326	5.8	1.0
Saskatchewan	40,276	1,757	341	4.4	0.8
Alberta	69,899	5,530	706	7.9	1.0
British Columbia	90,112	8,714	1,269	9.7	1.4
Territories	**1,479**	**173**	**32**	**11.7**	**2.2**
Yukon	748	147	30	19.7	4.0
Northwest Territories	731	26	10	3.6	1.4
Selected census metropolitan areas					
Montréal	70,552	5,093	858	7.2	1.2
Toronto	38,210	8,259	1,588	21.6	4.2
Vancouver	29,711	3,737	765	12.6	2.6
Type private dwelling					
Single-detached	329,517	24,987	2,172	7.6	0.7
Apartment in a building with five or more storeys	62,829	7,593	1,584	12.1	2.5
Other type of dwelling	272,067	28,707	2,102	10.6	0.8

At the provincial and territorial level in 2001, British Columbia had the highest rate of misclassification at 22.2%, followed by theYukon (18.2%), Quebec (14.6%), Alberta (14.5%), and Ontario (11.6%). The misclassification rates for the other provinces were fairly consistent, ranging from 4.2% in Newfoundland and Labrador to 7.9% in Nova Scotia.

Among the three largest CMAs, the 2001 rate of misclassification is very high in all three areas, with the rate in Toronto (43.9%) being somewhat higher than the rates in Vancouver (37.3%) or Montréal (20.0%).

Among the types of private dwellings classified in the census, the rate of misclassification is lowest in single-detached houses (11.0%) and highest in apartments in buildings of five or more storeys (47.7%). In the "Other type of dwelling" category, which includes semi-detached houses, row houses, duplexes, apartments in buildings with fewer than five storeys, mobile homes and other movable dwellings, the rate of misclassification is 21.3%.

Because of this classification error, a number of households and persons were not enumerated in the 2001 Census. However, among the estimated 114,603 dwellings misclassified as unoccupied, some had also been correctly enumerated by the Census Representative as occupied dwellings since they were listed twice in the Visitation Record, and some were occupied exclusively by temporary or foreign residents who, correctly, should not have been included in the census counts. Therefore, the actual number of underenumerated households was estimated at 111,628. This is the number of households that was added to the census counts via the Dwelling Classification Study (DCS). Table 7.3.1C shows the actual number of households and persons added to the census counts by the DCS.

Table 7.3.1C Number of Households and Persons Added to the 2001 Census Counts in 'Unoccupied' Dwellings by the 2001 Dwelling Classification Study [1]

Region / Private dwelling type	Number of dwellings classified as unoccupied in the 2001 census	Households added		Persons added	
		Total	Standard error	Total	Standard error
Canada	**904,236**	**111,628**	**6,815**	**222,720**	**11,711**
Urban	405,427	84,11	5,418	170,41	9,144
Rural	498,809	27,518	4,124	52,31	7,298
Atlantic provinces	**112,357**	**6,738**	**672**	**12,219**	**1,081**
Newfoundland and Labrador	35,896	1,407	258	2,439	419
Prince Edward Island	5,121	384	99	768	184
Nova Scotia	42,931	3,304	509	6,058	756
New Brunswick	28,409	1,643	341	2,954	622
Quebec	213,062	29,370	4,126	54,267	6,782
Ontario	316,474	36,188	4,617	78,393	8,204
Prairies	**150,294**	**15,047**	**1,721**	**29,610**	**3,293**
Manitoba	37,592	2,648	437	4,663	692
Saskatchewan	47,503	2,977	472	5,383	943
Alberta	65,199	9,422	1,597	19,564	3,078

Region / Private dwelling type	Number of dwellings classified as unoccupied in the 2001 census	Households added		Persons added	
		Total	Standard error	Total	Standard error
British Columbia	110,755	24,145	2,164	47,933	3,440
Territories	**1,294**	**140**	**27**	**298**	**55**
Yukon	395	72	10	151	24
Northwest Territories	899	68	25	147	50
Selected census metropolitan areas					
Montréal	42,174	8,392	1,720	16,231	3,142
Toronto	40,020	17,274	4,292	40,106	8,413
Vancouver	29,565	10,845	1,586	22,869	3,123
Type of private dwelling					
Single-detached	423,328	45,855	3,525	106,434	7,801
Apartment in a building with five or more storeys	39,365	17,315	3,894	28,028	6,011
Other type of dwelling	230,899	48,458	4,367	88,257	8,024

[1] Nunavut is not applicable.

B. Dwellings Not in the Housing Stock

The enumeration of unoccupied dwellings which fall outside the housing universe results in overcoverage of dwellings. A dwelling is considered to be outside the housing universe if it is used for commercial purposes, if it is not habitable year round, and if it is double-counted in the census. This last situation can happen when a dwelling is listed in the Visitation Record both as occupied and unoccupied.

In order for a dwelling to be considered suitable for year-round occupancy, it must have a source of heat or power and provide complete shelter from the elements. It is sometimes difficult to tell whether a dwelling is in fact habitable such as in the case of a cottage, a dwelling under construction that is almost complete, or a dwelling that has deteriorated. The question of suitability can therefore involve a degree of subjectivity, making it possible for census representatives to classify dwellings differently. For this reason, the estimates provided in Table 7.3.1D of the number of unoccupied dwellings identified in the Dwelling Classification Study as not being part of the housing stock should be used with caution. The census database is not adjusted to reflect these values.

Table 7.3.1D Estimated Number of Unoccupied Dwellings Not in the Housing Stock, 2001 Dwelling Classification Study [1]

Region	Number of dwellings classified as unoccupied in the 2001 census	Not in housing stock			
		Estimated total	Standard error	Rate (%)	Standard error (%)
Canada	**904,236**	**210,644**	**15,264**	**23.3**	**1.7**
Urban	405,427	82,915	7,368	20.5	1.8
Rural	498,809	127,729	13,239	25.6	2.7
Atlantic provinces	**112,357**	**16,710**	**2,401**	**14.9**	**2.1**
Newfoundland and Labrador	35,896	3,424	776	9.5	2.2
Prince Edward Island	5,121	2,132	436	41.6	8.5
Nova Scotia	42,931	5,748	926	13.4	2.2
New Brunswick	28,409	5,406	2,030	19.0	7.1
Quebec	213,062	73,410	9,422	34.5	4.4
Ontario	316,474	52,008	8,298	16.4	2.6
Prairies	**150,294**	**45,914**	**5,103**	**30.5**	**3.4**
Manitoba	37,592	8,466	2,328	22.5	6.2
Saskatchewan	47,503	13,837	2,584	29.1	5.4
Alberta	65,199	23,611	3,734	36.2	5.7
British Columbia	110,755	22,108	6,598	20.0	6.0
Territories	**1,294**	**493**	**159**	**38.1**	**12.3**
Yukon	395	56	13	14.2	3.3
Northwest Territories	899	437	158	48.6	17.6
Selected census metropolitan areas					
Montréal	42,174	12,311	3,296	29.2	7.8
Toronto	40,020	5,473	1,752	13.7	4.4
Vancouver	29,565	6,224	2,133	21.1	7.2

[1] Nunavut is not applicable.

At the Canada level, dwellings outside the housing stock account for 23.3% of all dwellings classified as unoccupied in the census. The problem is slightly more pronounced in rural areas (25.6%) than in urban areas (20.5%). At the province and territory level, the incidence of dwellings outside the housing stock being classified as unoccupied ranges from 9.5% in Newfoundland and Labrador to 48.6% in the Northwest Territories. Prince Edward Island is notably high at 41.6%.

Table 7.3.1E shows the dwelling classification error in the 2001 Census resulting from dwellings which should never have been part of census enumeration being classified as unoccupied. This error is estimated at 1.68% of all private dwellings in the 2001 Census. At the province and territory level, this error ranges from a low of 0.41% in the Yukon Territory to a high of 3.81% in Prince Edward Island. For the three largest CMAs, the error ranges from 0.33% in Toronto to 0.84% in Montréal.

Table 7.3.1E 2001 Census Dwelling Classification Error, Canada, Provinces, Territories and Selected Census Metropolitan Areas, 2001 Dwelling Classification Study

		Dwellings not in housing stock			
	Total dwellings	Estimated number	Standard error	Rate (%)	Standard error (%)
Canada	**12,548,588**	**210,644**	**15,264**	**1.68**	**0.12**
Newfoundland and Labrador	227,570	3,424	776	1.50	0.34
Prince Edward Island	55,992	2,132	436	3.81	0.78
Nova Scotia	403,819	5,748	926	1.42	0.23
New Brunswick	313,609	5,406	2,030	1.72	0.65
Quebec	3,230,196	73,410	9,422	2.27	0.29
Ontario	4,556,240	52,008	8,298	1.14	0.18
Manitoba	477,085	8,466	2,328	1.77	0.49
Saskatchewan	431,628	13,837	2,584	3.21	0.60
Alberta	1,171,841	23,611	3,734	2.01	0.32
British Columbia	1,643,969	22,108	6,598	1.34	0.40
Yukon	13,793	56	13	0.41	0.09
Northwest Territories	14,669	437	158	2.98	1.08
Selected census metropolitan areas					
Montréal	1,473,491	12,311	3,296	0.84	0.22
Toronto	1,671,087	5,473	1,752	0.33	0.10
Vancouver	786,277	6,224	2,133	0.79	0.27

7.3.2 Dwellings Listed as Non-response in the 2001 Census

Table 7.3.2 shows the number of non-response dwellings in the 2001 Census and gives the number of persons added to them through the Dwelling Classification Study. For an in-depth comparison of these results and those which would have been obtained using the old imputation method for non-response dwellings, see Dick (2002a) and Dick (2002b).

Table 7.3.2 Number of Persons in Dwellings Classified as Non-response in the 2001 Census, Canada, Provinces and Territories, 2001 Dwelling Classification Study

	Number of dwellings classified as non-response in the census	Occupied households		Persons	
		Total	Standard error	Total	Standard error
Canada	**179, 788**	**143,681**	**2,352**	**317,587**	**14,841**
Newfoundland and Labrador	1,431	1,185	67	2,268	179
Prince Edward Island	508	392	83	978	206
Nova Scotia	5,063	3,980	332	8,042	511
New Brunswick	3,303	2,676	60	5,501	66
Quebec	65,787	50,834	1,473	100,741	7,789
Ontario	59,784	48,396	1,686	124,825	12,282
Manitoba	3,798	3,254	142	6,602	161
Saskatchewan	3,246	2,313	144	5,142	172
Alberta	14,197	11,834	370	26,982	423
British Columbia	22,472	18,697	470	36,269	2,862
Yukon	118	90	10	177	12
Northwest Territories	81	30	21	60	29

8. Reverse Record Check

8.1 Introduction

Following each census since 1966, the Reverse Record Check (RRC) has been carried out to measure gross undercoverage, that is, to estimate the number of persons and households missed in the census. The RRC results are combined with the findings of the other coverage studies to calculate net undercoverage.

Population and household undercoverage is generally regarded as one of the most important sources of error affecting census data. It causes a downward bias to the extent that the census figures underestimate the true population and household counts. Overcoverage, on the other hand, results in an upward bias whereby census data overestimate the true population and household counts. These two sources of error may also distort the distribution of population and household characteristics estimated from census data if overcounted and uncounted persons do not have the same characteristics as enumerated individuals.

The main objectives of the 2001 RRC were:

(a) to study the effects of population undercoverage in the 2001 Census and produce estimates of undercoverage for provinces and territories and for some major subgroups of the population;

(b) to study the characteristics of individuals missed in the census and identify possible reasons for the errors; and

(c) to obtain an indication of the level of overcoverage of individuals living in private or collective dwellings on Census Day in the 2001 Census.

Producing census family, economic family and household estimates of undercoverage and overcoverage error was not part of the original scope of the 2001 coverage studies. Therefore, there are no results on coverage error for these population groups. The household undercoverage error in the 1996 Census was estimated to be 2.49% at the national level. The national overcoverage rate for households was estimated to be 0.30%, resulting in the net undercoverage rate of 2.19%. Readers can refer to the 1996 Census technical report entitled *Coverage* (Catalogue No. 92-370-XPB) and to its Table 10.3 for the most recent detailed statistics on the census household coverage error.

8.2 Methodology – Sample Preparation

The target population, which consisted of all persons who should have been enumerated in the 2001 Census, was formed from six sources or sampling frames. The first five frames were used to estimate undercoverage and overcoverage in the ten provinces, whereas estimates for the three territories were calculated exclusively on the basis of samples from the sixth frame. A total of 60,653 persons were selected for the sample. Table 8.2 shows the sample distribution by frame.

Table 8.2 Sample Frames, 2001 Reverse Record Check

Sample frame	Definition	Sample size (persons)
1. 1996 Census	All persons enumerated in the 1996 Census.	44,930
2. Missed	All persons from the 1996 Reverse Record Check (RRC) sample who were classified as "not enumerated". These persons kept the same weights as in the 1996 RRC.	2,223
3. Births	All children born between May 14, 1996 and May 14, 2001.	2,805
4. Immigrants	All landed immigrants who arrived in Canada between May 14, 1996 and May 14, 2001.	2,630
5. Non-permanent residents	All persons holding employment or student authorizations or Minister's permits (including extensions) and persons claiming refugee status who were in Canada on May 15, 2001.	1,060
6. Health care files	All persons listed in the health insurance files of the Yukon, the Northwest Territories and Nunavut who were in Canada on May 15, 2001.	7,005
Total		**60,653**

Sampling was carried out independently within each frame. The sample design varied from frame to frame depending on the nature of the list used. The sampling rates within frames were not uniform. To improve the efficiency of the sample, higher sampling rates were applied in subgroups for which high undercoverage or overcoverage or a lower tracing rate was expected.

In the 1996 Census frame, the sample design used was single-stage sampling with demographic stratification and optimum allocation based on historical undercoverage and overcoverage rates (new for 2001), historical tracing rate and stratum size. The population was stratified by province of residence, sex, age, and marital status. Persons enumerated on Indian reserves and settlements and in collective dwellings in the 1996 Census were placed in separate strata. A sample was then selected within each stratum in order to include the largest possible number of "missed" cases. The sampling fractions were not the same in all strata. For example, as was the case in 1996, single males aged 20 to 24 in 2001 had a greater probability of being selected since it had been observed in previous Reverse Record Checks (RRCs) that undercoverage and overcoverage were consistently higher in this stratum.

The Missed frame is a conceptual frame since there is no list of all the persons who were not enumerated in the 1996 Census. The sample for this frame consists of all cases classified as "not enumerated" in the 1996 RRC. The sample is not stratified as such, although there is implicit stratification since cases not enumerated in 1996 came from different frames and strata in the 1996 RRC.

For the Births frame, copies of all birth registrations for the intercensal period were obtained from vital statistics. The frame was then stratified by the mother's province of residence and the child's year of birth and single-stage sampling was then used.

The Immigrants frame was constructed using immigration records from Citizenship and Immigration Canada. It was stratified by province. For Quebec, Ontario and British Columbia, the three provinces having the highest influx of immigrants, three strata were formed based on year of arrival in Canada: a first for immigrants who arrived between 1996 and 1999, a second for year 2000 arrivals, and a final one for arrivals in 2001. Single-stage sampling was then used for each of these two frames.

The non-permanent residents frame was created using records from Citizenship and Immigration Canada. It was stratified by province. Strata were then formed for Quebec, Ontario and British Columbia, for three different categories of persons: those having a student authorization or a Minister's permit, those having an employment authorization and those claiming refugee status. This same stratification could not be applied in the other provinces, where numbers are smaller for these population groups. Single-stage samples were selected within each stratum.

Finally, for the three territories, the frames were constructed using the respective health care files and stratified by age, sex and region (urban or rural for the Yukon and the Northwest Territories, and Iqaluit or non-Iqaluit for Nunavut). Single-stage samples were selected within each stratum.

One problem with multiple frames is that persons may be listed on more than one frame. For example, a person in the Immigrants frame may have been in Canada on a work permit in May 1996 and thus would have been enumerable in the 1996 Census. That person would then be in the Immigrants frame and in the Census frame if he or she were enumerated, or represented by the Missed frame if he or she were not enumerated. It is important to identify all potential cases of frame overlap; if this is not done, estimates could be too high because people have been double-counted. Wherever possible, such cases are identified when the frames are constructed, but they can also be identified later on the basis of information provided by respondents.

Another problem is that none of the first five sample frames (see Table 8.2) cover persons who had emigrated or were out of the country at the time of the 1996 Census and returned during the intercensal period ("returning Canadians"). It is estimated that there are 192,310 such persons for whom the number of "missed" cases has not been estimated. The estimates of coverage error do not cover this group. In the 2001 RRC, however, unlike the 1996 RRC, persons who returned to a province after living in a territory (17,150) or an Indian reserve or settlement that was incompletely enumerated in 1996 (15,110) and enumerated in 2001 were considered "returning Canadians".

8.3 Methodology – Data Collection

8.3.1 Data Collection Environment

Head office (HO) staff in Ottawa worked closely with staff in five of the Statistics Canada regional offices (ROs) to collect data during the survey phase of the 2001 Reverse Record Check (RRC). These ROs were located in Halifax, Sherbrooke, Toronto, Winnipeg and Vancouver. The assignment of the sample cases to ROs was based on HO's "best guess" about where the selected person (SP) was residing during the collection period. Once a case was assigned to an RO, it was never transferred to another RO even if it was determined that the SP moved outside the RO collection area. Some RO workload statistics are shown in Table 8.3.1.

The main survey collection method was the computer-assisted telephone interviewing (CATI). The CATI application was developed using many of the standards set for all CATI surveys conducted at Statistics Canada. The application is constructed of various interrelated modules and was accessed through the Generic Interface for Regional Offices. This interface allowed senior RO staff to assign interviewers to specific types of cases, and to produce frequent progress reports.

If respondents were unwilling to conduct an interview by telephone, they received a paper version of the 2001 RRC Survey questionnaire. SPs then returned the paper questionnaire and an interviewer keyed the information into the CATI system.

The content of the RRC Survey questionnaire focused on the collection of addresses where the SP possibly lived on Census Day and in the month of May 2001. Names and demographic data were collected for all household members of the SP. Some SPs were asked additional questions related to addresses if they were students or workers studying or working away from their usual home in May 2001. A special set of questions relating to the subject of "brain-drain" was asked of a small portion of the RRC

Survey sample. Persons in the sample who were deceased at the time of the RRC Survey were included and a proxy respondent was contacted to conduct the survey. Proxy respondents were also used when the SP was under 15 years of age or when the selected adult person was not available during the survey period or was difficult to reach. The average duration of the CATI interview was 12 minutes.

Table 8.3.1 Regional Office Data Collection Workloads, 2001 Reverse Record Check Survey

Regional office	Field sample size	Interviewer hours	Completion rate (%)	Coverage areas
Halifax	12,294	18,113	95.0	Newfoundland and Labrador, Prince Edward Island, Nova Scotia, New Brunswick
Sherbrooke	9,921	16,134	83.2	Quebec, Nunavut
Toronto	11,367	15,448	90.1	Ontario
Winnipeg	15,272	16,693	91.5	Manitoba, Saskatchewan, Alberta, Northwest Territories
Vancouver	9,988	12,017	87.3	British Columbia, Yukon
Head office (not sent to a regional office)	1,811	…	…	Various
Total	**60,653**	**78,405**	**89.8**	**Canada**

… not applicable

8.3.2 Collection Operations

The computer-assisted telephone interviewing input data were loaded in separate waves based on when batch sample preparation was completed. Table 8.3.2A shows the distribution over time of the 58,842 field cases. Interviewing typically began in the regional office (RO) as soon as the new cases arrived and was attempted in all cases except 353 cases belonging to the Nunavut sample. Because the necessary frame information from Nunavut was received later than expected, the 353 cases were included in the Reverse Record Check Follow-up Survey at a later date using a paper questionnaire (see Section 8.3.5).

Selected persons (SPs) residing in the territories received letters asking them to call an RO for an interview before interviewers began the work of tracing and calling.

The 58,842 cases sent to the field represent 97.1% of the number of cases on the Sample Control File. The 1,811 cases that were not sent to the field were mainly SPs from the 1996 Census frame who were believed to be deceased (N = 1,068). As well, about 700 SPs in the Birth 2000 and 2001 sample were not available in time for field collection and were processed using head office (HO) resources. Thirty-eight cases in the Immigrants frame were deemed "out of scope" and not considered legitimate field cases as a result.

Survey data were sent electronically to HO from the five ROs each night after interviewing came to a halt. Transmission reports and collected survey data were reviewed each morning by HO staff. Cases considered unsuitable for processing were reactivated and sent back to the RO for field collection.

To increase response rates, RO managers contacted government agencies and departments in an attempt to obtain address and telephone number information for cases where contact had not been

established. HO also provided contact leads obtained from various administrative files to the ROs during the data collection process. Sustained efforts to interview persons who initially refused to participate in the survey also improved response rates.

Table 8.3.2A Commencement of the Field Collection of the 2001 Reverse Record Check Survey, by Start Date, Frame and Number of Cases

Start date	Frames	Number of cases	Number of interviewing days remaining
October 9, 2001	Census: all cases	43,863	193
	Birth: 1996-1998	1,554	
	Immigrant: 1996-2000	2,075	
	Missed: all cases	2,190	
November 28, 2001	Birth: 1999	554	145
January 8–14, 2002	Yukon and Northwest Territories cases	5,170	115-107
February 7, 2002	Immigrant: 2001	215	86
	Non-permanent residents: all cases	1,057	
	Immigrant: misc. (1996-2000)	15	
February 27, 2002	Head office methodology cases (poor quality CATI[1] input data available)	286	67
March 11, 2002	Nunavut: all cases	1,835	51
March 13, 2002	Head office methodology cases – part II	28	49
April 30–May 10, 2002	Data collection completed in the regional offices		

1. Computer-assisted telephone interviewing

Table 8.3.2B Number of Interviews Completed, by Month and Year, 2001 Reverse Record Check Survey

Month–Year	Oct–01	Nov–01	Dec–01	Jan–02	Feb–02	Mar–02	Apr–02	May–02	Total
Number of interviews completed	14,252	13,140	5,633	6,257	5,323	4,279	3,015	964	52,863

Two detailed management reports were created at HO to document the progress of the survey. One report gave statistics on the cases currently in the RO (unopened cases, completed cases, and opened cases not yet completed). The second report presented a summary of the number of completed interviews waiting for HO processing (see Section 8.3.4). This summary also included case completion projections for the ROs in order to help them meet their survey collection targets.

Data collected in the field were analyzed at HO for completeness and accuracy. Cases were rejected if data were insufficient, or if the data had mistakenly been gathered for someone other than the SP. Cases which were not rejected were compiled into batches, which were then passed on to the HO Reverse Record Check processing team.

8.3.3 Tracing

Tracing was a key aspect of the 2001 Reverse Record Check Survey. Tracing refers to the work done, mostly by the staff of regional offices (ROs) ,to find telephone and address contact leads for selected persons (SPs). Table 8.3.3 shows that 59.8% (51.9% + 7.9%) of all cases in the RRC Survey required tracing. They were cases where the SP or acceptable proxy person could not be reached with the original computer-assisted telephone interviewing (CATI) contact information. Table 8.3.3 shows that there were 19,247 cases where no CATI contact telephone number was given and where tracing was performed as a first step. In 39,595 cases (23,643 + 11,311 + 4,641), a contact telephone number was provided at the start of CATI. Of these, 11,311 were not contacted and subsequently went into tracing.

Table 8.3.3 Overview of Tracing Statistics, 2001 Reverse Record Check Survey

	Number	Percent of total
Cases sent to the regional offices	**58,842**	**100.0**
Cases not requiring use of the tracing application	23,643	40.2
Cases requiring use of the tracing application	30,558	51.9
Started in tracing	19,247	32.7
Required tracing	11,311	19.2
Cases remaining in tracing at Survey end	4,641	7.9

Interviewers used a variety of tracing tools, on-line electronic directories being the most popular. However, the most effective tracing leads came from the CATI application itself. Head office also supplied telephone and address information to the ROs as it became available.

8.3.4 Collection Statistics

Table 8.3.4 shows provincial level and territorial level completion rates for the 2001 Reverse Record Check Survey. The completion rate for Nunavut could be taken to be 66.0%—rather than 45.1%—if the Reverse Record Check Follow-up Survey cases are included. This would then bring the national level completion rate to 90.5%.

Table 8.3.4 Sample Sizes and Completion Statistics, Canada, Provinces and Territories, 2001 Reverse Record Check Survey

	Total sample size	Number of cases sent for field collection	Number of cases completed in the field	Completion rate (%)
Canada	**60,615[1]**	**58,842**	**52,863**	**89.8**
Newfoundland and Labrador	3,574	3,449	3,336	96.7
Prince Edward Island	2,809	2,716	2,606	95.9
Nova Scotia	3,501	3,388	3,186	94.0
New Brunswick	3,398	3,267	3,027	92.7
Quebec	8,370	8,115	7,470	92.1
Ontario	11,268	10,904	9,795	89.8
Manitoba	3,671	3,525	3,232	91.7
Saskatchewan	3,762	3,622	3,434	94.8
Alberta	5,044	4,885	4,345	88.9
British Columbia	8,213	7,966	6,965	87.4
Yukon	2,065	2,065	1,780	86.2
Northwest Territories	3,105	3,105	2,859	92.1
Nunavut	1,835	1,835	828	45.1

1. The effective Reverse Record Check Survey sample size was 60,615. An additional 38 cases deemed "out of scope" were selected for the Reverse Record Check sample. Therefore, the actual Reverse Record Check sample size was 60,653.

8.3.5 2001 Reverse Record Check Follow-up Survey

The main purpose of the 2001 Reverse Record Check Follow-up Survey (FOS) was to conduct follow-up or a second interview for a subset of selected persons (SPs) from the Reverse Record Check (RRC) Survey. Primarily, the FOS collected additional addresses from the selected persons who, following RRC processing, were not found to be listed on a 2001 Census questionnaire. These cases were referred to as "not found" cases. Secondly, the FOS received a number of cases, referred to as "unconfirmed death" cases, for confirmation of an SP's date of death. In addition, the FOS was responsible for conducting first-time interviews with a subset of SPs from Nunavut, cases referred to as "Nunavut" cases. Table 8.3.5.1 presents an overview of the main survey activities.

Table 8.3.5.1 2001 Reverse Record Check Follow-up Survey Schedule

Time frame / Dates	Reverse Record Check Follow-up Survey collection activity
Mar. 25–May 17, 2002	FOS[1] planning and systems development
May 10, 2002	Data collection begins for "unconfirmed death" and "not found" cases (39% of total FOS[1] sample)
May 28–July 5, 2002	Data collection begins for remaining "not found" cases and Nunavut cases
May 28–Aug. 15, 2002	Data capture by head office
July 31, 2002	Data collection completed
Aug. 13, 2002	Data file ready for head office processing
Sept. 25–Dec. 2002	Survey debriefing, documentation, analyses, federal–provincial–territorial workshop, etc.

1. Follow-up Survey

The 2001 FOS was a paper-and-pencil survey administered by telephone. The Sherbrooke Regional Office was responsible for conducting the interviews. Here, address leads obtained by automated matches between computer-assisted telephone interviewing data and Sample Control File input data to the 2001 Census database (see Section 8.4) were used for tracing purposes. In resolving any remaining cases, staff consulted year 2000 taxation file addresses for Nunavut and a file of Selected Person Identification Numbers, sorted by municipality name, for persons selected from the Yukon Territory. Completed cases, along with a progress report of data collection operations, were returned to head office by Sherbrooke staff on a weekly basis. Data capture of the completed questionnaires followed and a Statistical Analysis System program was used to verify the validity and consistency of responses. From these edited questionnaires, the RRC processing team created a cumulative text-format file to aid in the classification of SPs. Because each questionnaire was assessed individually when an edit failure occurred, verifying the validity and consistency of responses in preparing the file for the processing team was a time-consuming step.

Valuable feedback and recommendations for the 2006 FOS were obtained from two sources: the Interviewer Debriefing Report and the Regional Office Debriefing Session. A total of 3,122 hours, or approximately 416 person-days, were spent on the 2001 FOS in the Sherbrooke Regional Office. The FOS achieved a response rate of 81.4% (see Tables 8.3.5.2 and 8.3.5.3). In 178 (80.2%) of 222 "unconfirmed death" cases sent to the field, the SP's death was confirmed prior to May 15, 2001. In 39.5% of the "not found" cases (1,656 out of 4,192 cases), a potential new address lead was obtained for the SP. While it is true that the 2001 FOS response rate was lower than that of 1996 (87.3%), it is worth noting that the FOS met its primary objective of gathering additional information on the location of SP's on Census Day. A more detailed description of the 2001 FOS can be found in Armstrong et al. (2003).

Table 8.3.5.2 Completion Rates, Cases Sent and Completed by Case Type and Sampling Frame, 2001 Reverse Record Check Follow-up Survey

Cases sent and completed by case type	Sampling frame						
	Census	Birth	Immigrant	Missed	NPR[1]	Territories	Total
"Not found"							
Cases sent	2,667	108	230	316	234	637	4,192
Cases completed	2,258	91	184	248	139	535	3,455
Completion rate (%)	84.7	84.3	80.0	78.5	59.4	84.0	82.4
"Unconfirmed death"							
Cases sent	217	0	0	5	0	0	222
Cases completed	204	0	0	4	0	0	208
Completion rate (%)	94.0	…	…	80.0	…	…	93.7
"Nunavut"							
Cases sent	0	0	0	0	0	555	555
Cases completed	0	0	0	0	0	383	383
Completion rate (%)	…	…	…	…	…	69.0	69.0
Total cases sent	**2,884**	**108**	**230**	**321**	**234**	**1,192**	**4,969**
Total cases completed	**2,462**	**91**	**184**	**252**	**139**	**918**	**4,046**
Completion rate (%)	**85.4**	**84.3**	**80.0**	**78.5**	**59.4**	**77.0**	**81.4**

... not applicable

1. Non-permanent residents

Table 8.3.5.3 Completion Rates, Canada, Provinces and Territories, 2001 Reverse Record Check Follow-up Survey

	Cases sent	Cases completed	Completion rate (%)
Canada	**4,969**	**4,046**	**81.4**
Newfoundland and Labrador	244	212	86.9
Prince Edward Island	153	131	85.6
Nova Scotia	227	191	84.1
New Brunswick	203	178	87.7
Quebec	551	486	88.2
Ontario	898	725	80.7
Manitoba	225	181	80.4
Saskatchewan	251	210	83.7
Alberta	312	259	83.0
British Columbia	713	555	77.8
Yukon	178	136	76.4
Northwest Territories	377	326	86.5
Nunavut	637	456	71.6

8.4 Methodology – Address Processing and Selected Person Classification

The purpose of the Reverse Record Check's (RRC's) data processing phase is to classify—i.e., determine the status of—each selected person (SP). The results of this classification are then used in non-response adjustment and estimation (Section 8.5). The main objectives were to determine whether the SP belonged to the census target population and whether the SP was enumerated at the addresses collected, and to provide other information needed for non-response adjustment.

Hence, for each SP, the aim is to determine whether he/she can be classified as:

(a) a "listed" person;
(b) a "mobile" person;
(c) a person in the "census target population";
(d) an "enumerated" person.

For reference, an SP is "enumerated" if his/her name is listed in a census questionnaire (Form 2 or 3) and there is a corresponding record in the census database. The "census target population" includes the groups listed in Section 4.2 of this report. An SP is "mobile" if his/her address on Census Day is different from the addresses associated with him/her that have been obtained independently of RRC collection. Arbitrarily, out-of-scope persons are considered "mobile". When an SP can be classified independently of RRC collection, he/she is referred to as "listed". The "missed" classification covers persons in the target population who are not enumerated.

Enumerated non-mobile SPs are listed, and other SPs in the target population are unlisted.

SPs for whom one or more of characteristics (a) through (d) cannot be determined are considered non-respondents. More specifically, the SP is referred to as:

- not identified if it was impossible to determine whether he/she was listed;
- not contacted if it was impossible to determine whether he/she was mobile;
- not traced if it was impossible to determine whether he/she was in the census target population;
- not classified if it was impossible to determine whether he/she was enumerated.

The value of knowing whether an SP was enumerated is self-evident. We also want to be able to classify SPs by characteristics (a) through (c) so that we can best choose respondents who can represent non-respondents.

For the Census frame and the Missed frame:

- all not identified SPs are not traced;
- all not traced SPs are not contacted;
- all not contacted SPs are not classified;
- not classified SPs, in the restricted sense (i.e., not classified, but contacted), are in the target population and are mobile and unlisted;
- not contacted SPs, in the restricted sense (i.e., not contacted, but traced), are in the target population and are unlisted;
- not traced SPs, in the restricted sense (i.e., not traced, but identified), are unlisted.

For all other frames (excluding the Census and Missed frames):

- all not identified SPs are not contacted;
- almost all not contacted SPs are not traced (only seven not contacted SPs were traced);
- all not traced SPs are not classified;
- not classified SPs, in the restricted sense (i.e., not classified, but traced), are in the target population and are mobile and unlisted;
- not traced SPs, in the restricted sense (i.e., not traced, but contacted), are mobile and unlisted;
- not contacted SPs, in the restricted sense (i.e., not contacted, but identified), are unlisted.

Each in-sample unit classified as out-of-scope is assigned a reason for being out of scope: e.g., death, emigration or unit already represented by another sampling frame.

We also determine the address on Census Day of each classified SP in the target population, i.e., the address where, according to census instructions, the SP should be enumerated. If an SP in the target population is enumerated only once, the enumeration address is considered to be the SP's address on Census Day, even if other information raises doubts as to whether the instructions were followed. To measure overcoverage, we also count how many times classified SPs are enumerated, and we determine the addresses of the unwanted enumerations (double or triple enumerations, and enumerations of deceased persons or emigrants). To prevent double counting, we also check whether overcoverage cases identified in the RRC are in-scope for the other two overcoverage surveys, the Automated Match Study (AMS) and the Collective Dwelling Study (CDS). Cases identified as covered by one of these studies are excluded from the RRC and are not considered for the RRC overcoverage component estimates.

To produce that classification, we consult vital statistics registers (e.g., for a death), information obtained in the RRC interview and the Follow-up Survey (e.g., cases relating to emigration), sampling frames (e.g., cases where frames overlap) and other information sources such as the Internet. However, the bulk of the work lies in searching through census documents to determine whether the SP was enumerated at one of the addresses associated with him/her. The addresses associated with the SP come from various sources: the selection address in the sampling frame, addresses updated from tax records, addresses obtained in the RRC interview and the Follow-up Survey, and addresses obtained by matching with the Census frame by the birth date and sex of the SP and the members of his/her household.

Computer programs that exploit information such as the postal code of the address to be resolved can sometimes determine without outside assistance that one address duplicates another or (with a high degree of certainty) that the SP was enumerated at that address. In general, research staff do not see the addresses that are resolved automatically, just the ones that are not. In attempting to resolve one of the latter, they use electronic forms that often contain suggestions about which census questionnaires or enumeration areas may correspond to the address. These suggestions are provided by the above-mentioned computer programs. Sometimes the census questionnaire for the address to be resolved is found at one of those suggestions, but if not, other tools, such as electronic telephone books, are used. The search results are noted on forms, which are then edited electronically to minimize errors.

Subsequently, a file containing the search results for each address is produced. That file is used to identify cases requiring further verification, such as SPs found to have been enumerated at more than one address, SPs apparently omitted from a questionnaire, or non-response SPs. The file is also used to determine which units will be covered in the Follow-up Survey (see Section 8.3.5).

The distribution of the sample by SP classification and frame is presented in Table 8.4A. Of the 60,653 SPs initially in the sample, 3,164 (2,025 + 1,139) were missed by the 2001 Census, and 50,521 (46,705 + 3,816) were enumerated. To ensure that each SP is counted only once, "non-respondents" consist of not identified SPs, not traced SPs, not contacted SPs and not classified SPs, in the restricted sense. We note that 3,776 SPs were non-respondents (329 + 2,899 + 541 + 7), cases that had to be offset—in the estimation phase by non-response adjustment (see Section 8.5).

Table 8.4A Selected Person Classification by Sampling Frame, 2001 Reverse Record Check

	Census		Births		Immigrants		Missed		NPR[1]		Territories		Total	
Classification	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**
Enumerated, non-mobile (listed)	35,763	79.6	2,535	90.4	1,290	49.0	1,184	53.3	270	25.5	5,663	80.8	46,705	77.0
Enumerated, mobile (unlisted)	2,970	6.6	29	1.0	390	14.8	209	9.4	137	12.9	81	1.2	3,816	6.3
Missed, non-mobile (unlisted)	1,398	3.1	57	2.0	49	1.9	166	7.5	97	9.2	258	3.7	2,025	3.3
Missed, mobile (unlisted)	711	1.6	34	1.2	131	5.0	82	3.7	72	6.8	109	1.6	1,139	1.9
Deceased (mobile, listed)	1,782	4.0	8	0.3	3	0.1	55	2.5	0	0.0	22	0.3	1,870	3.1
Other out of scope, listed	67	0.1	0	0.0	226	8.6	18	0.8	51	4.8	138	2.0	500	0.8
Emigrated (mobile, unlisted)	271	0.6	6	0.2	21	0.8	78	3.5	51	4.8	2	0.0	429	0.7
Other out of scope, unlisted (mobile)	135	0.3	7	0.2	17	0.6	39	1.8	21	2.0	174	2.5	393	0.6
Not classified	76	0.2	4	0.1	4	0.2	16	0.7	21	2.0	208	3.0	329	0.5
Not traced	1,728	3.8	18	0.6	352	13.4	374	16.8	205	19.3	222	3.2	2,899	4.8
Not contacted	22	0.0	107	3.8	147	5.6	2	0.1	135	12.7	128	1.8	541	0.9
Not identified	7	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	7	0.0
Total	**44,930**	**100.0**	**2,805**	**100.0**	**2,630**	**100.0**	**2,223**	**100.0**	**1,060**	**100.0**	**7,005**	**100.0**	**60,653**	**100.0**

1. Non-permanent residents

The RRC's overcoverage results are shown in Table 8.4B. The total number of cases identified by the RRC can be divided into three categories: cases in-scope for the AMS, cases in-scope for the CDS, and cases that could only be identified by the RRC. Only the latter overcoverage cases are considered in the overall estimates of overcoverage. While the RRC measures the overcoverage already measured by the AMS and the CDS, the RRC's results are far less precise because of the small number of observations.

Table 8.4B Overcoverage Cases (Unweighted Results), Canada, Provinces and Territories, 2001 Reverse Record Check

	In scope for AMS[1]	In scope for CDS[2]	In scope for RRC[3] only	Total
Canada	**245**	**3**	**221**	**469**
Newfoundland and Labrador	18	0	10	28
Prince Edward Island	8	0	12	20
Nova Scotia	14	1	15	30
New Brunswick	6	1	11	18
Quebec	30	0	38	68
Ontario	62	0	42	104
Manitoba	17	1	10	28
Saskatchewan	12	0	17	29
Alberta	16	0	23	39
British Columbia	37	0	34	71
Yukon	2	0	4	6
Northwest Territories	17	0	5	22
Nunavut	6	0	0	6

1. Automated Match Study
2. Collective Dwelling Study
3. Reverse Record Check

For further information about the 2001 RRC classification, see Parenteau (2003).

8.5 Methodology – Estimation

The final weights of selected persons (SPs) are computed from the initial weights. The initial weight of an SP from the Missed frame is the final weight assigned to that person in the previous Reverse Record Check (RRC) (when the person was classified as missed). It should be noted that for the 2001 Census, the final weights in the 1996 RRC were recomputed using a methodology similar to the 2001 methodology. The Missed frame contains more SPs than in 1996. According to the new definition, the Missed frame includes persons who in 1996 were enumerated on a Form 4 (Missing or Incomplete Questionnaire Card) or were classified as non-respondents but were living in a dwelling for which a Form 4 was prepared. For SPs from other frames, the initial weights are usually the inverse of the probability of selection in sample, except for SPs from the Non-permanent residents frame. For the latter, the initial weight is set higher to account for non-permanent residents who were not in the frame when the sample was selected, which was the case for a small proportion of those residents. The exact counts for non-permanent residents were obtained after sample selection.

Then the initial weights have to be adjusted for non-response. The weight of non-respondents is redistributed among respondents. To the extent possible, the redistribution must ensure that the weight of

non-respondents with certain characteristics is redistributed among respondents with the same characteristics. The characteristics involved are the sample design stratum; an indication in the tax data that the SP can be included in the target population; and whether the SP is listed, mobile or a member of the target population. The redistribution, carried out by the Generalized Estimation System (GES), simulates a multi-phase sampling process in which each phase corresponds to the "selection" of a different category of SP: identified SP, contacted SP, traced SP and classified SP. For the Census and Missed frames, the set of contacted SPs forms a subset of the set of traced SPs; the opposite is true for the other frames. The order used by the GES varies with the sampling frame. Each phase includes one more characteristic than the previous phase. When a respondent with the same characteristics as a non-respondent cannot be found, the problem stratum is grouped with a similar stratum.

Traditionally, the unadjusted estimated number of persons enumerated in the territories has been smaller than the comparable census count; this appears to be due to undercoverage of the territories' sampling frame. To offset this undercoverage, the weights of the territories SPs are adjusted by a factor that makes the estimated number of persons enumerated equal to the comparable census count, for each of the three territories.

A unit that is respondent for the purposes of measuring undercoverage is not necessarily respondent for the purposes of measuring overcoverage. A unit can be classified as enumerated even if no RRC interview has been completed for it or even if not all the addresses associated with it have been resolved. Unresolved or unobtained addresses could be involved in overcoverage. As a result, two additional factors are used to inflate the variable of interest showing overcoverage. Further information on this topic is provided in Théberge (2002).

Table 8.5A contains weighted counts corresponding to the distribution of the sample. It should be noted that these figures reflect neither census count adjustments (see Chapter 7) nor overcoverage estimates, which affect the rate of net census undercoverage.

Table 8.5A Person Classification (Weighted Results), by Sampling Frame, 2001 Reverse Record Check

	Census		Birth		Immigrant		Missed		NPR[1]		Territories		Total	
Classification	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**
Enumerated, non-mobile (listed)	23,692,419	83.2	1,533,274	90.1	511,639	48.4	633,059	54.4	85,936	27.8	85,832	81.7	26,542,159	80.9
Enumerated, mobile (unlisted)	1,998,415	7.0	37,434	2.2	303,916	28.8	193,622	16.7	94,168	30.4	4,121	3.9	2,631,676	8.0
Missed, non-mobile (unlisted)	802,590	2.8	40,217	2.4	24,904	2.4	104,258	9.0	35,354	11.4	3,979	3.8	1,011,302	3.1
Missed, mobile (unlisted)	507,422	1.8	61,73	3.6	98,060	9.3	76,499	6.6	43,987	14.2	4,830	4.6	792,528	2.4
Deceased (mobile, listed)	1,054,204	3.7	3,912	0.2	1,142	0.1	43,217	3.7	0	0.0	348	0.3	1,102,823	3.4
Other out of scope listed	36,086	0.1	0	0.0	89,773	8.5	7,185	0.6	14,102	4.6	2,056	2.0	149,202	0.5
Emigrated (mobile, unlisted)	250,194	0.9	13,024	0.8	16,501	1.6	65,233	5.6	23,440	7.6	35	0.0	368,427	1.1
Other out of scope unlisted	123,857	0.4	11,526	0.7	10,488	1.0	39,796	3.4	12,333	4.0	3,885	3.7	201,885	0.6
Total	**28,465,187**	**100.0**	**1,701,117**	**100.0**	**1,056,423**	**100.0**	**1,162,869**	**100.0**	**309,320**	**100.0**	**105,086**	**100.0**	**32,800,002**	**100.0**

1. Non-permanent residents

Weighted RRC overcoverage results are presented in Table 8.5B.

Table 8.5B Overcoverage (Weighted Results), Canada, Province and Territories, 2001 Reverse Record Check

	In scope for AMS[1]	In scope for CDS[2]	In scope for RRC[3] only	RRC Total
Canada	**159,502**	**845**	**148,813**	**309,160**
Newfoundland and Labrador	3,844	0	1,615	5,459
Prince Edward Island	446	0	805	1,251
Nova Scotia	4,041	397	4,603	9,041
New Brunswick	1,588	58	3,908	5,554
Quebec	30,385	0	40,443	70,828
Ontario	76,137	0	49,543	125,68
Manitoba	5,652	390	5,050	11,092
Saskatchewan	3,232	0	6,885	10,117
Alberta	10,503	0	13,416	23,919
British Columbia	23,303	0	22,385	45,688
Yukon	36	0	79	115
Northwest Territories	248	0	81	329
Nunavut	87	0	0	87

1. Automated Match Study
2. Collective Dwelling Study
3. Reverse Record Check

With regard to computing the variance of the estimates, an important innovation in the 2001 RRC was the replacement of the random groups method with sample design methods, thanks to the GES. Under these methods, the RRC sample design is approximated by a stratified design with probabilities of selection proportional to size. The sizes selected are those which reproduce the final weights.

Further details on the estimation methods used in the 2001 RRC are available in Théberge and Liu (2003).

8.6 Methodology – Tabulation

In 2001, as in 1996, a series of tables was created and presented to the Federal–Provincial–Territorial Committee on Demography.

The estimates of coverage error are presented in four distinct series, as follows:

- Series A: Population coverage error by province and territory and demographic breakdown;
- Series B: Statistics relating to the classification of selected persons;
- Series C: Comparison of Reverse Record Check (RRC) estimates of the number of persons classified as enumerated to census estimates; and
- Series D: Comparison of RRC estimates of the number of persons classified as enumerated to those of independent sources (other than census).

The statistics in Table 5.1 and Chapter 11 are taken from some of the Series A tables.

Estimates of coverage error were produced using Statistics Canada's Generalized Estimation System (GES) and a series of programs written in SAS (Statistical Analysis System). Producing estimates in parallel using these two methods ensured consistency in the derivation. Estimates of the standard errors were produced using both GES and SAS. Tables of coverage error estimates and standard errors were produced using SAS in conjunction with Microsoft Excel. This approach for creating tables greatly reduced the time required in 2001 compared to1996.

9. Automated Match Study

9.1 Methodology

Overcoverage can occur in two ways. The first, and most common, is when a unit is enumerated more than once. The second is when a unit which should not be enumerated is enumerated. The level of overcoverage in the 2001 Census was measured by three studies: the Automated Match Study (AMS), the Collective Dwelling Study and the Reverse Record Check.

The AMS was designed to detect and estimate overcoverage due to duplication between different private households by searching the census database for pairs of households containing persons that had the same sex and full date of birth characteristics and were located in the same geographic region (Atlantic, Quebec, Ontario, Western Canada, which includes British Columbia and the territories). Two sets of programs were put in place. The first set identified pairs of census households located in the same region with at least two exact person matches (322,298 matches). The second set identified pairs in the same federal electoral district (FED) with only one exact person match (819,481 matches). The pairs of households were then stratified. The census questionnaires for a sample of pairs from each stratum were compared manually to determine if overcoverage had occurred. That is, the list of persons on one questionnaire was compared to the persons listed on the other household questionnaire. Whenever the same persons were on both questionnaires, overcoverage was said to have occurred.

In 1991, the automated matching operation was limited to comparing households within the same enumeration area. In 1996, the AMS was expanded substantially to compare households within the same geographic region. In 2001, a new operation was added to find pairs of households in the National Capital Region, with one household in Quebec and the other in Ontario. This is one large urban area that crosses two different geographic regions where a lot of mobility occurs.

9.2 Sample Design and Processing

The 1,141,779 pairs of households detected by the matching programs constituted the sampling frame. Stratification was guided by the desire to produce strata that were both meaningful and of a reasonable size. The goal was to group pairs of households with similar probabilities of containing overcoverage together. The frame of pairs of households was stratified on the basis of province, geographical proximity, the number of exact and near matches, and the number of persons in each household. The geographical proximity is defined as follows.

- the two households were within the same enumeration area (EA);
- the two households were within the same federal electoral district (FED), but in different EAs;
- the two households were within the same province, but in different FEDs; and
- the two households were within the same region (Atlantic, Quebec, Ontario, or Western Canada, which includes British Columbia and the territories), but in different provinces, or in the National Capital Region with one household in Quebec and the other in Ontario.

Pairs containing households with the highest level of similarity were considered to contain overcoverage without further review, based on 1991 and 1996 census experience. A random sample was drawn from the remaining pairs of households. The electronic copies of the questionnaires for the selected pairs of households were then checked to see whether or not the same persons were listed on two questionnaires. From a total sample size of 17,275 pairs of households, 35,224 persons in 10,953 pairs of households were found listed on two different census questionnaires. These persons were weighted according to the sampling design. Table 9.2.1 shows the number of pairs of households in the frame and the final AMS estimates, while Table 9.2.2 gives statistics for the sample.

Table 9.2.1 Number of Household Pairs and Final Estimates, Canada, Provinces and Territories, 2001 Automated Match Study

	Estimates		
	Number of pairs formed	Number of pairs containing overcoverage	Number of persons overcovered
Canada	**1,141,779**	**62,876**	**146,412**
Newfoundland and Labrador	9,408	774	1,657
Prince Edward Island	1,765	198	439
Nova Scotia	21,128	1,398	2,875
New Brunswick	14,295	1,403	2,608
Quebec	303,807	15,829	35,061
Ontario	488,083	21,709	53,378
Manitoba	30,224	1,842	4,128
Saskatchewan	21,684	1,748	3,629
Alberta	100,344	5,868	13,263
British Columbia	149,486	11,792	28,710
Yukon	402	85	173
Northwest Territories	631	129	325
Nunavut	522	100	165

Table 9.2.2 Sample Size and Processing Results, Canada, Provinces and Territories, 2001 Automated Match Study

	Sample		
	Number of pairs selected	Number of pairs containing overcoverage	Number of persons overcovered
Canada	**17,275**	**10,953**	**35,224**
Newfoundland and Labrador	423	199	591
Prince Edward Island	188	97	268
Nova Scotia	673	328	954
New Brunswick	515	246	670
Quebec	3,640	2,355	7,014
Ontario	5,616	3,984	13,159
Manitoba	681	281	885
Saskatchewan	568	234	698
Alberta	1,527	783	2,441
British Columbia	3,165	2,300	8,116
Yukon	76	41	106
Northwest Territories	118	71	227
Nunavut	85	34	95

10. Collective Dwelling Study

10.1 Methodology

This study covered all persons who were enumerated as usual residents (URs) in a non-institutional collective dwelling and measured overcoverage, which occurs when a respondent completes two census questionnaires at different dwellings. A collective dwelling is a dwelling of a commercial, institutional, or communal nature. Collective dwellings are of two types: institutional dwellings such as hospitals and jails, and non-institutional dwellings such as hotels, YMCAs, and school residences. Contrary to 1996, the 2001 Collective Dwelling Study (CDS) did not produce estimates for institutional collective dwellings. The source of overcoverage for these is now covered by the Reverse Record Check.

During the enumeration of collective dwellings on May 15, 2001, URs of non-institutional collective dwellings completed an individual census form on which they were asked to report an alternate address where they could have been enumerated. In the 1996 CDS, stratification was based solely on province or territory, while in the 2001 CDS, it also included dwelling type.

Each province and territory was assigned a sample whose size was proportional to the square root of its UR population, with a minimum of 75 persons per province and territory. This provided good precision for every province and territory without a serious loss of precision at the Canada level. Then a sample was allocated by dwelling type within each province and territory in proportion to its UR population, with a minimum of two persons per dwelling type. A sample of 4,500 URs was selected. The useful sample was 4,027 URs, as 473 were excluded because of insufficient demographic information.

10.2 Processing

The data were processed through steps similar to those applied to the Reverse Record Check. For each selected person (SP) in the sample, the questionnaires completed at the alternate address were verified to determine if the SP was enumerated a second time. If the SP was found on the questionnaire, there was overcoverage. Of all SPs, 347 provided an alternate address and 148 were also found enumerated at a private dwelling.

Initial weights, consisting of the inverse of the probability of selection, were adjusted to account for the 473 URs in the sample who were not considered because insufficient demographic information was available. Basically, the weights of the non-respondents were distributed among the respondents.

Table 10.2 shows the distribution of the final sample, the number of cases of overcoverage detected by the study and the weighted number of persons found overcovered.

Table 10.2 Sample Size, Number of Cases of Overcoverage and Weighted Number of Overcovered Persons, Canada, Provinces and Territories, 2001 Collective Dwelling Study

	Non-institutional collective dwellings		
	Number of persons in sample	Cases of overcoverage	Weighted number of persons found overcovered
Canada	**4,500**	**148**	**2,633**
Newfoundland and Labrador	155	1	31
Prince Edward Island	99	1	11
Nova Scotia	245	5	83
New Brunswick	257	37	141
Quebec	925	11	575
Ontario	946	27	851
Manitoba	350	5	79
Saskatchewan	205	12	58
Alberta	511	23	405
British Columbia	549	19	392
Yukon	107	2	7
Northwest Territories	76	1	0
Nunavut	75	4	0

11. Coverage Error Estimates

11.1 Methodology

Let

$\hat{M}$ be the estimate of the number of persons not enumerated at their usual place of residence, as obtained from the 2001 Reverse Record Check (RRC) by summing the adjusted weights of those persons in the RRC sample with a final classification of "not enumerated in the 2001 Census";

C be the published 2001 Census count;

$\hat{D}$ be the estimate of the number of persons living in dwellings classified by the Census Representative (CR) as "unoccupied" and the estimate of the number of persons living in dwellings classified as "non-response" by the CR, as obtained from the Dwelling Classification Study;

$\hat{O}_{AMS}$ be the estimate of the number of overcovered persons measured by the Automated Match Study;

$\hat{O}_{CDS}$ be the estimate of the number of overcovered persons measured by the Collective Dwelling Study;

$\hat{O}_{RRC}$ be the estimate of the number of overcovered persons measured by the RRC, net of the overcoverage included in the Automated Match Study and the Collective Dwelling Study; and

$\hat{O}$ be the estimate of the number of persons enumerated more than once or in error where

$$\hat{O} = \hat{O}_{AMS} + \hat{O}_{CDS} + \hat{O}_{RRC}$$

Note that C includes $\hat{D}$. That is, as described in Chapter 7, a process was carried out whereby an estimate of the number of persons not enumerated because they occupied dwellings classified by the CR as unoccupied was added to the count of persons actually enumerated to produce the published census counts.

The population undercoverage rate $\hat{R}_U$ indicating the proportion of persons missed by the 2001 Census as a proportion of the total number of persons who should have been enumerated was calculated as follows:

$$\hat{R}_U = \frac{\hat{M} - \hat{D}}{C + (\hat{M} - \hat{D}) - \hat{O}}$$

The number of persons who should have been enumerated is the total of the number of persons actually enumerated, $C - \hat{D}$, plus the number of persons who should have been enumerated but were not, $\hat{M}$, less the number of persons enumerated who should not have been enumerated, $\hat{O}$. $\hat{D}$ is subtracted from C because the published census count includes an estimate of the number of persons who were erroneously excluded because they occupied dwellings misclassified as unoccupied. The undercoverage rate represents the rate of persons not included in the official census counts that already include an adjustment for some categories of missed persons. In calculating, $\hat{R}_U$, it is necessary to subtract $\hat{D}$ from $\hat{M}$ because these persons are already included in as these persons are already included in C and $\hat{M}$ is an estimate of the total number of persons not enumerated at their usual place of residence.

The population overcoverage rate $\hat{R}_O$ was calculated as follows:

$$\hat{R}_O = \frac{\hat{O}}{C + (\hat{M} - \hat{D}) - \hat{O}}$$

The population net undercoverage rate $\hat{R}_N$ was calculated as follows:

$$\hat{R}_N = \hat{R}_U - \hat{R}_O = \frac{(\hat{M} - \hat{D}) - \hat{O}}{C + (\hat{M} - \hat{D}\ ) - \hat{O}}$$

11.2 Results

Tables 11.2.1, 11.2.2, and 11.2.3 give the main estimates from the coverage studies. Table 11.2.1 presents the number of persons missed, the undercoverage rate, the number of overcovered persons, the overcoverage rate, the net number of persons missed, and the net undercoverage rate for Canada and for the provinces and territories. Table 11.2.2 presents the same data by gender and age group. Table 11.2.3 presents the number of persons overcovered by province and territory of selection according to the source study.

11.2.1 Undercoverage (gross)

The rate of population gross undercoverage was highest for the Northwest Territories (9.10%). Provincially, the rate was highest for British Columbia (5.30%). Gross undercoverage was less than the national rate (3.95%) for all provinces east of Ontario whereas the rate for Ontario (4.56%) was slightly higher. There was greater variation in undercoverage between the gender and age groups. Gross undercoverage was higher for men (4.90%) than for women (3.02%), with the highest rates being for young adults. The rates were strikingly high for young persons aged 20 to 24 (9.85%), with 11.68% being the rate for males and 7.91%, the rate for females in this age group.

11.2.2 Overcoverage

Table 11.2.3 gives the contribution by province and territory of each overcoverage study to the estimate of total population overcoverage. The Automated Match Study (AMS) measured overcoverage from persons enumerated in two or more households in the same region (Atlantic, Quebec, Ontario, Western Canada, which includes British Columbia and the territories) and single-person households within the same federal electoral district. The estimate of this type of overcoverage is 146,412 persons. The Collective Dwelling Study (CDS) measured overcoverage from persons enumerated in a collective dwelling and a private dwelling. The estimate of this type of overcoverage is 2,633 persons. Although the Reverse Record Check (RRC) measures overcoverage from all sources by asking selected persons for all addresses where they could have been enumerated, only the overcoverage not measured by the AMS and the CDS contributes to the estimate of the total number of overcovered persons. Generally speaking, the RRC was used to measure the following types of overcoverage:

- single persons enumerated in different dwellings;
- persons enumerated in different dwellings but who reported different characteristics (i.e. sex and date of birth); and
- persons who were enumerated in different regions.

The estimate of this type of overcoverage is 148,811 persons.

According to Table 11.2.1, the national level coefficient of variation (CV) of the estimate of total overcoverage rose to 4.9% (the coefficient of variation is the standard error of an estimate divided by the estimate itself, expressed as a percentage). The AMS is the most precise of the three studies with a CV of 2.3%, while those of the RRC and the CDS are higher at 9.5% and 9.4% respectively. At the provincial and territorial level, the CVs for the AMS vary between 4.7% and 16.7%, while the CVs of the two other studies are higher. Overall the CVs of the estimates of total overcoverage vary from 9.4% to 21.6% in the provinces and territories.

The rate of population overcoverage was highest for British Columbia (1.26%). Overcoverage was higher for Saskatchewan (1.06%), Quebec (1.03%) and the Northwest Territories (1.00%) than for Canada. Overcoverage was lower for Nunavut (0.59%), Newfoundland and Labrador (0.63%), Manitoba (0.80%) and Nova Scotia (0.81%). Again, there was more variation between the gender and age groups. As for undercoverage, overcoverage was highest for young adults aged 20 to 24, also reflecting the higher degree of residence mobility. Unlike undercoverage, however, overcoverage was generally higher for females than for males. Apart from those aged 20 to 24, overcoverage was concentrated in children and youths aged 5 to 19 for both sexes. This phenomenon reflected the situation of children and youths with parents who did not reside in the same household as well as those who were enumerated more than once because their families moved shortly before or after Census Day.

11.2.3 Undercoverage (net)

The net effect of undercoverage and overcoverage is given by the net undercoverage rates. That is, missed persons are offset by those who were enumerated in error. Provincially, net undercoverage increased from east to west. The rate was highest for British Columbia (4.04%). In Ontario, net undercoverage was higher (3.68%) than the national rate (2.99%), whereas in Quebec, a high overcoverage rate (1.03%) offset an undercoverage rate (2.93%), resulting in a net undercoverage rate (1.90%) that was smaller than the national rate. For the territories, net undercoverage rates are higher than the national rate. The rate was highest for the Northwest Territories (8.11%), followed by the Yukon Territory (4.73%) and Nunavut (4.49%).

Net undercoverage was notably high for young adults aged 20 to 34, and even higher for males in this age group. In the younger group (20 to 24 years), a high overcoverage offset a very high undercoverage, whereas in the older group (25 to 34 years), an overcoverage rate close to the national rate only slightly offset a high undercoverage rate. As for undercoverage, net undercoverage was consistently higher for males compared to females.

Table 11.2.1 2001 Census Population Coverage Error Estimated Numbers and Rates, and Standard Errors, Canada, Provinces and Territories, 2001 Reverse Record Check

	Population undercoverage				Population overcoverage				Population net undercoverage			
	Estimated number	Standard error	Estimated rate (%)	Standard error (%)	Estimated number	Standard error	Estimated rate (%)	Standard error (%)	Estimated number	Standard error	Estimated rate (%)	Standard error (%)
Canada	**1,222,286**	**42,045**	**3.95**	**0.13**	**297,857**	**14,525**	**0.96**	**0.05**	**924,429**	**44,749**	**2.99**	**0.14**
Newfoundland and Labrador	12,704	1,689	2.43	0.32	3,303	544	0.63	0.10	9,401	1,782	1.80	0.33
Prince Edward Island	2,580	734	1.89	0.53	1,255	245	0.92	0.18	1,325	775	0.97	0.56
Nova Scotia	32,082	3,955	3.44	0.41	7,561	1,251	0.81	0.14	24,521	4,170	2.63	0.44
New Brunswick	26,751	3,229	3.57	0.42	6,656	1,436	0.89	0.19	20,095	3,555	2.68	0.46
Quebec	216,312	19,664	2.93	0.26	76,080	7,118	1.03	0.10	140,232	21,033	1.90	0.28
Ontario	540,120	31,601	4.56	0.25	103,771	10,466	0.88	0.09	436,349	33,472	3.68	0.27
Manitoba	40,160	5,115	3.49	0.43	9,257	1,711	0.80	0.15	30,903	5,423	2.69	0.46
Saskatchewan	31,803	3,840	3.18	0.37	10,572	1,954	1.06	0.20	21,231	4,333	2.12	0.42
Alberta	96,941	10,505	3.18	0.33	27,084	4,067	0.89	0.13	69,857	11,308	2.29	0.36
British Columbia	216,029	14,762	5.30	0.34	51,487	4,834	1.26	0.12	164,542	15,598	4.04	0.37
Yukon	1,683	369	5.59	1.16	260	48	0.86	0.16	1,423	372	4.73	1.18
Northwest Territories	3,700	360	9.10	0.80	405	44	1.00	0.11	3,295	362	8.11	0.82
Nunavut	1,421	410	5.07	1.39	165	27	0.59	0.10	1,256	411	4.49	1.40

Table 11.2.2 2001 Census Population Coverage Error Estimated Numbers and Rates, and Standard Errors, by Age Group and Sex, 2001 Reverse Record Check

Characteristics	Population undercoverage				Population overcoverage				Population net undercoverage			
	Estimated number	Standard error	Estimated rate (%)	Standard error (%)	Estimated number	Standard error	Estimated rate (%)	Standard error (%)	Estimated number	Standard error	Estimated rate (%)	Standard error (%)
Both Sexes	**1,222,286**	**42,045**	**3.95**	**0.13**	**297,857**	**14,525**	**0.96**	**0.05**	**924,429**	**44,749**	**2.99**	**0.14**
0–4 years	77,570	13,107	4.42	0.71	16,905	3,165	0.96	0.18	60,665	13,508	3.45	0.74
5–14 years	118,449	15,826	2.90	0.38	61,947	6,050	1.52	0.15	56,502	16,992	1.38	0.41
15–19 years	91,818	11,724	4.36	0.53	38,873	5,371	1.85	0.26	52,945	12,913	2.51	0.60
20–24 years	207,509	14,544	9.85	0.62	56,066	6,558	2.66	0.32	151,443	16,120	7.19	0.71
25–34 years	347,270	16,804	8.07	0.36	39,662	3,866	0.92	0.09	307,608	17,288	7.15	0.37
35–44 years	213,732	18,350	4.04	0.33	25,950	3,227	0.49	0.06	187,782	18,640	3.55	0.34
45–54 years	80,287	13,063	1.79	0.29	17,276	1,776	0.39	0.04	63,011	13,185	1.41	0.29
55–64 years	35,243	10,935	1.22	0.37	10,963	1,540	0.38	0.05	24,280	11,044	0.84	0.38
65 years and over	50,408	13,603	1.29	0.34	30,214	8,058	0.77	0.21	20,194	15,833	0.52	0.40
Males	**750,245**	**31,331**	**4.90**	**0.19**	**141,014**	**8,681**	**0.92**	**0.06**	**609,231**	**32,634**	**3.98**	**0.20**
0–4 years	29,978	8,176	3.36	0.89	6,124	640	0.69	0.07	23,854	8,201	2.67	0.89
5–14 years	49,412	10,408	2.38	0.49	33,144	4,367	1.59	0.21	16,268	11,306	0.78	0.54
15–19 years	60,171	9,297	5.49	0.80	15,947	3,319	1.45	0.31	44,224	9,877	4.03	0.86
20–24 years	126,419	11,222	11.68	0.92	26,406	4,735	2.44	0.45	100,013	12,308	9.24	1.03
25–34 years	232,261	13,432	10.67	0.55	22,477	2,905	1.03	0.14	209,784	13,773	9.64	0.57
35–44 years	151,739	14,310	5.71	0.51	12,110	1,512	0.46	0.06	139,629	14,397	5.26	0.51
45–54 years	55,863	9,999	2.50	0.44	7,676	768	0.34	0.03	48,187	10,029	2.16	0.44
55–64 years	19,212	7,819	1.35	0.54	4,759	528	0.33	0.04	14,453	7,837	1.01	0.54
65 years and over	25,191	9,095	1.50	0.53	12,371	3,429	0.74	0.21	12,820	9,726	0.77	0.58

Characteristics	Population undercoverage				Population overcoverage				Population net undercoverage			
	Estimated number	Standard error	Estimated rate (%)	Standard error (%)	Estimated number	Standard error	Estimated rate (%)	Standard error (%)	Estimated number	Standard error	Estimated rate (%)	Standard error (%)
Females	**472,041**	**29,016**	**3.02**	**0.18**	**156,842**	**11,591**	**1.00**	**0.08**	**315,199**	**31,393**	**2.02**	**0.20**
0–4 years	47,592	10,421	5.50	1.14	10,781	3,100	1.25	0.36	36,811	10,889	4.26	1.21
5–14 years	69,037	12,046	3.44	0.58	28,803	4,063	1.44	0.21	40,234	12,730	2.01	0.62
15–19 years	31,647	7,188	3.13	0.69	22,925	4,217	2.27	0.43	8,722	8,348	0.86	0.82
20–24 years	81,090	9,346	7.91	0.84	29,660	4,535	2.89	0.46	51,430	10,484	5.02	0.97
25–34 years	115,009	10,240	5.41	0.46	17,186	2,504	0.81	0.12	97,823	10,574	4.60	0.47
35–44 years	61,993	11,573	2.35	0.43	13,84	2,782	0.53	0.11	48,153	11,903	1.83	0.44
45–54 years	24,424	8,422	1.09	0.37	9,600	1,513	0.43	0.07	14,824	8,559	0.66	0.38
55–64 years	16,030	7,649	1.09	0.52	6,204	1,384	0.42	0.09	9,826	7,775	0.67	0.53
65 years and over	25,218	10,134	1.13	0.45	17,843	7,269	0.80	0.33	7,375	12,487	0.33	0.56

Table 11.2.3 Contribution of the 2001 Automated Match Study, 2001 Collective Dwelling Study, and 2001 Reverse Record Check to Total Population Overcoverage for Canada, Provinces and Territories

	Estimated number of overcovered persons			
	Automated Match Study	**Collective Dwelling Study**	**Reverse Record Check *(exclusive of other studies)***	**Total**
Canada[1]	**146,412**	**2,633**	**148,811**	**297,857**
Newfoundland and Labrador	1,657	31	1,615	3,303
Prince Edward Island	439	11	805	1,255
Nova Scotia	2,875	83	4,603	7,561
New Brunswick	2,608	141	3,908	6,656
Quebec	35,061	575	40,443	76,080
Ontario	53,378	851	49,543	103,771
Manitoba	4,128	79	5,050	9,257
Saskatchewan	3,629	58	6,885	10,572
Alberta	13,263	405	13,416	27,084
British Columbia	28,710	392	22,385	51,487
Yukon	173	7	79	260
Northwest Territories	325	0	81	405
Nunavut	165	0	0	165

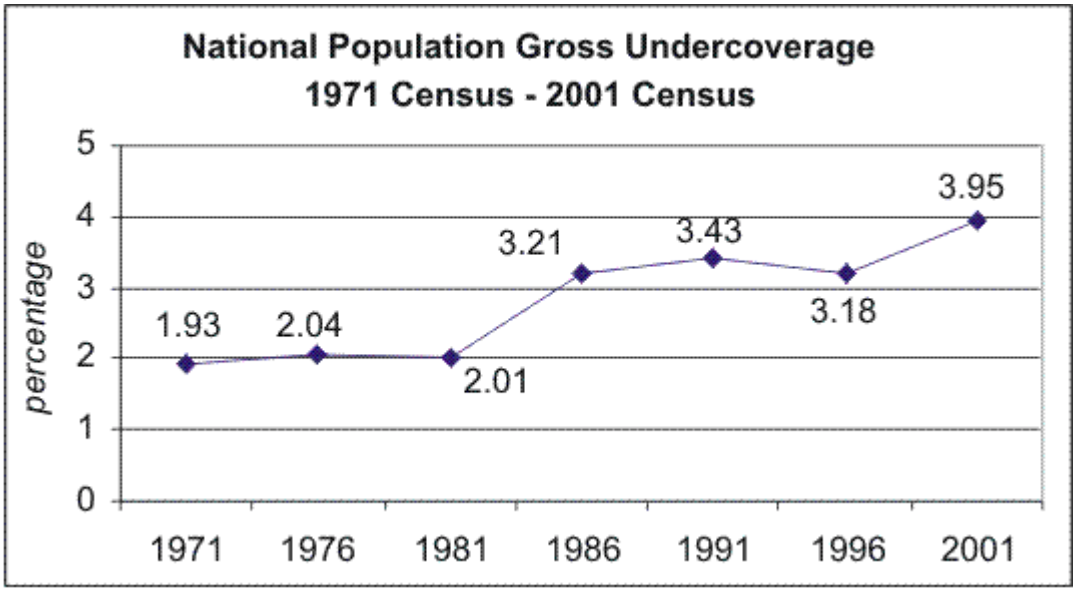
1. Rounding may account for small differences in the calculation of the Canada-level counts.

12. Historical Comparison

12.1 Comparisons with Published 2001 Census Counts

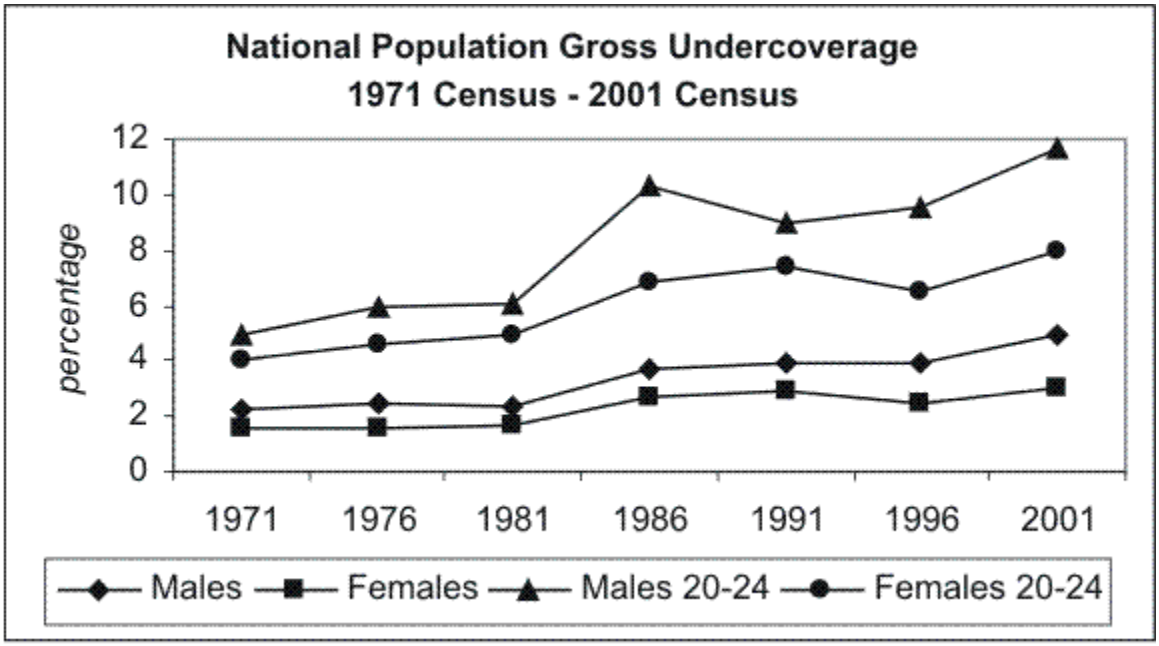
This section compares population gross undercoverage rates for 1971 to 2001. Estimates of net undercoverage are available only since 1991, because 1991 marked the first production of a comprehensive measure of overcoverage following an experimental study in 1986. Table 12.1A shows gross undercoverage rates for Canada and the provinces and territories from 1971 to 2001. Table 12.1B presents rates by age groups and by sex.

Figure 1. National Population Gross Undercoverage 1971 Census - 2001 Census



The national gross undercoverage rate was close to 2% for 1971, 1976, and 1981, but then rose to 3.21% in 1986 and to 3.95% in 2001. The latest increases are thought to be a reflection of both an increase in the construction of dwellings that are difficult to enumerate, such as renovated inner-city homes, and a change in the public mood towards government which reduced participation in the 2001 Census. As a result of the increase in 1986, coverage improvement initiatives were introduced for the 1991 Census. In particular, the use of the Address Register to provide a separate list of dwellings which should be enumerated helped to keep coverage error near the 1986 level. For the 1996 Census, the introduction of enumeration by a census representative, rather than by self-enumeration, in some large city enumeration areas served to minimize undercoverage. Also, moving Census Day from early June to mid-May helped to control undercoverage, because people were more likely to be at home and less likely to be moving. For the 2001 coverage studies, new automated systems were used to collect and process Reverse Record Check (RRC) data. While their use was based on improving efficiencies, these systems would not themselves account for increased rates of coverage error. It is felt however that they helped in achieving a more precise measurement of coverage error because of the number of additional addresses that were collected and processed in the 2001 RRC compared to previous RRC studies.

Figure 2. National Population Gross Undercoverage by Sex and Age Group, 1971 Census - 2001 Census



Several trends are seen from the data in Tables 12.1A and 12.1B.

(a) Among the provinces, British Columbia had the highest rate of gross undercoverage in every census from 1971 to 2001, excluding 1991 when Ontario had the highest rate.

(b) Gross undercoverage rates for the Atlantic and Prairie provinces tend to be lower than the national rate.

(c) There are two persistent demographic phenomena relating to gross undercoverage. The first is that the rate for males is higher than the rate for females in every census. Secondly, the highest rates of gross undercoverage are always for young adults in the 20-to-24 age group.

Minor differences in the design of the coverage studies over time mean that the rates in Tables 12.1A and 12.1B are not strictly comparable. Readers should note the following:

2001: [1]
(a) The institutional component of the Collective Dwelling Study was dropped and overcoverage estimates in this population were produced by the RRC.

[1] Estimates of the number of persons missed by the 1996 Census were produced to reflect the methodology of the 2001 RRC as closely as possible. The adjustments to the 1996 results include an allowance for changes in the non-response adjustment strategy and the estimation strategy (estimate for all persons not counted by the census after census count imputations are subtracted). The numbers given in this section do not reflect these revisions.

(b) As for 1996, the 2001 RRC did not estimate the number of persons missed in incompletely enumerated Indian reserves and Indian settlements. These reserves and settlements did however decrease in number in 2001. Nevertheless, the continued refusal of certain ones to participate in the census makes it increasingly difficult to estimate their population. Using a statistical model, it is estimated

that about 34,500 people lived on incompletely enumerated Indian reserves and Indian settlements (see Section 14).

(c) The RRC produced estimates for all persons who were not enumerated by the census, including estimates for the adjustments made to the census counts prior to publication. Those adjustments, based on the results of the Dwelling Classification Study, compensated for dwellings whose respondents could not be contacted (Form 4) and for occupied dwellings erroneously classified as unoccupied. The adjustments were subtracted from the census counts before population estimates for Canada and undercoverage and overcoverage rates were computed.

1996:
(a) The 1996 RRC did not estimate the persons missed on incompletely enumerated Indian reserves and Indian settlements. The continued refusal of some Indian reserves and settlements to participate in the census makes it increasingly difficult to select a representative sample. Using a statistical model, it is estimated that about 44,000 persons were missed in the 1996 Census on incompletely enumerated Indian reserves and Indian settlements.

(b) The Temporary Residents Study was cancelled for the 1996 Census because of concerns about the quality of the data, and because it was recognized that the RRC would measure this type of undercoverage with sufficient quality.

1991: [2] Non-permanent residents were included in the target population for the first time in 1991. This group had a higher than average gross undercoverage rate. Had this group not been included in the 1991 Census, it is estimated that the Canada-level rate of gross undercoverage would have been about 0.3 percentage points less. The gross undercoverage rates for Ontario, British Columbia and Quebec were particularly affected by the inclusion of non-permanent residents.

[2] Revised estimates of the number of persons missed in the 1991 Census have been produced as a result of analyzing the 1996 Census results. Revised estimates reflect the correction of some cases erroneously classified as missed, the removal of the estimate of persons missed in incompletely enumerated Indian reserves and Indian settlements, better estimates of the number of non-permanent residents, and a better measure of overcoverage based on the 1996 results. The numbers given in this section do not reflect these revisions. See Tourigny et al. (1998).

1986: The rates shown here for the 1986 Census differ from the results published in the *User's Guide to the Quality of 1986 Census Data: Coverage*. The rates shown in Tables 12.1A and 12.1B include revisions made after the 1986 publication when incompletely enumerated Indian reserves and Indian settlements were included as missed. In the original 1986 publication, they were included as "enumerated" since published provincial census counts included an estimate of persons missed on such reserves and settlements.

1976, 1971: Census counts for 1971 and 1976 did not include estimates from the Vacancy Check (VC) of persons missed in dwellings incorrectly classified as unoccupied. The 1981, 1986, 1991, and 1996 Census counts did include such a component. The 1976 population undercoverage rate would have been 1.78% had it included the results of the 1976 VC. There was no VC in the 1971 Census.

1971: For 1971, the age groups above age 24 differ from those shown for the other censuses.

Table 12.1A Estimated Population Undercoverage Rates and Standard Errors for Canada, Provinces and Territories, Reverse Record Check, 1971–2001[1]

	1971		1976		1981		1986		1991		1996		2001	
	Estimated rate (%)	S.E.[2] (%)	Estimated rate (%)	S.E.[2] (%)	Estimated rate (%)	S.E.[2] (%)	Estimated rate (%)	S.E.[2] (%)	Estimated rate (%)	S.E.[2] (%)	Estimated rate (%)	S.E.[2] (%)	Estimated rate (%)	S.E.[2] (%)
Canada	**1.93**	**0.09**	**2.04**	**0.10**	**2.01**	**0.09**	**3.21**	**0.13**	**3.43**	**0.12**	**3.18**	**0.09**	**3.95**	**0.13**
Newfoundland and Labrador[3]	2.25	0.72	1.10	0.39	1.74	0.45	1.92	0.33	2.47	0.30	2.45	0.29	2.43	0.32
Prince Edward Island	1.23	1.13	0.38	0.25	1.17	0.54	2.14	0.80	1.67	0.23	1.76	0.28	1.89	0.53
Nova Scotia	1.33	0.45	0.86	0.34	1.05	0.34	2.15	0.34	2.25	0.36	2.70	0.27	3.44	0.41
New Brunswick	1.65	0.56	2.16	-0.37	1.81	0.30	2.71	0.33	3.71	0.42	2.49	0.28	3.57	0.42
Quebec	2.10	0.19	2.95	0.25	1.91	0.21	2.91	0.31	3.18	0.20	2.46	0.18	2.93	0.26
Ontario	1.68	0.12	1.52	0.17	1.94	0.14	3.43	0.19	4.23	0.28	3.40	0.18	4.56	0.25
Manitoba	1.13	0.38	1.07	0.33	0.98	0.35	2.94	0.40	2.31	0.36	2.55	0.29	3.49	0.43
Saskatchewan	1.00	0.37	1.33	0.34	0.99	0.37	2.38	0.37	2.15	0.32	3.30	0.32	3.18	0.37
Alberta	2.55	0.44	1.49	0.26	2.54	0.36	3.00	0.32	2.51	0.27	2.99	0.24	3.18	0.33
British Columbia	2.89	0.39	3.13	0.31	3.16	0.33	4.48	0.36	3.42	0.24	4.58	0.24	5.30	0.34
Yukon	...	...	...	...	...	...	...	...	4.12	0.58	3.92	0.51	5.59	1.16
Northwest Territories[4]	...	...	...	...	...	...	...	...	5.73	0.57	5.68	0.47	9.10	0.80
Nunavut	...	...	...	...	...	...	...	...	...	...	...	...	5.07	1.39

... not available

1. Excludes incompletely enumerated Indian reserves and settlements. Includes non-permanent residents and territories in 1991, 1996 and 2001. Includes revisions to 1986 original publication. Excludes estimates of persons missed in dwellings incorrectly classified as unoccupied in 1971 and 1976.
2. Standard error
3. The official province name of Newfoundland was changed on December 6, 2001, to Newfoundland and Labrador.
4. Previous to the 2001 Census, Nunavut was included with the Northwest Territories.

Table 12.1B Estimated Population Undercoverage Rates and Standard Errors by Age Group and Sex, Reverse Record Check, 1971–2001[1]

Age group	1971[2]		1976		1981		1986		1991		1996		2001	
	Estimated rate (%)	S.E.[3] (%)	Estimated rate (%)	S.E.[3] (%)	Estimated rate (%)	S.E.[3] (%)	Estimated rate (%)	S.E.[3] (%)	Estimated rate (%)	S.E.[3] (%)	Estimated rate (%)	S.E.[3] (%)	Estimated rate (%)	S.E.[3] (%)
Both sexes, all ages	**1.93**	**0.09**	**2.04**	**0.10**	**2.01**	**0.09**	**3.21**	**0.13**	**3.43**	**0.12**	**3.18**	**0.09**	**3.95**	**0.13**
0–4 years	1.99	0.27	2.31	0.28	1.21	0.22	2.14	0.49	3.55	0.49	2.89	0.36	4.42	0.71
5–14 years	0.90	0.13	1.20	0.16	1.23	0.21	2.08	0.26	2.49	0.27	1.45	0.14	2.90	0.38
15–19 years	2.60	0.28	1.99	0.38	2.96	0.52	3.58	0.60	3.75	0.42	3.48	0.42	4.36	0.53
20–24 years	4.49	0.28	5.31	0.38	5.51	0.29	8.66	0.46	8.18	0.52	8.00	0.34	9.85	0.62
25–34 years	2.50	0.20	2.85	0.28	2.31	0.28	4.51	0.35	5.65	0.35	5.81	0.29	8.07	0.36
35–44 years	…	…	1.54	0.26	2.20	0.26	2.32	0.31	2.84	0.29	2.78	0.24	4.04	0.33
45–54 years	1.40	0.15	1.22	0.33	0.81	0.23	1.58	0.29	1.61	0.27	1.90	0.21	1.79	0.29
55–64 years	1.22	0.18	0.92	0.20	0.91	0.29	2.06	0.31	1.69	0.28	2.23	0.34	1.22	0.37
65 years and over	…	…	1.20	0.25	0.71	0.30	1.76	0.31	1.51	0.28	1.52	0.26	1.29	0.34
Males, all ages	**2.27**	**0.12**	**2.46**	**0.17**	**2.37**	**0.13**	**3.75**	**0.16**	**3.95**	**0.16**	**3.89**	**0.14**	**4.90**	**0.19**
0–4 years	1.73	0.34	2.53	0.46	1.32	0.33	2.22	0.67	2.79	0.58	2.56	0.47	3.36	0.89
5–14 years	0.93	0.18	1.14	0.21	1.27	0.29	1.98	0.32	2.32	0.34	1.46	0.24	2.38	0.49
15–19 years	2.71	0.39	1.93	0.48	3.12	0.68	4.09	0.74	3.55	0.60	3.68	0.43	5.49	0.80
20–24 years	4.97	0.40	5.99	0.52	6.03	0.48	10.36	0.57	8.98	0.81	9.48	0.50	11.68	0.92
25–34 years	3.38	0.31	3.64	0.46	2.70	0.44	5.43	0.45	7.28	0.56	7.74	0.42	10.67	0.55
35–44 years	…	…	2.33	0.48	3.42	0.40	3.29	0.51	3.65	0.41	3.94	0.39	5.71	0.51
45–54 years	1.90	0.24	1.63	0.41	1.21	0.38	1.95	0.52	2.05	0.45	2.12	0.27	2.50	0.44
55–64 years	1.37	0.28	1.28	0.34	0.91	0.40	1.88	0.47	2.04	0.44	2.50	0.54	1.35	0.54
65 years and	…	…	1.90	0.44	0.69	0.47	1.57	0.50	1.41	0.50	1.64	0.45	1.50	0.53

Age group	1971[2]		1976		1981		1986		1991		1996		2001	
	Estimated rate (%)	S.E.[3] (%)	Estimated rate (%)	S.E.[3] (%)	Estimated rate (%)	S.E.[3] (%)	Estimated rate (%)	S.E.[3] (%)	Estimated rate (%)	S.E.[3] (%)	Estimated rate (%)	S.E.[3] (%)	Estimated rate (%)	S.E.[3] (%)
over														
Females, all ages	**1.59**	**0.11**	**1.61**	**0.10**	**1.65**	**0.12**	**2.68**	**0.17**	**2.93**	**0.17**	**2.49**	**0.12**	**3.02**	**0.18**
0–4 years	2.25	0.40	2.07	0.36	1.10	0.33	2.06	0.62	4.35	0.71	3.24	0.55	5.50	1.14
5–14 years	0.87	0.17	1.26	0.27	1.19	0.31	2.20	0.33	2.65	0.39	1.45	0.22	3.44	0.58
15–19 years	2.49	0.38	2.05	0.51	2.80	0.73	3.05	0.76	3.96	0.54	3.28	0.55	3.13	0.69
20–24 years	4.01	0.37	4.62	0.48	4.98	0.43	6.89	0.72	7.36	0.71	6.45	0.48	7.91	0.84
25–34 years	1.58	0.22	2.03	0.38	1.92	0.32	3.59	0.45	3.98	0.37	3.84	0.40	5.41	0.46
35–44 years	…	…	0.72	0.24	0.93	0.31	1.33	0.32	2.01	0.35	1.62	0.28	2.35	0.43
45–54 years	0.90	0.17	0.81	0.38	0.41	0.26	1.20	0.35	1.16	0.34	1.68	0.33	1.09	0.37
55–64 years	1.10	0.24	0.58	0.25	0.92	0.34	2.23	0.50	1.35	0.33	1.97	0.40	1.09	0.52
65 years and over	…	…	0.64	0.38	0.71	0.42	1.89	0.44	1.58	0.36	1.43	0.32	1.13	0.45

... not applicable

1. Excludes incompletely enumerated Indian reserves and settlements. Includes non-permanent residents and territories in 1991, 1996 and 2001. Includes revisions to 1986 original publication. Excludes estimates of persons missed in dwellings incorrectly classified as unoccupied in 1971 and 1976.
2. Age groups for 24 years and over: 25–39, 40–59, 60 years and over.
3. Standard error

13. Evaluation of the Reverse Record Check

13.1 Introduction

The results of the largest coverage study, the Reverse Record Check (RRC), can be evaluated by comparing its estimates with data on the same characteristic from other sources such as the 2001 Census database. Comparisons with RRC estimates serve to evaluate RRC estimates and to quantify conceptual and measurement differences.

In spite of some conceptual differences between the RRC and the 2001 Census, the RRC estimate of persons enumerated in the 2001 Census can be compared with the count from the 2001 Census database. In order to render the two numbers comparable however, some adjustments were made to the census counts prior to comparing them.

Intercensal components of growth estimates can be compared with RRC estimates. In particular, the RRC estimate of persons who died between the 1996 Census and the 2001 Census can be compared with the count from vital statistics files.

Estimates of counts of net interprovincial migration from Canada Customs and Revenue Agency data can be compared with RRC estimates. It is not possible, however, to construct strict comparisons for this characteristic since reasonable adjustments for conceptual differences cannot be derived.

Finally, the RRC (and the census) can produce estimates of the components of population growth that can be compared with comparable components of population estimates derived from other sources. This publication contains figures for all provincial components of growth.

13.2 Comparisons With Published 2001 Census Counts

Since the Reverse Record Check (RRC) single-stage, stratified sampling design results in unbiased estimators, differences between RRC estimates and estimates from the census are due to sampling error on the part of the RRC estimates, conceptual differences between the two sources, and/or biases in the two sources which result in a systematic underestimation or overestimation of the studied characteristic.

13.2.1 Enumerated

The provincial and national comparisons are given in Table 13.2.1 along with the standard error of the Reverse Record Check (RRC) estimate and the *t*-value for testing the hypothesis that there is no difference between the RRC estimate and the comparable census figure. The following adjustments were made to published census counts to account for conceptual differences between the two sources:

- Adjustments from the Dwelling Classification Study were removed since they are included in the census database but are not part of the RRC estimate of enumerated persons.
- 2001 Census overcoverage is subtracted, as the census database contains overcoverage whereas the RRC estimate is based on the number of persons who are enumerated at least once.
- The census count of persons living outside Canada five years ago (excluding immigrants and non-permanent residents) is subtracted, as the RRC frame does not include these persons.
- Lastly, 1996 Census overcoverage is added since it is contained in the RRC estimate via the initial weights for the 1996 Census sampling frame which were not adjusted for this overcoverage.

Table 13.2.1 Comparison of 2001 Reverse Record Check Estimate and 2001 Census Count of Enumerated Persons, Canada, Provinces and Territories

	Reverse Record Check		Published census count	Comparable census figure[1]	Difference	*t*-value[2]
	Estimated number	Standard error				
Canada	**29,173,838**	**57,922**	**30,007,094**	**29,152,400**	**21,438**	**0.37**
Newfoundland and Labrador	510,247	5,712	512,930	507,577	2,670	0.47
Prince Edward Island	129,299	1,500	135,294	132,740	-3,441	-2.29
Nova Scotia	908,030	10,668	908,007	886,503	21,527	2.02
New Brunswick	706,494	6,588	729,498	715,450	-8,956	-1.36
Quebec	7,080,809	30,439	7,237,479	7,038,074	42,735	1.40
Ontario	11,042,521	47,884	11,410,046	11,074,981	-32,460	-0.68
Manitoba	1,092,493	10,440	1,119,583	1,102,426	-9,933	-0.95
Saskatchewan	963,583	9,644	978,933	958,663	4,920	0.51
Alberta	2,889,840	21,291	2,974,807	2,883,176	6,664	0.31
British Columbia	3,760,570	24,448	3,907,738	3,762,859	-2,289	-0.09
Yukon	27,599	0	28,674	27,599	0	…
Northwest Territories	36,487	0	37,360	36,487	0	…
Nunavut	25,866	0	26,745	25,866	0	…

... not applicable

1. The comparable census figure is calculated as: Comparable census figure = Published census count – Dwelling Classification Study – 2001 Overcoverage – Outside Canada in 1996 + 1996 Overcoverage.
2. A *t*-value either greater than 1.96 or less than -1.96 indicates the difference is significant at the 95% level.

Nationally, the RRC estimate of persons enumerated in the 2001 Census is slightly higher (0.07%) than the comparable 2001 Census figure. In 1996, the RRC underestimated the comparable census figure by 0.08% and in 1991, it did so by 0.46%. Provincially, the difference for Prince Edward Island is statistically significant at the 95% level (-2.29). The difference for Nova Scotia is just barely significant (2.02). All other gaps have a *t*-value less than 1.5 and in the case of six provinces, less than 1.0 (Newfoundland and Labrador, Ontario, Manitoba, Saskatchewan, Alberta and British Columbia).

The significant gaps are of some concern since they may indicate a bias in the RRC classification (including, for example, the province of residence on Census Day). However, other factors are important. Apart from sampling errors, biases in the adjustments applied to the published census figure to arrive at a conceptually equivalent figure may explain the gap. Also, the RRC non-response bias may affect this gap since the adjustment for non-response is chosen to give the best result for estimating missed persons rather than enumerated persons.

Prince Edward Island's rate is sensitive to interprovincial migration because persons selected outside the province have higher sampling weights than persons selected inside the province, and the presence or absence of interprovincial migration in the sample has a definite impact on the estimates. In Nova Scotia's case, a number of persons who were selected in other provinces and migrated to Nova Scotia were noted in the sample.

13.3 Comparison With Population Estimates

13.3.1 Deceased Persons

Table 13.3.1 compares the estimate of persons who died in the intercensal period according to the 2001 Reverse Record Check (RRC) province of classification with counts from Vital Statistics (VS). At the national level, the RRC overestimates the VS count by 16,764 (1.5%). The greatest relative gaps were for New Brunswick (3,811 vs. 30,388, a difference of 12.5%) and Saskatchewan (4,624 vs. 44,520, a difference of 10.4%). With respect to the t-value, the most significant differences were for New Brunswick (1.34) and for Ontario and Saskatchewan (1.15 respectively). All the estimates are well within the 95% confidence interval, and the results raise no concerns about the RRC.

Table 13.3.1 Comparison of Estimates of the Number of Deceased Persons from Census Day 1996 to Census Day 2001, by Province of Classification, 2001 Reverse Record Check and Vital Statistics

Province	Reverse Record Check			Vital statistics	Difference	*t*-value[1]
	Sample size	Estimated number	Standard error			
Provincial Total	**1,848**	**1,102,475**	**31,195**	**1,085,711**	**16,764**	**0.54**
Newfoundland and Labrador	134	21,794	1,936	21,068	726	0.37
Prince Edward Island	109	5,489	530	5,815	-326	-0.61
Nova Scotia	129	37,534	3,353	39,389	-1,855	-0.55
New Brunswick	147	34,199	2,852	30,388	3,811	1.34
Quebec	270	253,278	15,825	269,756	-16,478	-1.04
Ontario	367	429,043	22,957	402,707	26,336	1.15
Manitoba	139	48,092	4,255	48,690	-598	-0.14
Saskatchewan	155	49,144	4,032	44,520	4,624	1.15
Alberta	139	85,654	7,643	84,724	930	0.12
British Columbia	259	138,249	9,048	138,654	-405	-0.04

1. A *t*-value either greater than 1.96 or less than -1.96 indicates that the difference is significant at the 95% level.

13.3.2 Interprovincial Migration

Table 13.3.2 compares the Reverse Record Check (RRC) estimates of net interprovincial migration for the intercensal period with comparable figures derived from Canada Customs and Revenue Agency (CCRA) files. In general, interprovincial migration statistics are not comparable since the RRC only takes into account migration flows which occurred between the sampling frame reference date and Census Day 2001. Estimates derived from CCRA data, on the other hand, take account of annual migration. As a result, only net migration estimates are shown below.

The difference is significant for Prince Edward Island (t-value of -3.72). While the RRC estimate denotes a net migration loss, the estimate derived from CCRA data denotes a slight gain. For all other provinces, with the exception of Nova Scotia, both series of estimates show either a loss or a gain in net migration. For Nova Scotia, the RRC estimate shows positive net migration (5,047), while the figure derived from

CCRA administrative data shows negative net migration (-5,678). The distribution of the t-values for this characteristic does not suggest that the results derived from the RRC should be questioned.

Table 13.3.2 Comparison of 2001 Reverse Record Check Estimates of Net Interprovincial Migration[1] With Comparable Figures Derived from Canada Customs and Revenue Agency Data, by Provinces

Province	Reverse Record Check[1]			CCRA[2] figure	Difference	*t*-value[3]
	Sample size	Estimated number	Standard error			
Newfoundland and Labrador	384	-39,705	5,033	-32,502	-7,203	-1.43
Prince Edward Island	210	-4,368	1,268	344	-4,712	-3.72
Nova Scotia	420	5,047	9,255	-5,678	10,725	1.16
New Brunswick	280	-17,813	5,050	-8,658	-9,155	-1.81
Quebec	195	-50,247	12,956	-69,345	19,098	1.47
Ontario	787	58,954	20,146	65,317	-6,363	-0.32
Manitoba	283	-25,870	7,785	-20,696	-5,174	-0.66
Saskatchewan	374	-32,055	7,977	-24,720	-7,335	-0.92
Alberta	851	140,848	15,819	136,764	4,084	0.26
British Columbia	618	-35,802	15,339	-33,847	-1,955	-0.13

1. Excludes persons who lived in one of the three territories in 1996, but in a province in 2001.
2. Canada Customs and Revenue Agency
3. A *t*-value either greater than 1.96 or less than -1.96 indicates that the difference is significant at the 95% level.

13.4 Components of Population Growth

An extensive comparison of Reverse Record Check (RRC) estimates of the intercensal components of population growth and population estimates from administrative data sources was produced by the members of the Demography Division. The RRC estimates of the components of population growth are a by-product of the RRC and therefore are not necessarily very precise. Table 13.4 compares 2001 RRC (and census) estimates of total growth and estimates derived from administrative data sources.

The estimates are a combination of many other estimates of population growth (births, deaths, the number of immigrants, internal migration, emigration, the net number of non-permanent residents, and the increase in the population of incompletely enumerated Indian reserves and Indian settlements), which are subject to errors whose magnitude depends on the source of the data. The estimate of the net number of non-permanent residents is a case in point. It is also important to keep in mind that the RRC was not designed to produce these kinds of estimates and that these estimates are actually by-products. As a result, differences between the two series of estimates are to be expected.

There is a difference between the estimates from the two sources at the national level (1.7%). The two estimates are very similar for Alberta and British Columbia. The largest differences appear in the estimates for Quebec (43,057) and Ontario (-41,317), and the difference is positive in one case and

negative in the other. When taken as a percentage of the RRC estimate, the differences look much larger (22% and 5.3% respectively) for Quebec.

Table 13.4 Comparison of Estimates of Population Growth between Census Counts (Adjusted with Reverse Record Check Results) to Vital Statistics Counts, by Provincial Total and Provinces

Province	Reverse Record Check	Vital statistics	Difference
	Estimated number	Estimated number	
Provincial Total	**1,428,406**	**1,453,268**	**-24,862**
Newfoundland and Labrador	-35,628	-26,882	-8,746
Prince Edward Island	-2,275	2,647	-4,922
Nova Scotia	23,647	11,154	12,493
New Brunswick	-11,570	2,025	-13,595
Quebec	191,697	148,640	43,057
Ontario	772,183	813,500	-41,317
Ontario	7,709	16,525	-8,816
Saskatchewan	-13,171	-2,139	-11,032
Alberta	287,507	281,558	5,949
British Columbia	208,308	206,240	2,068

14. Refusal Indian Reserves and Settlements

14.1 Introduction

On some Indian reserves and settlements in the 2001 Census, enumeration was not permitted or was interrupted before it was completed. These areas, a total of 30 in the Census, are called "incompletely enumerated Indian reserves and Indian settlements". Census data for these areas are not available and therefore have not been included in any census tabulation.

These areas present unique problems for the coverage studies and for the Population Estimates Program. The Reverse Record Check (RRC) target population is the same as the census one, which is described in Section 4.2 of this document. However, the population that is actually covered by the RRC does not include those residents for which the 2001 Census did not attempt to collect any data. For purposes of the RRC, these residents would be considered to be out-of-scope. However, the Population Estimates Program requires an estimate of the population living in these areas. Consequently, since neither the census nor the RRC is in a position to produce such an estimate, another approach is required. This chapter summarizes the approach taken to estimate this population.

Table 14.3.1 shows the name of all the Indian reserves and Indian settlements defined by census collection to be "incompletely enumerated" in the 2001 Census. The table also shows the enumeration status of the listed reserves and settlements in the 1996 Census. Note that of the 30 reserves and settlements, nine were considered to have complete enumerations in 1996 while the other 21 were considered to be "incompletely enumerated" or "refusal". These results are in contrast to the 1996 Census when 77 reserves and settlements, with approximately 50,000 persons, where classified as "incompletely enumerated" (Hamel 1997).

14.2 Methodology

A two-step estimation model was developed to estimate the 2001 incompletely enumerated Indian reserve or Indian settlement population. First, a simple linear regression was built to predict the census count in 2001. Then, in order to be consistent with the Reverse Record Check (RRC) results, an adjustment was made to the estimated "census" count to account for net undercoverage that all census counts are subject to.

The linear regression was constructed using all Indian reserves and settlements that were completely enumerated in both the 1996 and the 2001 Censuses. The basic model assumes a linear growth from 1996 to 2001 for all provinces. However, for the intercept and the regression parameters, separate estimates were produced for each province for this simple model. The model was evaluated for the basic regression assumptions of independence of errors, homogeneity of variances and normality of errors. Further details can be found in Dick and So (2004).

As the input variable for each incompletely enumerated Indian reserve and settlement in 2001, the regression model used either the actual census count in 1996 or the best predicted census count from the model developed by Hamel (1997). The output of the model was the estimated census count in 2001.

These census counts, in order to be consistent with all population estimates from the RRC, have to be adjusted for net undercoverage. The net missed persons were estimated by calculating the net undercoverage rate for all completely enumerated Indian reserves and settlements in each province. This provincial rate was then applied to the estimated "census" count of all the incompletely enumerated Indian reserves and settlements in the province.

The estimated "census" count and the "estimated net missed persons" count in each reserve or settlement were then summed to create an "estimated" population for the incompletely enumerated Indian reserve or settlement.

14.3 Results in 2001

The results of the regression model are shown in Table 14.3.1 in the column called "Estimated census count in 2001". The 2001 estimates based on the regression model are approximately 10% larger than the 1996 estimates. The number of missed persons in each reserve or settlement is estimated by applying the net undercoverage rates to the estimated census counts. The final total of 34,541 represents the estimated population living on incompletely enumerated Indian reserves and settlements in 2001.

Table 14.3.1 Results for Incompletely Enumerated Indian Reserves and Indian Settlements in 2001[1] by Province and Reserve or Settlement Name

Province / Reserve or settlement name	Status in 1996 Census	Census count or estimate in 1996	Estimated census count in 2001	Undercoverage rate[2] in 2001 (%)	Estimated missed in 2001[3]	Population estimate in 2001
Quebec[4]		**10,717**	**11,717**	7.4	**931**	**12,648**
Akwesasne (Partie)	Incomplete	2,265	2,471		196	2,667
Doncaster 17	Complete	0	4		0	4
Kahnawake 14	Incomplete	7,087	7,757		617	8,374
Kanesatake	Incomplete	1,137	1,235		98	1,333
Lac-Rapide	Complete	228	250		20	270
Ontario[4]		**13,176**	**14,345**	**10.1**	**1,615**	**15,960**
Akwesasne (Part) 59	Incomplete	1,229	1,335		150	1,485
Bear Island 1	Complete	153	169		19	188
Chippewas of the Thames First Nation 42	Incomplete	777	839		94	933
Goulais Bay 15A	Incomplete	46	54		6	60
Marten Falls 65	Complete	204	228		26	254
Moose Factory 68	Complete	0	4		0	4
Munsee-Delaware Nation 1	Incomplete	214	235		26	261
Ojibway Nation of Saugeen (Savant Lake)	Incomplete	203	224		25	249
Oneida 41	Incomplete	1,226	1,332		150	1,482
Pikangikum 14	Complete	1,17	1,261		142	1,403
Rankin Location 15D	Incomplete	670	723		81	804
Six Nations (Part) 40[5]	Incomplete	632	681		77	758
Six Nations (Part) 40	Incomplete	4,731	5,174		583	5,757
Tyendinaga Mohawk Territory	Incomplete	1,15	1,249		141	1,390
Wahta Mohawk Territory	Incomplete	154	171		19	190
Whitefish Bay 32A	Incomplete	502	538		61	599
Whitesand	Complete	115	128		14	142

Province / Reserve or settlement name	Status in 1996 Census	Census count or estimate in 1996	Estimated census count in 2001	Undercoverage rate[2] in 2001 (%)	Estimated missed in 2001[3]	Population estimate in 2001
Manitoba[4]		**85**	**97**	**12.1**	**13**	**110**
Dakota Tipi 1	Incomplete	85	97		13	110
Saskatchewan[4]		**451**	**491**	**15.6**	**90**	**581**
Big Head 124	Incomplete	451	491		90	581
Alberta[4]		**3,775**	**4,115**	**17.3**	**862**	**4,977**
Ermineskin 138	Incomplete	1,300	1,413		296	1,709
Little Buffalo	Incomplete	221	243		51	294
Saddle Lake 125	Incomplete	2,254	2,459		515	2,974
British Columbia[4]		**225**	**255**	**2.9**	**8**	**263**
Esquimalt	Incomplete	82	93		3	96
Marble Canyon 3	Complete	67	76		2	78
Pavilion 1	Complete	76	86		3	89
Provincial Total[4]		**28,429**	**31,02**	**10.4**	**3,521**	**34,541**

1. Cautionary Note: The 2001 estimates in Table 14.3.1 are purely "model-based estimates". The validity of these estimates depends heavily on whether the model has correctly captured the true underlying situation. This is an unverified assumption: consequently these estimates should be used with extreme caution.
2. The net undercoverage rate was estimated from completely enumerated Indian reserves in each province. An Indian reserve was defined as all census subdivision types coded as Indian Reserve, Indian Government District, Indian settlement, *Terre Inuit, Village nordique, Village Cri*.
3. Estimated missed = Estimated census count / (1-Estimated missed rate).
4. Rounding may account for small differences in the calculation of the totals.
5. There are two *Six Nations (Part) 40*. The first listed refers to Standard Geographic Code of 3528037 and the second is 3529020.

Appendix A. Acronyms

The following acronyms are used in this report.

AMS	Automated Match Study
CA	Census Agglomeration
CANCEIS	Canadian Census Edit and Imputation System
CATI	Computer-assisted Telephone Interviewing
CC	Census Commissioner
CCD	Census Commissioner District
CCRA	Canada Customs and Revenue Agency
CDS	Collective Dwelling Study
CMA	Census Metropolitan Area
CR	Census Representative
CV	Coefficient of Variation
DCS	Dwelling Classification Study
EA	Enumeration Area
FED	Federal Electoral District
FOS	Follow-up Survey
GES	Generalized Estimation System
HO	Head Office (of Statistics Canada)S
NAFTA	North American Free Trade Agreement
NAICS	North American Industry Classification System
NIM	Nearest-neighbour Imputation Method
NPR	Non-permanent Resident
RO	Regional Office (of Statistics Canada)
RRC	Reverse Record Check
SAS	Statistical Analysis System
SIC	Standard Industrial Classification
SP	Selected Person
SPIDER	System for Processing Instructions from Directly Entered Requirements
UR	Usual Resident
VC	Vacancy Check
VR	Visitation Record
VS	Vital Statistics

Appendix B. Glossary of Terms

The definitions of census terms, variables and concepts appear in the *2001 Census Dictionary* (Catalogue No. 92-378-XIE). Users should refer to the *2001 Census Dictionary* for full definitions and additional remarks related to any concepts, such as information on direct and derived variables and their respective universe. The following terms do not appear in the *2001 Census Dictionary* but are relevant to this report.

Census Overcoverage: Overcoverage occurs when a unit that is not part of a census universe is included in the census. Overcoverage may occur in two ways. First, and most common, is the situation where a unit that is part of a census universe is enumerated more than once. Secondly, a unit outside the census universe (e.g., a foreign resident, a fictitious person, or an unoccupied marginal dwelling) may be erroneously enumerated.

Census Undercoverage: Undercoverage occurs when a unit that is a part of a census universe is completely missed by the census.

Net Undercoverage: Net undercoverage is the difference between the total undercoverage and the total overcoverage for a census universe.

Appendix C. 2001 Census Products and Services

The census is a reliable source for describing the characteristics of Canada's people and dwellings. The range of products and services derived from census information is designed to produce statistics that will be useful, understandable and accessible to all users. Sources, such as the *2001 Census Catalogue*, the Statistics Canada Web site (http://www.statcan.ca) and, specifically, the On-Line Catalogue, contain detailed information about the full range of 2001 Census products and services.

There are several new product and service features for the 2001 Census:

1. Media

- The Internet is the preferred medium for disseminating standard data products and reference products.
- More census data are available to the public free of charge via the Internet.

2. Content

- Data tables for the 2001 Census are released by **topics**, that is, groups of variables on related subjects.
- Wherever possible, the language and vocabulary used in 2001 Census products available on the Internet is simplified to make the information accessible to more people.
- Users are offered various methods of searching and navigating through **census standard products** (including **reference products** on the Internet).

3. Geography

- Geographic units such as dissemination areas, urban areas, designated places and metropolitan influenced zones were added to the standard products line. Some new units, such as dissemination areas, replace others.

4. Variables

- Information on the following new subjects was collected in the 2001 Census: birthplace of parents, other languages spoken at home and language of work. The 2001 questionnaire also included the question on religion, which is asked in every decennial census. The family structure variable was broadened to include same-sex couples.

Bibliography

Armstrong, M., H. Farr and H. Richards. 2003. 2001 Reverse Record Check Survey: Data Collection Report. Unpublished document. Ottawa. Statistics Canada.

Clark, C., M. Armstrong and C. Thibault. 2003. Measurement and Innovations in the 2001 Census Coverage Studies. Proceedings of the Statistical Society of Canada Annual Meeting. Halifax. June 8–11.

Dick, P. 2002a. The Census of Canada: The Dwelling Classification Study. Paper presented at the Joint Statistical Meeting of the American Statistical Society. New York. August 11–15.

———. 2002b. The Dwelling Classification Study: Results and Methods. Presentation prepared for Statistics Canada's Federal–Provincial–Territorial Meeting on Demography. Ottawa. September 12.

Dick, P. and G. So. 2004. Estimating the Population Count on Incompletely Enumerated Indian Reserves in the 2001 Census. Unpublished document. Ottawa. Statistics Canada.

Hamel, D. 1997. Estimating the Population Count and the Number of Occupied Private Dwellings on Indian Reserves Partially Enumerated in the 1996 Census. Unpublished document. Ottawa. Statistics Canada.

Morel, J. and G. Kleim. 2003a. 2001 Census Coverage Studies: Report on Overcoverage: Evaluation. Unpublished document. Ottawa. Statistics Canada.

———. 2003b. 2001 Census Coverage Studies: Report on Overcoverage: Review of the Study Methodology. Unpublished document. Ottawa. Statistics Canada.

Parenteau, M. 2003. 2001 Reverse Record Check Classification. Unpublished document. Ottawa. Statistics Canada.

Statistics Canada. 2002a. 2001 Census Dictionary. 2001 Census Reference Series. Catalogue No. 92-378-XIE. Ottawa.

———. 2002b. 2001 Census Handbook. 2001 Census Reference Series. Catalogue No. 92-379-XIE. Ottawa.

———. 1999. Coverage. 1996 Census Technical Reports Series. Catalogue No. 92-370-XPB. Ottawa.

Théberge, A. 2002. Facteurs de rajustement pour la surcouverture dans la CVD. [Adjustment factors for overcoverage in the Reverse Record Check]. Unpublished document. Ottawa. Statistics Canada.

Théberge, A. and W.C. Liu. 2003. Non-response adjustment and variance estimation for the 2001 Reverse Record Check. Unpublished document. Ottawa. Statistics Canada.

Thibault, C. 2003. 2001 Census of Population: Coverage Study Evaluation Report. Unpublished document. Ottawa. Statistics Canada.

Tourigny, J., M. Bureau and C. Clark. 1998. Revised Direct Estimates of the 1991 Census Coverage Error. Unpublished document. Ottawa. Statistics Canada.