RESULTS OF THE 1991 CENSUS COVERAGE ERROR MEASUREMENT PROGRAM

Marie-France Germain and Claude Julien
Statistics Canada

ABSTRACT

The objective of the 1991 Census Coverage Error Measurement Program was to investigate the incidence of coverage errors with respect to the census universes of population, households and dwellings. Four studies were conducted under the framework of the program. The Vacancy Check Study estimated the number of households and persons missed because their dwelling was misclassified as unoccupied. The Temporary Residents Study produced estimates of persons missed because they were temporarily away from their usual place of residence. As in previous censuses the results of these two studies were used to adjust the final census counts.

The Reverse Record Check Study estimated population and household undercoverage at the national and provincial/territorial levels and studied the characteristics of persons and households missed. In 1991, some changes to the scope and the methodology of the study were made. The Overcoverage Study provided estimates of population and household overcoverage and covered both duplicate enumerations and erroneous inclusions. This study was divided into four sub-components, each directed at estimating different types of overcoverage.

This paper describes each of the above studies but focuses mainly on the Reverse Record Check and the Overcoverage Study. It presents a summary of their methodology and their results and discusses some of their limitations.

KEYWORDS

Census, Coverage Error, Reverse Record Check, Overcoverage Study

1. Introduction

1.1 The Canadian Census

The Census of Population, conducted every five years, provides a wide range of demographic data on the Canadian Population. It is also the basis for a number of other statistical series produced by Statistics Canada, including the quarterly and annual population estimates. The recent Census is of good overall quality but, as

with all large statistical surveys, the resulting data are subject to a certain degree of error.

Errors can arise at virtually every stage of the census collection and processing operations. Checks built into the census operations help to control the number and magnitude of these errors, but beyond a certain point such quality control procedures cease to be cost effective and some small amount of error must be accepted.

In Canada, over 98% of all census data are collected through "self-enumeration": that is each household completes its own census form. Only in remote northern areas and most Indian reserves are data obtained by interview. In self-enumeration areas, Census Representatives (CRs) list all dwellings in their Enumeration Area (EA) on a Visitation Record (VR) and drop off the forms. They ask the respondents to complete and mail them back. After any necessary follow-up for missing information, the forms are sent to regional processing sites where they undergo basic checks. The data are then transferred onto a computer file and sent to Ottawa where further checks and corrections are made.

1.2 Coverage Errors

Among the errors that occur in a census, coverage errors are particularly serious since they affect the basic population counts. Undercoverage occurs when households or persons within households are missed. For instance, dwellings which appear to be unoccupied when the CR drops off the census forms, may in fact be occupied on Census Day. The latter is such a common problem that a Vacancy Check Study has been developed to revisit a sample of dwellings listed as unoccupied, verify their occupancy status and adjust the census population counts accordingly.

The Canadian Census is conducted on a de jure basis, which means people should be enumerated at their usual place of residence. In dwellings correctly identified as occupied, individual residents can be missed. For example, persons away at school or on business or vacation can be omitted from the census form. To counteract this problem, persons staying somewhere other than their usual place of residence are asked to complete a special "temporary residents" census form on which they report the address of their usual place of residence. A Temporary Residents Study has been developed to take a sample of these "temporary residents", verify wether they were enumerated at their usual residence and adjust the census population counts accordingly.

These adjustments for unoccupied dwellings and temporary residents reduce undercoverage but do not eliminate it. There are many other reasons why dwellings are missed and persons within dwellings are omitted. Dwellings can be missed because the CR fails to identify the boundary of the EA correctly, or because some dwellings are hidden. Also, some individuals may be missed because they have no

usual residence and did not spend Census Night in any dwelling. The Reverse Record Check has been developed to estimate the total undercoverage, including that portion of undercoverage already adjusted for by the Vacancy Check and Temporary Residents Study.

On the other hand, some errors can result in overcoverage. This occurs when a member of the census target population is counted more than once, or persons who are not in the target population, such as fictitious persons, pets, foreign visitors, etc. are enumerated (hereafter they are referred to as erroneous inclusions). Duplicate enumerations can be caused by factors related to the respondent, such as moving close to Census Day, or having more than one residence. As well, procedural errors like delivering two forms to the same dwelling can result in overcoverage. The Overcoverage Study has been developed to measure this type of error.

In both the Reverse Record Check and the Overcoverage Study, we do not consider a geographic error alone to be a coverage error. If a person is found to be enumerated in the wrong location, the person is not considered as being missed at the right location and overcovered at the wrong one. Persons are enumerated if they are found on a census form anywhere in Canada, even if they should have been enumerated somewhere else according to census residence rules. Likewise, persons are double counted only if they are found on more than one census form anywhere in Canada.

The remainder of this paper describes the methodology and results of the four coverage measurement studies that were conducted for the 1991 Census of Canada. Sections 2 and 3 briefly describe the two studies whose results were incorporated into the final census counts, namely the Vacancy Check and Temporary Residents Study. Sections 4 and 5 describe the Reverse Record Check and the Overcoverage Study respectively. Section 6 describes how the results of all four studies are combined to provide the estimates of net undercoverage and presents some results. Finally, section 7 concludes with the current status on the issue of adjusting the post-censal population estimates for net undercoverage and the future work to be done on the coverage error studies.

2. The Vacancy Check

In this study, a sample of 1,400 EAs was selected. Within each selected EA, every dwelling listed as unoccupied by the Census Representative was revisited shortly after the Census by a specially trained interviewer. Interviews were conducted with occupants, landlords and neighbours to determine the correct occupancy status on Census Day, how many persons were living there and whether there was somewhere else they might have been enumerated.

The dwellings and persons that were missed because of misclassified

unoccupied dwellings were weighted to produce the estimates. Of the roughly 702,000 dwellings originally listed as unoccupied, an estimated 505,000 were listed correctly, 62,000 were actually occupied and 135,000 should not have been included in the housing stock (e.g., businesses, unhabitable dwellings, etc.). Some 127,000 persons were missed in the 62,000 occupied dwellings. The undercoverage rate was 0.60% for missed households and 0.45% for missed persons, a slight increase compared to the 1986 rates of 0.53% and 0.37%.

3. The Temporary Residents Study

In this study, a sample of 12,000 census forms completed by persons listed as temporary residents, in a private or collective dwelling, was selected. The census form completed at the address of their usual place of residence (reported on the census form completed where they were selected) was checked to determine wether the person was enumerated or missed.

The persons who were missed were weighted to produce the estimates. Of the roughly 576,000 temporary residents listed in the Census, 93,000 of them were estimated to have been missed at their usual residence. The undercoverage rate was 0.33%, a slight increase compared to the 1986 rate of 0.29%.

For further details on the methodology of the Vacancy Check and Temporary Residents Study, see Statistics Canada (1990).

4. The Reverse Record Check

4.1 Overview

The Reverse Record Check (RRC) is the main study of undercoverage of persons and households in the Canadian Census. It has been the source of official estimates of census undercoverage since 1966. In 1991, for the first time, its results combined with those of the Overcoverage Study have been used to produce estimates of net undercoverage.

Also for the first time, the Yukon and Northwest Territories were included in the study as well as non-permanent residents. The non-permanent residents, that is, persons in Canada holding student or employment authorizations, Minister's permits (including extensions) and persons claiming refugee status, were included in the census target population for the first time in 1991.

The RRC methodology involves the creation of a comprehensive list of all persons who should have been enumerated in the Census from sources independent of the current census. A sample is selected

from this list and a number of tracing operations are then undertaken to determine the census day address of each selected person. These operations include a computer linkage to administrative files to update the selected person's (SP) address, a search in census forms carried out during Regional Office Processing, telephone tracing, and searches in various administrative records. Once the census day address is determined, a search in the census documents is carried out to determine whether the person was enumerated at the census day address. The results of the tracing and searching operations lead to the classification of each selected person as either enumerated in the Census, missed in the Census, deceased, emigrated or abroad prior to Census Day, or not traced. The results are then coded, captured, edited and weighted and estimates of undercoverage are produced.

For a discussion on the limitations of the Reverse Record Check methodology, see Burgess (1988).

4.2 Methodology

4.2.1 Frame Construction and Sample Selection

The target population, which contains all persons who should have been enumerated in the 1991 Census, was constructed from the following six basic sources or frames:

(a) 1986 Census Frame : all persons who were enumerated in the 1986 Census, including persons enumerated without names.

(b) Birth Frame: all babies born during the intercensal period, i.e. between June 3rd, 1986 and June 3rd 1991 inclusive. The sample is drawn from provincial records of registered births.

(c) Immigrant Frame: all landed immigrants (persons who immigrate with the intention of becoming Canadian citizens) who entered Canada between June 3, 1986 and June 3, 1991. The sample is drawn from administrative records of Employment and Immigration Canada.

(d) Missed Frame: All persons not enumerated in the 1986 Census. While no complete list of these persons exists, a sample is available from the 1986 Reverse Record Check.

(e) Permit Holder and Refugee Frame: All persons in Canada on June 4, 1991 on student or work authorization, Minister's permit (and extensions) or claiming refugee status. A sample of these persons is selected from Employment and Immigration Canada records.

(f) Health Care Plan Frame: All persons listed on the Yukon and the Northwest Territories Health Care files as of June 4, 1991.

The first five frames were used to estimate undercoverage in the ten provinces whereas the samples for the two territories were drawn from the last frame only.

Sampling was carried out independently within each frame. The sample design varied from frame to frame, depending on the format of the list available. A total of 55,912 persons were selected for the sample and distributed as follows:

Frame	Sample size (persons)
Census	45,300
Birth	2,344
Immigrant	1,447
Missed	1,522
Permit holders and refugees	799
Health Care files	4,500
Total	55,912

The sampling rates within frames were not uniform. Higher rates were used in certain subgroups for which high undercoverage was expected in order to improve the sample design. Examples of such subgroups are persons aged 15-19 in the 1986 Census frame or more recent immigrants in the immigrant frame.

The various frames were stratified by province at time of selection (when this was available) to ensure a good geographic spread of cases and, within each frame, by variables that are known to be related to undercoverage.

The census frame was first stratified by 1986 province of residence. Within provinces, the population was further stratified by 1986 method of enumeration (mail-back, pick-up, canvasser) and by size of municipality. Special types of enumeration areas such as large collective dwellings or Indian Reserves formed separate strata as they have had higher undercoverage in past censuses. A two-stage sample was selected within each stratum. A sample of 1986 Census enumeration areas (EA) was selected for the first stage. Within each selected EA, a systematic sample of approximately 10 persons (SP) was then selected.

The birth frame was first stratified by province of residence of the SP's mother. Year of birth was also used as a second stratification variable since it was observed that babies born in the two years prior to Census showed a higher undercoverage rate.

For the immigrant frame, the stratification used year of arrival in Canada since there is a higher undercoverage among more recent immigrants. Province of landing was not used as a proxy for province of residence because of the high mobility of the immigrants. For the birth and immigrant frames, systematic samples were selected from within each stratum.

The missed frame is a conceptual frame since there exists no list

of all persons missed in the 1986 Census. The sample from that frame consists of all persons classified as missed in the 1986 Reverse Record Check. The sample is not stratified as such although there is an implicit stratification since cases missed in 1986 come from different frames and strata in 1981.

The permit holder and refugee frame was divided by type and duration of permit for permit holders and the refugees formed a stratum on their own. Since this is the first time that we measured undercoverage for this new population, no prior information was available, however type and duration of permit were believed to have an effect on undercoverage. Again, systematic samples were selected within each stratum.

For each of the first five frames, the samples were selected in five replicates for variance estimation purposes. The replication method allowed for a better variance estimation in the presence of non-response adjustments and post-stratification and to take into account the two-stage design used in the census frame.

Finally, age, sex, type of area (rural vs urban) and aboriginal status were used to form strata within each of the two territories of the Health Care Plan frame. For this frame, the sample was not selected using replicates as the design was simple and no post-stratification was used.

4.2.2 Tracing, Searching and Classification Operations

The purpose of the various Reverse Record Check operations was to classify each selected person as one of the following:

(a) enumerated in the 1991 Census;
(b) missed in the 1991 Census;
(c) died before the 1991 Census;
(d) emigrated or abroad before the 1991 Census;
(e) out-of-scope, i.e. should not be included in the 1991 Census (eg. babies born after June 3, 1991, permit holders who were no longer in Canada on June 4, 1991).
(f) not traced, i.e. the classification of the SP on Census Day is unknown.

Addresses obtained at the time of selection of the sample were generally out of date. Consequently, it was necessary to establish the address of each SP on June 4, 1991 (a process known as "tracing") so that the 1991 census form for that address could be searched.

The first step in the tracing was a computer linkage to administrative files in an effort to update the SP's address. The match was carried out for SPs and household members from the 1986 census frame and the missed frame, and for parents of children born in the years 86-89 for the birth frame. The addresses obtained from the match were addresses of SPs or household members in early 1990.

Then, in order to proceed to the next step, the most likely 1991 EAs in which the SP's last known address was located were determined.

The next operation consisted of a manual match carried out as part of Regional Office processing as the 1991 EA boxes were received from the field. This involved a search of the 1991 census documents (visitation records and census forms) to determine whether the SP had been enumerated at his or her last known address. All cases where the SP was found listed on a census form in the Regional Office processing match were classified as "enumerated" and considered closed. They were sent back to Head Office in Ottawa.

For cases which were not found during the previous step, telephone tracing from the Regional Offices was attempted. The interviewers traced the SP's address through several sources such as telephone and city directories, voter lists, provincial agencies, schools, ethnic associations, neighbors, superintendents, etc. Then, they contacted the SP and asked for his/her whereabouts on Census Day as well as for previous or temporary addresses in that period and also for characteristics asked on the census form.

The next step was an automated match (using date of birth and sex) to the 1991 census database in the EAs identified as containing the SP's census day address. This was followed by census form verification of names and addresses for the matched records or, for non-matches, by a complete manual search in census documents. These searches were undertaken to determine whether the SP had in fact been enumerated at the traced address. They were conducted as part of the Head Office processing operations of the Census.

Finally, a search was made in the administrative records of Health and Welfare Canada (family allowance and old age security) and Revenue Canada to obtain new addresses for traced SPs which had not been found on a census form after the various searches described above or for SPs who had not been traced. When new possible addresses were obtained, searching of the census forms took place and a classification result was assigned.

A verification of the final classification of cases was also undertaken. Enumerated cases were matched against the census database to make sure that they were in fact on the database. A review of all missed cases was done to ensure that all possible census day addresses had been obtained and searched correctly. At the same time, reasons why they were missed were coded. For deceased cases, a match to the death register was carried out to verify that these persons had in fact died prior to June 4, 1991.

No verification could be carried out for SPs traced as "having emigrated prior to June 4, 1991", since no emigration records exist in Canada. Persons were classified in the above category only if the source of information was deemed to be reliable. Finally, a search of the last known address of residence in Canada was undertaken for "abroad" cases to ensure that these persons had not

been listed on a census form by other persons.

4.2.3 Data Processing and Production of Results

The processing of the data collected and results of searches was carried out in four main steps:

(a) coding and data capture;
(b) computer edit, manual review and correction of errors;
(c) weight adjustments;
(d) calculation of final estimates of undercoverage and their standard errors.

The first two steps were carried out simultaneously during the searches of Head Office processing and the verification of the classification whereas the last two were done once the classification had been finalized.

The third step consisted of two weight adjustment procedures. The first adjustment was a non-response adjustment done in three stages which involved a redistribution of the initial weight (the inverse of the probability of selection) of cases not traced to cases traced within certain subgroups in each replicate of the sample. The subgroups varied from frame to frame. The second weight adjustment ensured consistency with known frame totals. For both adjustments, the subgroups were defined in terms of information obtained at the time of selection.

Estimates of population undercoverage were then obtained by summing the final adjusted weights. Estimates of household undercoverage were similarly obtained, although this required a further weight adjustment using the household size at the time of the 1991 Census.

The undercoverage rate $\hat{U}_G$ was derived as follows:

$$\hat{U}_G = \frac{\hat{M} - R}{C + (\hat{M} - R) - \hat{O}}$$

where $\hat{M}$ is the RRC estimate of persons missed in the Census;

R is the number of persons added to the census counts by the Vacancy Check and the Temporary Residents Study;

C is the published census count; and

$\hat{O}$ is the estimate from the Overcoverage Study of persons

included erroneously or counted more than once.

Note that the random additions are netted out of the undercoverage because they have been included in the final census counts.

4.3 Results

4.3.1 Final Classification and Weight Adjustments

The results of the various tracing and searching activities conducted in the Reverse Record Check led to the classification of the 55,912 selected persons into the 5 categories as shown in Table I.

The "non-response" (not traced) rate of 4.8% included three types of non-response: a) cases that were not searched because of the absence of information at selection (persons "assigned" in the 1986 Census or persons enumerated without names); b) cases not traced during regional office tracing; c) cases traced in the field but for which the census day address was too vague to assign an EA (eg. the address reported was "SP lived in Toronto" or "SP had no fixed address").

The first type of non-response is due to frame data problems. It affected 334 cases and accounted for 0.6% of the 4.8%. The weights of these cases were redistributed over all other cases in the sample.

The second type of non-response arose from refusals and non-contacts in the field. The 1,788 cases (3.2%) not traced in the field did not distribute evenly among frames; there were larger proportions in the immigrant and the permit holder and refugee frames. The weights of these cases were distributed over traced cases from the other five classifications.

Finally, 582 cases fell into the third type of non-response. For these cases, a contact was made in the field and confirmed that the SP was alive and in Canada on Census Day. But because the SP's census day address could not be located, these cases could not be assigned to either the "enumerated" or "missed" category. They had their weights redistributed over enumerated or missed traced cases.

The non-response adjustment lead to an implied undercoverage rate for no-trace cases of 12.2%, this rate being calculated as

$\hat{M} / (\hat{E}_{RRC} + \hat{M})$ where $\hat{E}_{RRC}$ is the RRC estimate of the number of

persons enumerated in the Census. The adjustment increased the national undercoverage rate by approximately one third of a percentage point.

The second weight adjustment was an adjustment to known frame

Table I
Number of Cases in Each Final Category by Frame
Number and Per Cent (in Brackets)

| FRAME | CLASSIFICATION RESULT | | | | | |
	ENUMERATED	MISSED	DECEASED	EMIG./ABR./ OUT-OF-SCOPE	NOT TRACED	TOTAL
CENSUS	39,739 (87.7)	1,737 (3.8)	1,945 (4.3)	356 (0.8)	1,523 (3.4)	45,300 (100.0)
BIRTH	2,110 (90.0)	81 (3.5)	23 (1.0)	20 (0.9)	110 (4.7)	2,344 (100.0)
IMMIGRANT	1,044 (72.1)	127 (8.8)	5 (0.3)	69 (4.8)	202 (14.0)	1,477 (100.0)
MISSED	1,069 (70.2)	203 (13.3)	45 (3.0)	46 (3.0)	159 (10.4)	1,522 (100.0)
NON-PERMANENT RESIDENTS	320 (40.1)	133 (16.6)	0 (0.0)	49 (6.1)	297 (37.2)	799 (100.0)
HEALTH CARE	3,809 (84.6)	246 (5.5)	0 (0.0)	32 (0.7)	413 (9.2)	4500 (100.0)
TOTAL	48,091 (86.0)	2,527 (4.5)	2,018 (3.6)	572 (1.0)	2,704 (4.8)	55,912 (100.0)

totals for variables such as age and sex. It affected the national undercoverage rate only minimally.

4.3.2 Population Undercoverage

After the reweighting was done, the results showed that, at the national level, undercoverage was 3.8% with a standard error of 0.13%, an increase of 0.4% from the rate observed in 1986[1]. This increase is explained primarily by the addition of the permit holders and refugees into the census population and into the measure of undercoverage (0.3%). If we adjust the 1986 and 1991

---

[1] Many Indian Reserves refused to answer the 1986 and 1991 censuses. In 1986, a reliable estimate of the size of their population was produced and added to the census counts. Thus, the population on these reserves was considered as "enumerated" in the 1986 Census. Unfortunately, such a reliable estimate was not available for the 1991 Census. Hence, in order to make the 1986 rate comparable to the 1991 rate, the 1986 rate had to be recalculated with the population on these reserves now considered as "missed". This recalculation increased the national undercoverage rate from 3.2% (the official 1986 rate) to 3.4%.

rates to make them directly comparable, i.e. excluding the territories and the permit holder and refugee frame, the rates are approximately the same.

The rates varied among the ten provinces from 1.9% in Prince Edward Island to 4.7% in Ontario. They were high also for the territories with 4.4% in the Yukon and 6.4% in the Northwest Territories, partly due to the high mobility of the territorial population.

As observed in the previous censuses, young people aged 20-29 years in 1991 tended to be missed at a higher rate (8.0%) than other age groups, and males showed a higher rate than females in that group (9.4% vs 6.5%) and in general. Single (15 years and over) and divorced persons had an undercoverage rate of 7.4%, the rate for males (8.7%) being again higher than for females (5.9%). Persons whose mother tongue is neither English nor French were missed at almost twice the rate (6.1%) of persons who reported either one of the two official languages (3.3%).

Persons living in large urban centres as well as in rural areas were missed more often than others with both rates equal to 3.9%. The rates for the three largest census metropolitan areas ranged from 5.2% in Toronto to 3.3% in Montréal with Vancouver at 3.5%. Households who rented their dwelling were missed much more often than those who owned it with rates of 5.5% and 1.2% respectively.

A preliminary analysis of the results showed that 54% of the undercoverage resulted from missing the SP's household whereas 41% of missed SP's were missed in enumerated private households. Another 5% were missed in collective households (dwellings).

Of cases where a complete household was missed, one fifth came from the SP's dwelling being missed; for another 23%, the SP's dwelling had been classified as being vacant on Census Day. In 5% of cases, the SP's household had been incorrectly classified as a foreign or temporary resident household and therefore not counted. Another 5% of household undercoverage was due to refusal of some Indian reserves to participate in the Census. For the remaining 47% missed households, another household was found at the address reported as the SP's census day address. A further investigation of these cases is planned as part of the evaluation studies to take place later this year. Among the possible reasons for this are the household moving around Census Day, two households living at the same address, another dwelling located in the same building but missed by the census representative, etc.

Undercoverage occurring within enumerated households was more difficult to explain. More frequent reasons were SPs who moved around Census Day, SPs being away temporarily, working, at school or outside Canada, SPs living with non-relatives who forgot to include them on the census form, SPs having no usual residence on Census Day and SPs identified as temporary or foreign residents. Again, an evaluation of the reasons for within household undercoverage is planned.

Further evaluations of results and of methods used for non-response adjustment and post-stratification as well as for sampling design and estimation will take place later this year.

5. The Overcoverage Study

5.1 Overview

The Overcoverage Study (OCS), as its name implies, is the main study of overcoverage of persons and households in the Canadian Census. It was first carried out on an experimental basis for the 1986 Census. For the 1991 Census, it was carried out on a much larger scale and its results were combined with those of the RRC to produce estimates of net undercoverage.

The OCS consists of three surveys that detects persons who should not have been counted in the Census (erroneous inclusions) and persons who were counted more than one. In the three surveys, a sample of census enumerations is selected. Data are collected to determine the erroneous inclusions and to obtain any addresses, other than where the selected census enumerations took place, where the persons might also be enumerated. A search of census forms determines whether or not the persons are enumerated more than once.

The OCS also involves an Automated Match Study (AMS) that detects persons who were counted more than once within the same Enumeration Area, probably at the same dwelling. The AMS does not involve any additional data collection. A computer search of the census database detects pairs of households having members with identical or similar characteristics. A clerical verification of their census forms then determines if the same persons were counted at both households.

For a more complete overview, see Dibbs and Royce (1990).

5.2 Methodology

5.2.1 Frame Construction and Sample Selection

The target population, which contains all 1991 census enumerations, was covered by the following components:

(a) Private Dwelling Study (PDS): all persons who were enumerated in a private dwelling. Indian reserves and remote northern areas were excluded in order to control the cost of data collection for the study.

(b) Institutional Collective Dwelling Study (ICS): all persons who

were enumerated as a usual resident of an hospital, a prison, etc.

(c) Non-Institutional Collective Dwelling Study (NICS): all persons who were enumerated as a usual resident in a hotel, a school residence, etc.

The PDS used a two-stage sample of approximately 30,000 households, with a first stage sample of 2,000 Enumeration Areas (EAs) and 15 households chosen within each selected EA. The EAs were stratified by province and territory (hereafter referred to as province only) and size of municipality. Within each stratum, eight replicates of an equal number of EAs were selected without replacement.

The ICS used a cluster sample. A list of dwellings was created prior to the Census using the 1986 Census information and administrative updates. The list was stratified by province and by four types of dwellings. A sample of 562 dwellings was allocated among the strata. The dwellings within each stratum were sorted by the number of usual residents (estimated prior to the Census) and a systematic sample of dwellings was selected. All usual residents of the selected dwellings were included in the sample.

Usual residents of non-institutional collective dwellings completed a special census form. On it, they were asked: "Is there any other address where someone may have included you in the 1991 Census of Canada?". The 8,920 persons who answered "yes" to this question provided the frame for the NICS. They were stratified by province. A sample of 1,232 persons was allocated to the strata. Within each stratum, the usual residents were sorted by dwelling type and a systematic sample was selected.

5.2.2 Data Collection

For the PDS, a list of selected EAs was sent to the census processing sites. After census data collection, the 15 households were chosen from the Visitation Record and information was transcribed from the census form onto a PDS questionnaire. This information included the selection address and the telephone number and the household members' names, sex and dates of birth.

The households were contacted by telephone and, when necessary, by a personal visit. The respondents were asked questions to determine if they were erroneous inclusions and to report any other addresses where they might have been enumerated (hereafter called reported address). Among the addresses that the interviewers probed for were the previous and present residence of persons who had moved close to Census Day, school residences, cottages or any places where they might have been included one someone else's census form (e.g., students included on parents' form).

For the ICS, the data collection was done along with the regular census collection. The Census Representatives assigned to the

selected dwellings were given a special form on which they entered the dwelling's address (selection address) and then listed all the usual residents along with any other addresses available from the institution's administrative records (reported addresses). No direct contact with the usual residents was required.

For the NICS, the reported addresses were taken from the special census form completed by the usual residents.

5.2.3 Data processing

The data processing consisted in determining the enumeration status of each sampled person at their reported and selection addresses.

The census form that was completed at the reported addresses was searched using an automated match to the 1991 census database and a census form verification similar what is described in 4.2.2. Each reported address was classified as follows:

(a) not located (address level non-response);
(b) same as the selection address;
(c) located but the person was not listed there;
(d) located and the person was listed there.

The first category included addresses that were not precise enough to locate (e.g., just the city or the province). For the PDS, it also included addresses that were not reported (e.g., the respondent reported that a roommate lived elsewhere before Census Day but did not know where or did not want to report it).

At the selection address, each sampled person was classified as one of the following:

(a) correctly enumerated;
(b) erroneously included;
(c) non-response (EA, household, dwelling or person level);
(d) out-of-scope.

The selection of the census enumerations for the PDS and the data collection for the ICS and NICS took place either during or after the census collection but before the checks and data capture of the census information. As a result, some persons and households that were sampled were deleted later from the census population because they were found to be temporary or foreign residents. They were identified by matching the sampled persons to the census database at their selection address and were classified as "out-of-scope". This operation also identified persons who were either added to the census population after data collection or missed during data collection. They were classified as person level non-response.

For the ICS and NICS, persons were classified as "out-of-scope" because, according to the information on the census database, they were not covered by the study. An example would be persons in ICS

dwellings that were not an institution any more and who were covered by either the PDS or the NICS component. Erroneous inclusions were identified in the PDS only. Given the limited amount of information collected, erroneous inclusions were not identified in the ICS and the NICS.

For the PDS, the household level non-response consisted of no-contacts and refusals. For the ICS, it consisted of dwellings that were not processed or not found by the CR, as well as those that did not have any other addresses available on their administrative records.

The enumeration status at the selection and reported addresses were combined and an "overcoverage value" was assigned to each person as follows: 0, for persons correctly enumerated only once; 1, for erroneous inclusions; 1/2, for persons enumerated at the selection address and at one reported address; 2/3, for persons enumerated at the selection address and two reported addresses. The fractions take into account the multiple frame feature of the Overcoverage Study, i.e. the fact that persons who are enumerated more than once in the Census have more than one chance of being selected in the study. It is also important to mention that persons who are counted more than once within the same enumeration area get an overcoverage value of 0 because this specific type of overcoverage is covered by the Automated Match Study.

5.2.4 Weighting and Production of results

For the PDS, design weights were calculated by using the formulas for a two-stage design adjusted for EA and household level non-response. A first adjustment ensured consistency with known population totals for certain subgroups while compensating for the person level non-response. Subgroups of persons were defined according to province, stratum, age and relationship to other members in the household. A second adjustment compensated for the address level non-response. The reported addresses were assigned the adjusted weight of the persons reporting them. The adjustment involved the redistribution of the adjusted weight of non-response addresses to addresses that were located within certain subgroups. Subgroups were defined according to region (Atlantic provinces, Quebec and Ontario, the rest of Canada), the type of address (previous residence, school residence, etc.) and the persons' characteristics. The adjustment required collapsing all the strata and replicates together. The estimate of population overcoverage was calculated by summing the product of the initial weight, the two weights adjustments and the overcoverage value. The estimate of standard error was calculated using the replication technique.

For the ICS, the person level non-response was treated as a random sample selected within the dwellings and a two-stage sample formula was used to assign each person an initial weight. Persons whose address was not located were also treated like non-response persons. The initial weights were adjusted using the ratio method

to ensure that they added up to known population totals. Subgroups were defined by province and the type of collective dwelling.

For the NICS, the estimates were based on a two-phase sample with stratification approach. The sampled persons were divided into three strata: "out-of-scope", "reported address same as selection address" and the others. Only persons in the latter stratum had a chance of being overcovered. The persons whose address was not located were treated as a random sample selected from the third stratum.

5.2.5 The Automated Match Study (AMS)

The AMS covered all private dwellings in Canada. It was designed to detect double-counting due primarily to errors by the Census Representative. For example, the CR might follow-up and complete a census form for a non-response household, only to receive the original form in the mail a few days later. If not detected, this situation results in overcoverage.

The study combined a two-phase and a two-stage design. A sample of 9,500 EAs was allocated among the regions (Atlantic provinces, Quebec, Ontario, Prairie provinces, British-Columbia and the Territories). Every pair of private households within a selected EA was matched to determine the number of similar persons between them. Similar persons were defined as two persons with the same sex and at least two components of their date of birth exactly the same (e.g. a male born on 01/01/62 and a male born on 10/01/62 were similar). Names and addresses were not used because they are not captured in the Canadian Census. Obviously, household pairs with many similar persons had a much higher chance of representing overcoverage than household pairs with little or no similar persons.

Within an EA, the household pairs were grouped into eight categories based on the number of similar persons, the size of the households and the proximity of the households. The EAs were stratified according to the categories of household pairs that they contained and a (second phase) subsample of EAs was allocated among the strata. The subsample rate ranged between 10% and 60%, with the higher rates applied to the strata of EAs containing household pairs that were more likely to be overcoverage. Within each subsampled EA, a second stage sample of household pairs was selected from each category and clerks verified the census form completed by both households to determine the persons who were counted twice. The estimates of overcoverage were produced using the formulas appropriate to this design.

For further details on the Automated Match Study, see Julien (1991).

5.3 Results

5.3.1 PDS, ICS and NICS

Table II presents the initial sample size (obtained after data collection) and the final sample size (remaining after data processing) for the PDS, the ICS and the NICS. It also provides the estimates of population overcoverage and the standard error.

Table II
Results from the PDS, ICS and NICS

	PDS	ICS	NICS
HOUSEHOLDS			
INITIAL SAMPLE	29,736	562	N/A
OUT-OF-SCOPE	173	293	N/A
NON-RESPONSE	700	92	N/A
FINAL SAMPLE	28,863	177	N/A
PERSONS			
INITIAL SAMPLE	76,846	6,729	1,232
OUT-OF-SCOPE	90	0	53
NON-RESPONSE	350	145	0
FINAL SAMPLE	76,406	6,584	1,179
ERRONEOUS INCLUSIONS	22	N/A	N/A
ADDRESSES			
REPORTED	11,222	5,147	1,179
NOT LOCATED	3,441	569	54
DOUBLE COUNTED	397	403	336
ESTIMATES			
OVERCOVERAGE	103,236	7,574	2,548
STANDARD ERROR	9,942	881	125

For the PDS, the household and person level response rates, after having excluded the "out-of-scope", are 98% and 99.5% respectively. Although an attempt was made to improve the inexact addresses, by matching the persons to administrative files, the address level response rate is only 69.3%. The 103,236 overcovered persons estimated are distributed as follows: 9,640 erroneous inclusions, 68,246 persons who were counted more than once, 1,232 persons added

by the first adjustment and 24,056 persons added by the second adjustment. The second adjustment, for address level non-response, accounts for 23% of the PDS estimate.

For the ICS, only 177 of the 562 dwellings were processed and in scope of the study. Most of the "out-of-scope" dwellings had no usual residents (174) or were not enumerated as institutional collective dwellings (113). 62 of the 92 "non-response" dwellings had no alternative address available for all usual residents. These results illustrate the poor quality of the sampling frame. These dwellings are not collected consistently from one census to another. The dwelling types and, especially, the number of usual residents change dramatically. Hence, it is very difficult to develop an optimal design. In the future, simple and more robust designs must be considered.

The 2,548 overcovered persons estimated by the NICS represent more than 28% of the 8,920 persons who reported an address on their census form. It assumes that the other 103,340 usual residents, who did not report an address, were enumerated only once. A study is currently under way to investigate this assumption further by matching a sample of these persons to administrative records to obtain addresses that they should have reported.

5.3.2 **AMS**

The principal goal of the AMS was to provide precise estimates of overcoverage while at the same time having to verify the least number of household pairs. In order to better allocate the clerical resources available for the verification operation, the household pairs with no similar persons were excluded from this study and were assumed to represent no overcoverage. Nonetheless, the matching operation carried out on the 9,500 EAs still produced close to 2.8 million household pairs with at least one similar person.

Among the 2,300 EAs selected for verification, 14,901 household pairs were verified, of which 1,037 contained some overcoverage. In the class of household pairs that were most likely to be overcoverage, 394 of the 398 household pairs verified contained overcoverage. In the least likely class, no overcoverage was found in the 1,864 household pairs verified. The 1,037 household pairs with overcoverage accounted for 2,850 overcovered persons (unweighted).

This component produced an estimate of 44,636 overcovered persons with a standard error of 2,525. This fairly precise estimate was obtained by accepting a slight bias (underestimate). This study relies on the assumption that the overcovered persons report similar characteristics. This is necessary for them to be detected at the matching operation. An overcovered person who reports different characteristics can still be detected at the verification operation as long as another overcovered person in the household is

detected at the matching operation. 3% of the overcovered persons had different characteristics, which indicates that there is some overcoverage among the household pairs that were excluded. However, a huge sample of them would have to be verified to detect one or two cases of overcoverage at the expense of producing a much less precise estimate of overcoverage. Given the choice between a precise but slightly biased estimate and a unbiased but very imprecise estimate, we chose the former.

5.3.3 Population Overcoverage

The Overcoverage Study, when all four components were combined, showed that the overcoverage rate was 0.56%. The rate varied little among the 10 provinces from 0.35% in Saskatchewan to 0.72% in Prince Edward Island. The rates were smaller in the territories with 0.30% in both the Yukon and the Northwest Territories.

Similar to the RRC results, young people aged 20-29 years tended to be overcovered at a higher rate (0.9%) than other age groups, but males and females had similar rates for all age groups. Single person (15 years and over) had an overcoverage rate of 0.86%. Persons whose mother tongue was French, English or neither all had similar rates. More than half the overcoverage was that of a whole household counted at two different places. Among these cases, the rate was higher for renters (0.72%) than for owners (0.27%).

In the PDS, the overcoverage rate increases as the person's relationship to other household members decreases. The person who completed the census form (Person 1) and immediate family (spouse and children) have an overcoverage rate of 0.3%. This rate increases to 1.0% for other persons related to Person 1 (e.g., siblings, in-laws, etc.) and rises to 1.8% for persons who are not related to Person 1 (e.g., lodgers, roommates, etc.).

The types of overcoverage detected are, in decreasing order of frequency: persons who were counted at two or more different private dwellings (56%), persons who were counted twice within the same EA, usually at the same private dwelling, (28%), erroneous inclusions (9%) and persons who were counted at both a collective and a private dwelling (7%). In the PDS, approximately two thirds of the overcoverage occurred at the address where the persons had moved to (or from) close to Census Day.

Analysis of the reasons for overcoverage in the AMS has just begun, but early results indicate that 20% of the overcoverage occurred because of households being added during the census follow-up. Compared to the PDS, the AMS detected relatively fewer persons aged 20-29 years and more persons aged 30-44 years and also more persons in the larger metropolitan areas.

6. Net undercoverage

As mentioned earlier, for the first time in 1991, the estimates of population undercoverage were combined with that of overcoverage to produce estimates of net undercoverage rates as follows:

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

where $\hat{M}$ is the estimate of undercoverage from the Reverse Record Check, R is random additions from the Vacancy Check and Temporary Residents Study, $\hat{O}$ is the estimate of overcoverage from the Overcoverage Study and C is the census count.

$\hat{M}$ and $\hat{O}$ can be treated as independent estimates and the variance of the estimated net undercoverage rate is estimated using the Taylor linearization technique as follows

$$var(\hat{U}) = \frac{(1 - \hat{U})^2 \; (var(\hat{M}) + var(\hat{O}))}{(C + (\hat{M} - R) - \hat{O})^2}$$

where $var(\hat{M})$ and $var(\hat{O})$ are the variance of the undercoverage and overcoverage estimates. Although R is also an estimate, its variance is negligible compared to that of $\hat{M}$ and $\hat{O}$ and is ignored.

Table III gives the net undercoverage rate for several important domains. The net undercoverage rate for Canada (3.2%) is the same as the 1986 undercoverage rate, however these results are not directly comparable, as already discussed in section 4.3.2.

7. The Adjustment of Population Estimates

In late January 1993, the Chief Statistician of Canada announced that Statistics Canada's Population Estimates Program, which provides post-censal estimates of the population between censuses, will be revised to incorporate the agency's estimates of net undercoverage for the 1991 Census. This decision followed several years of ongoing consultations with the major users of the estimates program, with external advisory committees on Demography and Statistical Methods, and with the National Statistics Council, the major advisory body to Statistics Canada.

The population estimates are used in a variety of ways, including the transfer of billions of dollars between various levels of

Table III
Net Undercoverage Rates

DOMAIN	NET UNDERCOVERAGE RATE	STANDARD DEVIATION
CANADA	3.2	0.13
ATLANTIC	2.5	0.20
QUÉBEC	2.8	0.22
ONTARIO	4.1	0.31
PRAIRIES	2.2	0.22
BRITISH COLUMBIA	3.1	0.29
TERRITORIES	5.5	0.60
ALL MALES	3.7	0.17
MALES 0 - 19 YEARS	2.4	0.31
MALES 20 - 29 YEARS	8.4	0.58
MALES 30 - 44 YEARS	4.6	0.35
MALES 45 AND OVER	1.5	0.26
ALL FEMALES	2.7	0.19
FEMALES 0 - 19 YEARS	3.1	0.30
FEMALES 20 - 29 YEARS	5.7	0.58
FEMALES 30 - 44 YEARS	2.2	0.31
FEMALES 45 AND OVER	1.3	0.25
MALES AND FEMALES DIVORCED OR NEVER MARRIED	6.2	0.35
MALES AND FEMALES MARRIED, SEPARATED OR WIDOWED	1.7	0.12
MOTHER TONGUE: ENGLISH	2.8	0.22
MOTHER TONGUE: FRENCH	2.7	0.24
MOTHER TONGUE: OTHER	5.5	0.33
RENTERS	4.8	0.30
OWNERS	0.9	0.10

government, the weighting of current sample surveys, and for important demographic analyses. Because of the financial implications in particular, agreement on adjustment has not been unanimous on the part of the stakeholder community. Nevertheless, the availability of reliable estimates of net undercoverage has now

made it possible to improve the quality of the post-censal estimates. Close to a million persons will be added to the Canadian population by this adjustment.

The new estimates will be released beginning in September 1993. As part of the development of the revised estimates during the next few months, we will be carefully reviewing the results of the Reverse Record Check and the Overcoverage Study to determine where improvements to the estimates of net undercoverage can be made. In addition, we will be developing estimates of net undercoverage for sub-provincial areas and, where feasible, for previous censuses.

The decision to include estimates of net undercoverage in the Population Estimates Program marks a new and an important use of the coverage measurement studies, one which will carry forward into the 1996 and future censuses. Work has already begun on planning the coverage measurement studies for 1996, examining ways of making the results more accurate while consolidating the improvements that were achieved in 1991. This new use of the coverage measurement program will provide an ongoing challenge to us all.

8. References

Burgess, R. (1988), "Evaluation of the Reverse Record Check Estimates of Undercoverage in the Canadian Census of Population", Survey Methodology, 14, 137-156.

Dibbs M.R., and Royce, D. (1990), "Measuring Overcoverage in the 1991 Census of Canada", in Proceedings of the Section on Government Statistics, American Statistical Association, pp. 24-27.

Julien, C. (1991), "Estimation of Overcoverage in the Census of Canada Using an Automated Approach", in Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 431-436.

Statistics Canada (1990), User's Guide to the Quality of the 1986 Data: Coverage, Publication 99-135.