

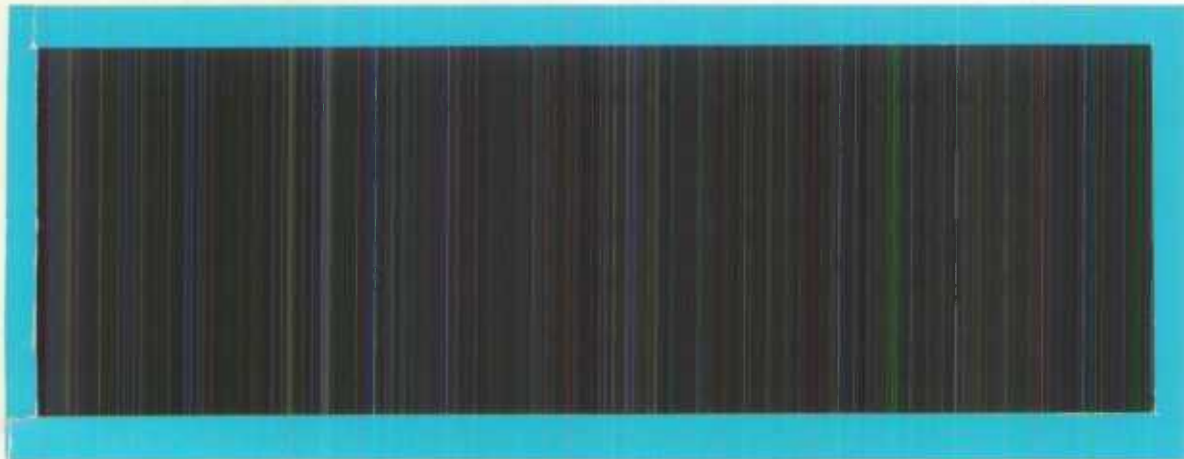
11-613E

no.90-11

c. 2

Statistics  
Canada

Statistique  
Canada



Methodology Branch

Social Survey  
Methods Division

Direction de la méthodologie

Division des méthodes  
d'enquêtes sociales

Canada



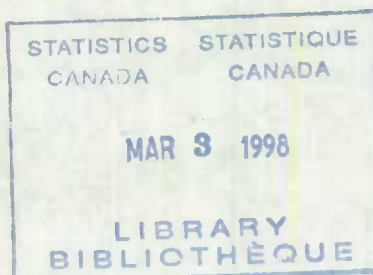
CAHIER DE TRAVAIL NO. SSMD-90-011 E

Methodology

**MEASURING OVERCOVERAGE IN THE 1991 CENSUS OF CANADA**

SSMD-90-011 E

**Ruth Dibbs and Don Royce  
Social Survey Methods Division  
Statistics Canada**



September 1990

ST. JOHN'S UNIVERSITY  
LIBRARY  
MAR 3 1988  
LIBRARY  
ST. JOHN'S UNIVERSITY

## ABSTRACT

Key Words: Post enumeration study, Coverage evaluation

Since 1966, Statistics Canada has used the Reverse Record Check to estimate the numbers and characteristics of persons and households missed by the Canadian census. Until 1986, there had been no Canadian study to measure overcoverage, since it was believed that it tends to be quite small compared to undercoverage. As part of the 1986 Census, an experimental study was carried out to obtain an approximate estimate of the level of overcoverage. For the 1991 Census, the option of an integrated post enumeration survey to estimate net undercoverage was investigated, but considered inapplicable to the current Canadian census environment. Therefore, our 1991 objective is to produce reliable estimates of overcoverage which, when combined with estimates of undercoverage from the Reverse Record Check, will yield estimates of net undercoverage for the first time.

This paper describes the planned 1991 overcoverage studies which will include a survey of persons enumerated in private dwellings, extensive automated matches of the characteristics of persons on the census database and studies in collective dwellings.

## RÉSUMÉ

Mots clés: étude postcensitaire, évaluation de la couverture

Depuis 1966, Statistique Canada a utilisé la contre-vérification des dossiers comme méthode d'estimation du nombre et des caractéristiques des personnes et des ménages non dénombrés par le recensement du Canada. Ce n'est qu'en 1986 qu'une étude a été menée afin de mesurer le surdénombrement, car on croyait qu'il était peu important comparativement au sous-dénombrement. Dans le cadre du recensement de 1986, on a effectué une étude expérimentale afin d'estimer le niveau de surdénombrement. La possibilité d'intégrer au recensement de 1991 une enquête postcensitaire visant à estimer le sous-dénombrement net a été étudiée mais on a établi qu'il était difficile de le faire dans le contexte actuel du recensement du Canada. Notre objectif pour 1991 est alors de produire des estimations fiables du surdénombrement qui, réunies à celles du sous-dénombrement établies à partir de la contre-vérification des dossiers, donneront pour la première fois une estimation du sous-dénombrement net.

L'article décrit les études de surdénombrement prévues pour 1991, qui comprendront une enquête auprès des personnes dénombrées dans les logements privés, des appariement automatisées approfondies des caractéristiques des personnes sur la base de données du recensement et des études portant sur les logements collectifs.





## MEASURING OVERCOVERAGE IN THE 1991 CENSUS OF CANADA

Ruth Dibbs and Don Royce  
Statistics Canada

Presented at the 1990 Joint Statistical Meetings

---

---

### 1. Background

This paper outlines plans for measuring overcoverage as part of the 1991 Census of Canada coverage evaluation program. The overall objectives of the program are:

- A) to measure net undercoverage
- B) to measure the relative contribution of various components and causes of coverage error, such as gross undercoverage and overcoverage among whole households and within households.

Since 1966, Statistics Canada has used the Reverse Record Check (RRC) methodology to produce official estimates of the numbers and characteristics of persons and households missed by the Canadian Census. Until 1986, however, there had been no Canadian study to measure overcoverage, since it was believed that it tends to be quite small compared to undercoverage. As part of the 1986 Census, an experimental overcoverage study was carried out to gain some familiarity with the methods and procedures used, and to obtain an approximate estimate of the level of overcoverage. The Canada level estimate from this study was 45,600 persons enumerated more than once, compared to the 839,000 persons estimated to have been missed by the census (a gross undercoverage rate of 3.2%). This rate of 3.2% represented a rise in undercoverage from the level of about 2% in the three previous censuses. Because of this increase in undercoverage in the 1986 Census, Statistics Canada is dedicating considerable resources in the 1991 Census to coverage improvement procedures in census operations, and to enhancements to the coverage measurement program.

Although the Reverse Record Check is a well developed methodology to measure gross undercoverage, it is, like any coverage measurement technique, subject to limitations (Burgess, 1988). For example, the distinction between whole household undercoverage and persons missed within households depends on some fairly strong assumptions. It also does not give much information on the reasons persons were missed in the census. The 1986 experimental

Overcoverage Study had some operational problems and excluded some components of the institutional population, suggesting that the estimate of 45,600 persons overcounted was low. (Statistics Canada, 1990).

To supplement the information provided by the RRC, and because of the limitations of the 1986 Overcoverage Study, a number of other coverage measurement approaches were investigated for the 1991 Census.

One approach considered was the demographic method which is used in a number of countries. A description of the methodology and its potential application in the Canadian context is given in Romaniuc (1988). However, this method yields only estimates of net coverage error, with no breakdown into gross undercoverage and gross overcoverage. It also gives no indication of the reasons for coverage errors. Nevertheless, further investigation of this method is continuing.

Another option which was examined in detail was an integrated post enumeration survey (PES), modelled on the methodology planned by the United States Bureau of the Census (USBC) for the 1990 census. Such an approach would permit a direct estimate of net undercoverage using a dual system estimate, and would also provide useful information on the components of coverage error. In addition, this would give us some experience with a PES methodology, in the event that the reverse record check methodology became unworkable. If, for example, there were no 1996 Census, (the mid-decade census is a legal, not a constitutional requirement), a 2001 study would be faced with the problem of tracing a sample of persons whose addresses, in most cases, were ten years out of date.

### **1.1 Consideration of an Integrated Post Enumeration Survey**

The 1990 US Census Bureau approach is based on two overlapping samples of geographic blocks. A post enumeration survey is carried out in the selected blocks to create an independent listing of dwelling units and persons (the P-sample). A two-way match is conducted of the survey respondents, to the persons enumerated in the Census for the same blocks, (the E-sample). If necessary, a followup interview is carried out to obtain missing P-sample information or to reconcile E-sample non-matches. The E-sample is also used to search for duplicate enumerations (Hogan, 1986), (Childers et al., 1987).



After careful consideration however, it was decided that this methodology was inapplicable to the current Canadian census environment. One of the reasons is that it implies the use of an area frame based on the census geographic structure. The United States has developed the TIGER system, based on a comprehensive nationwide geographic block structure. The block, which has between 40-75 households, is the basic sampling unit. In Canada, the smallest census area unit is the Enumeration Area (EA), which corresponds to the workload that can be handled by one enumerator. In urban areas, an EA has an average of 350 households, which is too large to serve as a cluster of households. Subsampling of areas within EAs and identifying the census households within the selected areas appeared to be very difficult with this geographic structure. The problem of subsampling is even more severe in the more rural areas of Canada which cover 25 % of the population, and which lack precise addresses. The development of an area frame for areas smaller than the EA would have been very costly and labour-intensive. As well, methods for accurately assigning census enumerations to PES areal units were felt to pose major problems. Moreover, it would not be possible to test this design prior to 1991.

Another reason for concern about adopting a PES methodology in 1991 was related to the time required for the development and testing of two-way matching procedures and to their integration into the census production cycle. Because of the constraints imposed by the census processing schedule, the followup interviews to reconcile inconsistencies could not be carried out until six to nine months after census day, with the concomitant problems of respondents moving and recall difficulties.

As a result of the challenges involved in mounting a timely and accurate matching operation, and the costs and problems related to sample design, it was decided that the option of a US-modelled PES involved too much risk and expense to adopt in 1991. Moreover, because the Reverse Record Check is well established and successful, it had already been decided that it would remain the basic method of measuring gross undercoverage. Therefore, for the 1991 Census, it was decided to carry out a postcensal survey that would measure only gross overcoverage. Our objective is to produce reliable estimates of overcoverage which, when combined with estimates of undercoverage from the Reverse Record Check, will yield estimates

of net undercoverage for the first time. In addition, we will be able to obtain some information on the components and causes of overcoverage.

### **1.2 Definition and Causes of Overcoverage**

Overcoverage in a census occurs when a member of the target population is counted more than once, or persons who are not in the target population, such as fictitious or out-of-scope persons, are enumerated. Unlike the USBC, 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, this is not considered to be an erroneous enumeration. This is in keeping with the approach of the Reverse Record Check which considers persons enumerated if they are found on a census questionnaire anywhere in Canada, even if they should have been enumerated somewhere else according to census residence rules.

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 questionnaires to the same dwelling can result in overcoverage. In the plans for the 1991 study, separate methodologies will be implemented to identify these different sources of errors. Thus, the methodology for the 1991 Overcoverage Study includes three components:

1. a private dwellings survey of a sample of persons and households enumerated in the census, to determine if the persons listed on the census form have been correctly included and to obtain alternative addresses to search for possible double counting;
2. obtaining alternative addresses for a sample of persons in collective dwellings such as student residences, work camps, hospitals, etc. to search for duplicate enumerations.
3. an automated search of the census database of households that are geographically close, matching respondents on date of birth and sex, to identify duplicate inclusions of households or persons;

The major focus of our efforts will be to identify cases where persons have been enumerated more than once. In addition, the private dwelling study will also try to distinguish census



fabrications and persons who are out of scope for the census, which under 1991 coverage rules are mainly foreign visitors.

## **2. The Private Dwelling Study**

### **2.1 Methodology**

A national two-stage sample of about 30,000 households (80,000 persons) will be selected, with the first stage being a sample of 2,000 enumeration areas, with 15 households chosen within each selected EA. The sample size has been chosen to meet the criteria that the estimate of net undercoverage will have a CV less than 25% for all provinces, assuming that the undercoverage rates would be similar to 1986 and the ratio of overcoverage to undercoverage would be at most, between 0.25 to 0.4. Under these assumptions, the CV for a gross overcoverage rate of 0.005, would be about 10% at the national level, with provincial CVs ranging from 20 to 40%. The sample will initially be allocated to the provinces proportional to the square root of their population size, assuming a constant overcoverage rate for all provinces. It will then be slightly redistributed in order to improve the reliability of the estimates for small political entities such as Prince Edward Island and the two northern territories.

The sampling frame will be census Visitation Records, a list of dwellings that is created by the enumerator for each enumeration area at the time of the delivery of census questionnaires. The actual selection will be carried out in the final phase of the collection operation in July and early August at regional Field Collection Units. Names, addresses, telephone numbers and basic demographic information will be transcribed from census forms to Overcoverage Study questionnaires. A telephone interview will be then be carried out to determine respondents' eligibility for the census, to obtain any other addresses where household members might have been enumerated, and to establish the usual place of residence on census day. There will be a trace of movers who are expected to be about five percent of the cases. Where necessary, there will be field follow-up for non-response cases and households without telephones. For non-contact, no-trace cases it is hoped that enough information can be obtained from a knowledgeable source to determine whether the household ever existed and resided at the sampled address on census day.

Where there is contact and the respondents are in scope for the census, the persons will be considered to have been correctly enumerated and no further matching is required. A search will be carried out in Head Office Processing for the cases where another address has been obtained. The census questionnaire at this alternative address will be checked to identify duplicate enumerations.

In the 1991 Census, an automated search and matching procedures will be used for both the Overcoverage Study and the Reverse Record Check (Statistics Canada, 1989). Names and addresses are not captured in the Canadian Census but addresses will be captured for the coverage studies. An automated geographic search will first identify the enumeration area of an address. This will be implemented by converting the postal code portion of the address to one or more EA numbers. If a postal code has not been provided, an automated procedure can be used to provide one. The second part of the search is an automated match on date of birth and sex of the respondent in the coverage study and the census enumerations in the EA or EAs that have been indicated by the automated geographic search. Based on tests, it is expected that 80-90 percent of the cases where a match exists can be found by the automated match. All matches will be confirmed by a manual check of names and addresses. Cases that cannot be resolved by the automated match will be sent to a clerical operation.

## **2.2 Advantages**

The planned methodology is less expensive and involves less risk than the alternative of a PES based on the American design. First, only one interview is required; there is no need for a followup to reconcile non-matches between a household listing established at the time of the survey and the census respondents. Second, matching to the census is only required for the cases where an alternative address has been obtained. This is expected to be between five to ten percent of our cases. As well, because we can obtain up-to-date telephone numbers from the census questionnaires, we can use a telephone survey methodology for the majority of the sample. This reduces the field costs and also means we can afford more callbacks, thus increasing the final response rate. Ninety-eight percent of Canadian households have telephones, and the telephone number, given by the respondent on the census form, can be a more accurate identifier of the sample than addresses which may be imprecise or names which can have many duplicates. Finally, this approach involves the least risk to meeting the basic objective, that of estimating gross overcoverage. Most of the aspects of an integrated PES



methodology are untested in the Canadian context and would have required more development time than is available before the 1991 Census.

### **2.3 Disadvantages and Risks**

The major disadvantage of this approach is that we have had to limit our objectives to measuring gross overcoverage and will not obtain additional information on the extent or causes of undercoverage. For 1991 the Reverse Record Check will remain the basic and only method of estimating gross undercoverage. Moreover, another disadvantage is that we will not be able to get any experience with an area-based PES.

In addition, there are some areas of concern with the methodology of the private dwelling survey itself that must be addressed. The first is that the transcription of data from the census questionnaires must be of high quality. To ensure this, quality control of this operation is planned. Secondly, because interviewing cannot be carried out until August or September, it may be difficult to trace census respondents who have moved, or to obtain reliable information to determine whether or not they existed or were fictitious (created by curbstoning). This problem would also have existed with the PES methodology with the timing of followup interviews to reconcile inconsistencies. Finally, we must be cognisant of the issue of respondent sensitivity and avoid giving the impression that we are checking up on the answers given by the respondent in the census.

Despite these concerns, it is believed that the methodology that has been selected is the best available at this time for a survey-based estimate of overcoverage in private dwellings.

### **3. Studies in Collective Dwellings**

The objective of the survey is to identify overcoverage of persons who were enumerated in private dwellings. We are also developing special procedures to estimate overcoverage of persons whose usual place of residence was in collective dwellings. These are classified into two types: "institutions", with hospitals, old age homes and penal establishments as examples, and "non-institutional" collectives such as student residences, hotels, YMCAs etc. In the 1986 Census about 450,000 persons, or about two percent of the Canadian population, were permanent residents of collectives.

The 1986 study was conducted in a sample of 39 dwellings drawn from only three types of collectives, namely, general hospitals, institutions for the physically handicapped, and jails. The estimated overcoverage rate was very high, (26 %), although much of this resulted from incorrect census procedures being carried out in one institution in the sample. Nevertheless, it is believed that overcoverage may generally be high in collective dwellings. In institutions, a reason may be that the census residency rule which specifies that persons should not be enumerated unless they have resided there for six months or more is not correctly applied. In non-institutional collectives, examples of populations at risk of overcoverage are persons in student residences or work camps who may be enumerated both there and at the private dwelling of a parent or spouse.

Census information on residents in institutions is obtained from records whereas in other types of collective dwellings the respondents themselves complete individual census questionnaires. In both cases, we have added a section to the census forms, asking for an alternative address, for example, a home address for a resident of a hospital. These responses will be sampled and the census questionnaire at the alternative address searched to see if the sampled person was enumerated there. We are also considering an automated match of residents in some institutions of the same type, in the same locality. Examples are residents of chronic care facilities or homes for the aged in the same city. This would identify persons who might have been transferred close to census day and were recorded as residents at each place.

#### **4. Automated Match**

As part of the 1986 experimental overcoverage check, a study was carried out to estimate the number of persons who were enumerated more than once in private dwellings in the same EA. This study was implemented by matching a sample of 50,000 persons with the persons in neighbouring households. The selection of households to be matched was determined by their closeness as listed in the VR, with the difference in their assigned household numbers not exceeding four. The estimate of 16,000 cases of double counting, that is, 36% of the total overcoverage estimate, suggested that this was a very promising method (Statistics Canada, 1990). Therefore the automated match will be expanded for the 1991 Census. The sample size will be doubled and the possibility of matching within an entire EA, and matching complete households is being investigated.



In 1986, the automated match was made on year and month of birth, sex and marital status. In 1991, 'day of birth' will also be captured for the first time, making it possible to match on the exact date of birth. We are carrying out developmental work to determine the most efficient matching strategy, given this more precise matching variable. The objective is to develop criteria that will decrease the number of false matches without being so rigorous that they give rise to too few true matches. We are also investigating the possibility of overlapping the automated match sample with the private dwelling study in order to derive a more precise overall estimate of overcoverage. The components of the Overcoverage Study complement each other, with the automated match identifying overcoverage resulting mainly from procedural causes. Alternative addresses which are beyond the matching area will be obtained from respondents by the survey, which in addition, is the vehicle for distinguishing cases of fictitious or out-of- scope enumerations.

## **5. Conclusion**

Although the feasibility of conducting a post-enumeration survey to measure net undercoverage directly was investigated, for 1991 the Reverse Record Check will continue as the sole method to estimate gross undercoverage. The plans for measuring overcoverage in the 1991 Census build upon what we learned in the experimental study of 1986 and are being developed to provide an efficient and complete way of measuring a very rare population. The chosen methodology should meet the basic objective of obtaining an accurate estimate of gross overcoverage in the 1991 Census. This estimate will be combined with the results of the Reverse Record Check to provide, for the first time, an estimate of net undercoverage in the Canadian Census.

## 6. References

- BURGESS, R.D. (1988), "Evaluation of Reverse Record Check Estimates of Undercoverage in the Canadian Census of Population," Survey Methodology, 14:137-156.
- CHILDERS, D., DIFFENDAL, G., HOGAN, H., SCHENKER, N., and WOLTER, K. (1987), "The Technical Feasibility of Correcting the 1990 Census" Proceedings of the Social Statistics Section, American Statistical Association, 36-45.
- FAY, ROBERT E., PASSELL, JEFFREY and ROBINSON, J. GREGORY (1988), The Coverage of Population in the 1980 Census, U.S. Bureau of the Census, PHC80-E4
- HOGAN, HOWARD (1986), "Weaknesses of the 1980 PEP and Solutions for 1990" paper presented to the Joint Advisory Committee of the Census Advisory and Population Statistics Committees of the American Statistical Association.
- JULIEN, C. (1990), "La possibilité d'estimer le taux d'erreur nette de couverture lors du recensement de 1991" Recueil des textes des présentations du colloque sur les méthodes et domaines de la statistique, 58<sup>e</sup> congrès de l'ACFAS, Bureau de la Statistique du Québec, (à être publié en 1991).
- ROMANIUC, A. (1988), "A Demographic Approach to the Evaluation of the 1986 Census and the Estimates of Canada's Population" Survey Methodology, 14: 157-171.
- STATISTICS CANADA (1989), "Integration and Automation Report: Head Office Processing", Internal Report, 19910-6-80.
- STATISTICS CANADA (1990), User's Guide to the Quality of 1986 Data: Coverage, Publication 99-135.
- STATISTICS CANADA (1990), "Evaluation of Options for the 1991 Post-Enumeration Survey", Internal Report, Social Survey Methods Division, Data Quality Project.



STATISTICS CANADA LIBRARY  
STATISTICS CANADA LIBRARY



1010257814

*c.2*

00S

FR 1

