Statistics Canada | Statistique Canada

Methodology Branch

Social Survey
Methods Division

Direction de la méthodologie

Division des méthodes
d'enquêtes sociales

Canada

WORKING PAPER NO. SSMD-87-009E

METHODOLOGY BRANCH

CAHIER DE TRAVAIL NO. SSMD-87-009E

MÉTHODOLOGIE

RESEARCH INTO A REGISTER OF RESIDENTIAL ADDRESSES FOR URBAN AREAS OF CANADA

by

J. Douglas Drew
John B. Armstrong
Ruth Dibbs

SSMD

May, 1987

* This is a preliminary version. Do not quote without author's permission. Comments are welcome.

Abstract

As part of the research program for the 1991 Census of Population and Housing, a study of the feasibility of creating an automated list or register of residential addresses for Canadian urban areas is underway at Statistics Canada. Initial work of the project team is reported in this paper. The methodology used to construct pilot registers using information from several administrative records systems is described. The results of a coverage evaluation study for pilot registers constructed for test areas in Ottawa are presented. Alternative uses of an address register in the census are discussed.

Résumé

Une étude de faisabilité concernant la création d'une liste automatiseé ou un registre d'adresses résidentielles pour les régions urbaines du Canada est actuellement en opération à Statistique Canada dans le cadre du programme de recherche pour le recensement de la population et des ménages de 1991. Dans cet article, on discute des travaux amorcés par l'équipe de projet. On décrit la méthodologie utilisée lors de la construction des registres pilotes créés à l'aide d'information recueillie sur divers fichiers administratifs. Les résultats d'une étude d'évaluation de la converture des registres pilotes construits pour des régions tests à Ottawa sont présentés. On discute également des utilisations alternatives d'un registre d'adresses pour le recensement.

1. Introduction

As part of its programme of research and testing in preparation for the 1991 Census of Population and Housing, Statistics Canada has initiated research into the creation of an automated list or register of residential addresses in urban areas of Canada, and into the investigation of uses of such a list in the conduct of the census.

In order to set the stage for the discussion of address register research we briefly describe the current census collection methodology.

Currently in taking the Canadian Census, address lists are compiled manually by approximately 40,000 Census Representatives (CRs) each responsible for an Enumeration Area (EA) containing 200-300 dwellings. This listing of addresses in a control document (the visitation record) is carried out co-incident with drop-off of census questionnaires. In urban areas, the questionnaires are mailed back to the local CR. At present the questionnaires are picked up by the CR in rural areas, although consideration is being given to mailback for these areas as well in 1991.

The CR is responsible for basic edits ensuring completeness and consistency of the data, for telephone or personal follow-up of edit failure cases, and for personal follow-up of nonresponding households. When completed, work for the EA is returned via a supervisory network to regional sites where data capture takes place. The EA remains a unit for control during data capture and subsequent processing steps. While address ranges for block faces are data captured, individual addresses are not.

As can be seen from the preceding description, the initial steps in census taking in Canada may be characterized as being manual, highly decentralized, and heavily reliant on the quality of work by a large staff of temporary employees. An automated address register (AR) would be a pre-requisite to increased automation and centralization, which on the surface would appear to have good potential for reducing costs and increasing quality. An AR either in combination with increased automation and centralization of front end census steps, or on its own also has the potential for reducing the dependence of census coverage on the temporary field staff. Use of an AR may emerge as an effective means for the agency to deal with uncertainties regarding the future availability of a qualified short term work force. Automation and centralization are being investigated separately in other 1991 Census research (Royce and Pryor 1987). An AR also has potential to yield better coverage of dwellings than the current methodology.

Since the topic of AR research at Statistics Canada is not a new one, a brief description of previous work follows below as background to the current work.

As part of the research programme for the 1971 Census, Fellegi and Krotki (1967) constructed and evaluated address registers for two medium sized cities: Kitchener-Waterloo and London. In addition to consideration of an AR as a vehicle for conducting a mail-out census, other potential uses forseen at this time included use of an AR in automated assignment of geographical co-ordinates to census data, and use as a sampling frame for household surveys.

The approach followed by Fellegi and Krotki in the Kitchener-Waterloo test was similar to that currently under investigation; viz, creation of a register from the merging and unduplication of address information from multiple administrative data sources. In their case, sources included municipal assessment rolls, the 1961 Census lists of households, and electricity billing lists. Due to technological limitations of the day, construction of the register was largely a manual process.

Interestingly, a 97% coverage of addresses was obtained for the AR created in this fashion, which is comparable to current results as described in Section 4. This level of coverage implied that a field check, either by Statistics Canada personnel or by post office letter carriers, would be required to improve coverage prior to use of the AR in a mail-out census. It was decided not to proceed with AR development for the 1971 Census.

During the 1970's, a series of studies, summarized by Booth (1976), were conducted concerning creation and maintenance of an AR for use in mailing out the 1981 Census. The approach considered in this case involved initially data capturing address lists from a previous census, with updates based on information supplied by Canada Post. It was found that such a register yielded dwelling coverage comparable to that of traditional census methods. However, the high initial data capture costs were viewed as problematic despite estimated longer term savings, and the research was terminated.

It was decided to re-examine the feasibility of an AR for the 1991 Census, since it was felt that environmental factors are now more conducive to such an undertaking than was the case in previous decades. Royce (1986) has enumerated these factors, and identified research issues and a time frame for their examination relative to decision points for the 1991 Census. The environmental factors include the increased availability of machine readable administrative record systems with address information, the widespread use of postal codes which facilitates linkage of address information and standard geography; the increased power and decreased cost of computers; and the availability of improved record linkage methods and software.

Remaining sections of the paper include: the methodology being used for AR construction (Section 2), the evaluation of pilot registers in Ottawa and Vancouver (Section 3), a description of possible uses of an AR in the census (Section 4) and directions of further research (Section 5).

2. Methodology for Address Register Construction

In this section the methodology used to construct pilot address registers for test sites in Ottawa and Vancouver is described. Two pilot registers were constructed in each city - one for comparison to the results of the 1986 Census and another to enable us to evaluate coverage of an address register relative to that of the Canadian Labour Force Survey (LFS). The methodology employed is similar to the approach investigated by Fellegi and Kortki (1967). It involves the merging and unduplication of adddress information from several administrative sources.

Based on the field evaluation of the coverage of the pilot registers, which will be reported in the next section, and the resource requirements of the methodology described here, we expect that a similar procedure could be used on a national scale. Before the methodology is reviewed the issue of address register contents is examined and the administrative sources used for the pilot tests are mentioned. Subsections 2.3 and 2.4 contain more detailed discussions of address standardization and record linkage, two methodological problems involved in address register construction. The section concludes with a list of possible enhancements to the construction methodology, suggested by the pilot study experience, that may be included in future address register work.

2.1 Address Register Contents and Administrative Sources

Dwelling address is of course the most important item that should appear on an addresss register. Depending on the eventual uses of a register it may be appropriate to include both physical location and mailing addresses. Failure to differentiate between these two addresses or appropriately link them would lead to significant overcoverage. From the operational point of view, obtaining information needed to link location and mailing addresses is difficult. Among the administrative files used in the pilot tests, only the municipal assessment files included both location and mailing addresses corresponding to the same dwelling. However, the mailing address referred to the property owner rather than the resident.

Only location addresses were included on the pilot registers. This choice is the appropriate one for an address register that will be used in conjunction with the current census delivery methodology - either to provide preliminary lists or as a coverage improvement tool. Each address was stored on the register in two forms. In addition to the free format version obtained from the original administrative source an address search key was stored. The search key is a standardized version of the address produced by Statistics Canada software and used for exact matching and record linkage purposes. Other contents of the pilot registers included the name and creation date for the administrative file from which the free format version of the address was obtained and information about other files that included addresses for which the same search key was generated. Telephone number and dwelling type were also stored when available. The Federal Electoral District - Enumeration Area number, providing a link to census geography, was stored on those pilot registers constructed for census test sites.

Three nationwide administrative files - Revenue Canada, Family Allowance and Old Age Security - are made available to Statistics Canada on a regular basis. For the pilot tests these files were supplemented by additional files purchased specifically for the project. In Ottawa the telephone company's file of billing addresses for listed numbers was used and in Vancouver, the electrical utility company's file of billing addresses was employed. Municipal assessment files, containing mailing addresses for assessment notices, were also used in both test cities.

2.2 Steps in Address Register Construction

Before the address register construction process began, lists of postal codes corresponding to the test areas were obtained. For the census tests, this was done using Statistics Canada's files linking postal code and census geography. For the Labour Force Survey tests, postal code information obtained by LFS interviewers was verified by clerical staff.

Construction of the pilot registers using these postal code lists involved five steps:

- (i) Selection of addresses that should correspond to test areas from administrative files using exact matching on postal code.
- (ii) Address standardization (creation of search keys).
- (iii) Elimination of exact duplicates.
- (iv) Postal code verification.
- (v) Elimination of probable duplicates using record linkage techniques.

Given that exact matching on postal code is used in the first step of this procedure, a high incidence of correct postal codes on the administrative files is of obvious importance for the methodology. Fortunately, due to delivery time and postage rate incentives offered by

Canada Post, the postal codes system is in almost universal use in Canada. On most of the administrative files used in the study, a postal code appeared on over 99% of the addresses. Accuracy studies for various files have indicated rates of correct postal code use considerably higher than 90%. In addition, most errors occur in the final three digits of postal code, which determine the exact location of an address within a relatively small geographic area. Cases in which an address contains a postal code correponding to another city are relatively uncommon.

The second step of address register construction involves the use of address standardization software developed at Statistics Canada. The software accepts free format addresses and constructs an address search key including the following components:

(i) province
(ii) municipality
(iii) street name
(iv) street number
(v) street number suffix
(for example, A in the case of 10A Main St)
(vi) street type
(vii) street direction
(viii) postal code
(ix) postal designator type
(box, apartment and rural route are examples)
(x) postal designator number.

More details concerning the address standardization process and developmental work currently in progress are given in subsection 2.3.

In the third address register construction step, the addresses that have been selected from the various files are merged and those with search keys that agree exactly for street name, street number, street number suffix, apartment number and postal code are eliminated. In the next step, postal codes for addresses on the merged file are validated using Statistics Canada's postal coding software and those that are assigned a postal code outside the test area are eliminated.

After the elimination of exact duplicates and addresses with postal codes outside the test area, an address register may still contains some address pairs that refer to the same dwelling due to minor errors on the administrative files. An example of such a pair is

Apt 1604 1591 Riverside Dr Ottawa Ont K1G 4B7, and

(1)

Apt 1604 1591 Rivesride Dr Ottawa Ont K1G 4B6.

The final step of address register construction involves elimination of such duplicates using record linkage techniques. For the construction of pilot registers in Ottawa and Vancouver, address search key fields were used for record linkage. A discussion of the record linkage methodology can be found in subsection 2.4.

For each pair of duplicate addresses identified by record linkage, the address to eliminate was chosen automatically using information on the address register concerning the number of files on which the address appeared. If one address appeared on fewer files than the other, the address that appeared least frequently was deleted. When each address appeared equally often, the choice of record to eliminate was made randomly.

2.3 Address Standardization

The address standardization software used for construction of the pilot address registers was developed at Statistics Canada during the 1970s. It accepts free format address information and identifies various components. The initial step of the algorithm involves scanning the input string and breaking it into "tokens". Each token is either a numeric or an alphabetic string that does not include blanks. For example, the address

Apt 1604 1591 Riverside Dr Ottawa Ont K1G 4B7 (2)

includes 9 tokens. Tables are searched in a attempt to identify the alphabetic tokens. In this case Dr is a street designator, Apt is a postal designator, Ont is the province and the other strings cannot be identified from the tables. In the second stage, the numerics and the alphabetic tokens that were not found in the tables are identified by matching a series of complete address patterns to the available information.

A new address standardization package is currently under development. The new software, which is written in MPL, a compiler generator, offers a number of advantages over the package now available, including improved alphabetic token tables and a parsing algorithm that uses approximately 100 production rules to process various parts of the input address. Standardization success rates for urban addresses are considerably higher with the new package than with the old software. For example, in a test using a systematic nationwide sample of 357 urban addresses from the Revenue Canada file the new software success rate was 98% compared to 92% for the old package.

Unlike many other syntax parsing problems, such as mathematical equation parsing, address standardization is a process for which the correct answer is sometimes unknown. Consequently, certain cases will cause difficulty for any software package. Some examples include

(i) 10 Main St Fort St John BC V6E 2G7, and

(ii) 10 Paul Anka West Pickering Ont M9Z 2E1

In the first case the street designator St appears in the city name "Fort St John". In the second case it is not clear whether West is a street direction or part of the city name. The new software identifies addresses like (i) and (ii) as ambiguous and consults tables of municipality names in an attempt to resolve them.

The addresses

(iii) Apt A 10 Main St Ottawa Ont K2A 9B1, and

(iv) 10A Main St Ottawa Ont K2A 9B1

refer to the same dwelling. When the new software finds address (iii) on an administrative file it generates two records - one with "A" as an apartment designator and a synthetic duplicate with "A" as a street number suffix. If address (iv) is later found on another file it can de identified as an exact match with (iii) by means of the synthetic duplicate. More details concerning address standardization work at Statistics Canada can be found in Armstrong et al (1987).

2.4 Record Linkage

Record linkage work involved in address register construction was done using Statistics Canada's Generalized Iterative Record Linkage System (Hill and Pring-Mill 1985), which is a record linkage package based on the methodology proposed by Fellegi and Sunter (1969). The Fellegi-Sunter methodology involves assigning weights to each pair of records according to the probability that the addresses both refer to the same dwelling. The total weight, W, for the record pair (a,b) is obtained by adding weights for individual linkage fields. That is,

$$W=w_1+w_2+\dots w_k, \tag{3}$$

where each w_i is a log-odds ratio,

$$w_i=\log\ (P(O_i)|(a,b)\in M)\ /\ P(O_i)|(a,b)\in U). \tag{4}$$

M is the set of address pairs that refer to the same dwelling (true matches) and U is the set of pairs that do not refer to same dwelling (true non-matches). O_i is the outcome of the comparison of linkage field i for addresses a and b. It may be agreement, disagreement or some form of partial agreement.

In practice, since the sets M and U are unknown for the file of addresses of interest, so are the conditional probabilities that appear in (4). Estimates of the weights can be obtained if a calibration file is available for which the sets M and U have been determined manually. Creation of such a calibration file is a process that is both expensive and somewhat error-prone. Recently

Jaro (1986) has experimented with a method of weight calculation suggested by Fellegi and Sunter that does not need a calibration file, but instead, involves solution of nonlinear equation systems.

The Generalized Iterative Record Linkage System (GIRLS) uses another procedure that avoids the calibration file requirement. Initial weights are calculated assuming that the linkage fields on the file of interest are free of errors. Under this assumption weights can be computed using the frequency of occurrence of various values of each linkage field. These weights are used to rank address pairs that show some evidence of agreement, that is, might be true matches. This list of pairs is examined by the user who makes a judgement concerning a "cut-off" weight. Pairs with a weight higher than the cutoff are temporarily considered to be true matches. Estimates of error rates for the linkage fields are calculated using this set of pairs. These estimates are used to adjust the initial weights and the process of ranking pairs and selecting a cut-off is repeated. This procedure can be continued until differences between consecutive iterations are small.

In order to reduce the number of comparisons between address pairs, GIRLS allows the user to specify one or more blocking fields. Each set of values for the blocking fields defines a "pocket". During the iterative process only addresses in the same pocket, that is, addresses for which all the blocking fields agree, are compared.

In pilot AR construction, address search keys fields were used as input to GIRLS. Street number and apartment number were used as blocking fields. Partial agreement outcomes were defined for street and city names using a simple character string comparison algorithm that allows for character transposition and as well as random insertion and deletion of characters. For postal code, partial agreement outcomes were defined based on the number of characters that agreed and the positions of those characters. According to these definitions, the addresses

Apt 1604 1591 Riverside Dr Ottawa Ont K1G 4B7, and

(1)

Apt 1604 1591 Rivesride Dr Ottawa Ont K1G 4B6 partially agree on both street name and postal code.

2.5 Future Methodological Improvements

Future address register construction work will involve some enhancements to the methodology described here, such as

(i) use of the new address standardization package currently under development,

(ii) use of telephone number in record linkage (names may also be used but they will not be included on the final register),

(iii) matching on phonetically encoded versions of street and city names, and

(iv) imputation for apartment buildings (if a building includes units 101, 201 and 401 we have strong evidence that 301 has been missed)

3. Evaluation of Pilot Address Registers

In this section results of the evaluation study for the pilot registers constructed for test sites in Ottawa are reported. Some details concerning the test sites and relevant LFS and census procedures are given in subsection 3.1. The methodology used to evaluate the pilot registers is described in subsection 3.2. In addition, overall coverage, coverage by structural type of dwelling, and the potential for improving LFS and census coverage using an address register are examined. Lessons learned about the feasibilty of constructing a national address register using the pilot study methodology are summarized in the final subsection. A similar evaluation study is in progress for Vancouver test sites and will be described later.

3.1 Ottawa Test Sites

Two pilot registers were evaluated in Ottawa. The first register was constructed for areas covered by active Labour Force Survey clusters.

Clusters, which consist of small numbers of adjacent block faces, are the penultimate LFS sampling units (households are the ultimate units). A cluster is considered active if households from the cluster are included in the current LFS sample. Once a cluster becomes active it remains active until all dwellings in the cluster have been included in the LFS sample. This process typically takes at least a year and can require several years. When a cluster becomes active it is visited by an LFS interviewer, who produces a dwelling list without making contact with residents. This list is updated if an LFS interviewer, making contact with the residents of a dwelling for survey purposes, discovers that the same building contains dwellings that were missed during the original no-contact listing. (In fact, dwellings found in this manner are included in the sample.) It is also updated to allow for new construction. However, demolished dwellings are not removed from the lists. All the clusters in the Ottawa test were listed after October 1984. Only clusters from the Ottawa area frame were included in the test. This frame excludes apartment buildings with more than four floors and more than thirty units.

While the area covered by the LFS pilot register represents a probability sample of block faces, the second Ottawa pilot register constructed for four tracts from the June 1986 census was a purposive sample. Three of the census tracts selected were from older central parts of the city, and the fourth was from a very affluent area. They were chosen because they represented types of areas in which census coverage has traditionally been below average.

The methodology described in section two was used to construct the pilot registers. Five administrative files were used - Revenue Canada, Family Allowance, Old Age Security, the telephone company's billing list for listed numbers and the municipal assessment billing lists. According to the LFS cluster lists, the dwelling count for the LFS test area was 3655. The 1986 Census dwelling count for the census test area, based on the visitation records (control forms completed by census enumerators) was 5954. Dwelling counts for the pilot address registers constructed for the two areas were 3670 and 6643 respectively.

3.2 Coverage

The first step in evaluation of each pilot register involved matching the register to the dwelling lists from the alternative source. Addresses on the LFS lists and the 1986 Census visitation records were standardized and postal codes were assigned. Both exact matching and record linkage techniques were used to match these lists and the pilot registers. Unmatched addresses from the LFS lists and the visitation records were checked and postal coding and standardization errors were corrected manually. The matching process was then repeated. For each test area, it was assumed that all matched dwellings were valid. A field check was conducted to resolve unmatched addresses. For each test site a list of street names and numbers that corresponded to unmatched addresses from either the address register or the alternative source was determined. Head office personnel were asked to visit each of these street numbers and report the total number of associated dwellings. Budgetary restrictions did not permit a 100% enumeration of the test areas. However, the field check methodology did result in verification of some dwellings that did not directly correspond to unmatched addresses. For example, when a street number for an apartment building was included in the field check the total number of units in the building was reported, even if only one or two units corresponded to unmatched addresses. The personnel conducting the field check were instructed to avoid contact with dwelling residents.

The results of the matching operation are reported in Table 1. The match rates in this tables are calculated relative to our "best guess" at the true list of dwellings, based on the sets of matched dwellings and the field check results. Some dwellings that did not appear on either the address register or the alternative list may be excluded from this "best guess". In the case of the census test 4660 of the 5942 dwellings in the test area (78.4%) were found as a result of this matching operation. The match rate for the LFS pilot register was 90.2%.

Coverage information for the pilot register and the alternative list for the two test sites are given in Table 2 (LFS test) and Table 3 (census test). The tables also include information about the coverage of the LFS and the 1986

Census for the test areas. All cases in which more than one dwelling is associated with a street number are classified as multiples so this category includes townhouse developments as well as apartment buildings.

Before we comment on the coverage of the address registers, some remarks concerning LFS and census coverage rates are appropriate. The LFS single family undercoverage rate of 0.5% is similar to results obtained using a listing check incorporated in the LFS re-interview program (Cyr 1984). This study reported a national LFS undercoverage rate of 0.44% (averaged over all dwelling types). The multiple undercoverage rate of 5.3% reported in Table 2 should be carefully interpreted. The rate of LFS multiple detection at the time of interview is substantial. Multiples missed at the time of listing typically account for between 1% and 1.5% percent of the total LFS sample. For the LFS test, the missed multiples amount to 1.9% of the total dwelling count. If one assumes that none of the dwellings in the cluster lists have been included in the LFS sample, the multiple undercoverage rate obtained here is similar to Cyr's national average rate. [Information concerning the number of dwellings in the cluster lists that have been included in the LFS sample was not available in time for inclusion in this draft.]

The census undercoverage rate of 3.7% appears somewhat high when compared to certain undercoverage rates for the 1981 Census estimated using a reverse record check methodology (Burgess 1983). For example, the 1981 reverse record check estimate of national undercoverage for urban households (excluding some remote areas) was 1.71%, with a standard error of 0.13%. However, one must remember that the census tracts included in the test were selected because relatively poor coverage was expected. Over 47% of the dwellings in the test area (2812 out of 5942) were either apartments in buildings with less than five storeys, townhouses, or duplexes. Information concerning the number of townhouses in the test is not available but, given the types of areas included, we expect that there were very few such dwellings. Census undercoverage for the other two structural types has traditionally been relatively high. The 1981 reverse record check estimates were 5.81% for apartments in small buildings and 3.02% for duplexes.

By far the largest factor contributing to the overall LFS overcoverage rate of 5.5% was cases in which the cluster lists indicated more dwellings at an address than did the field check. Such situations accounted for 124 of the 195 units incorrectly listed by the LFS. Discrepancies of this magnitude underline the difficulties associated with determining the number of dwellings in certain types of small multiples using a no-contact field check. Demolitions, which accounted for 51 dwellings incorrectly listed by the LFS, was the next largest source of overcoverage.

The census overcoverage rate of 3.9% is surprising. It is important to note, however, that the field check was conducted in March 1987, ten months after the 1986 Census. In addition, each address listed on census visitation records was considered a dwelling covered by the census, regardless of whether or not a census questionnaire was returned from that address. A breakdown of the various factors contributing to the 3.9% is given in Table 4. As one might expect, cases in which the census visitation records listed more dwellings at a street number than our field check are the most important source, accounting for over 41% of the dwellings apparently incorrectly listed in the 1986 Census. The second most important category is "Incorrect EA/Invalid Address" which accounts for over 31% of overcoverage. We expect that some dwellings listed in this category should have been included in "Demolished". [Further study of this overcoverage is in progress. For example, census questionnaires will be examined to determine how many of the overcovered dwellings returned questionnaires. Results will be reported in a later draft of this paper].

The overall address register coverage rates of 96.0% and 95.5% for the LFS and census tests, respectively, are comparable to the LFS and 1986 Census coverage rates of 97.7% and 96.3%. An analysis of the factors contributing to address register undercoverage for these tests is not available, although some of the undercoverage can be attributed to problems with certain computer procedures involved in address register construction - postal code verification, address standardization and record linkage. Some information concerning coverage errors due to the address standardization software is reported in the next subsection.

Note that the overcoverage rate of 7.0% for the address register constructed for the LFS test area is considerably lower than the 16.3% rate for the census test register. Part of this overcoverage difference is due to the fact that the address register constructed for the LFS test was verified manually while the census test register was not manually checked. Table 4 contains information about the various reasons contributing to address register overcoverage for the census test. Cases in which the address register listed more dwellings at a street number than our field check are the most important factor contributing to overcoverage (54.3% of overcoverage). Many of the dwellings in the "Incorrect EA/Invalid Address" category (26.7%) may be included due to postal coding errors that were not detected by Statistics Canada's software. Analysis of the contributions of the five administrative files to the address register indicates that the telephone company file, which includes relatively unreliable postal codes, is the largest single source of dwellings in this category. The "Nonresidential" overcoverage (16.7%) is apparently due, to some extent, to individuals who prefer to use the address of their place of employment for correspondence concerning money matters and may be difficult to reduce.

Use of an address register as a coverage improvement tool in conjunction with the current field operations is one of the options that is currently under consideration for both census and the LFS. In view of this possibility, the information reported in Table 6 is quite promising. For the Ottawa LFS test area, use of an address register would reduce undercoverage on initial no-contact dwelling lists from 2.3% to 1.4%. For the census test area, the coverage improvement potential is greater. Use of an address register would have reduced undercoverage in the test areas from 3.7% to 1.5% assuming a comparable success in a production mode. This represents almost a 60% decrease in census undercoverage. It must be cautioned that with full scale implementation in a census, the reduction in coverage could be less. For example if implementation consisted of matching an independently derived VR with an AR, the matching would be a manual operation, with resolution of discrepancies by temporary staff, unlike the test which involved automated matching and use of experienced staff to resolve discrepancies.

3.3 Feasibility of National Application of the Methodology

The address register coverage information reported in the previous sub-section suggests that an address register could be used in the census and Statistics Canada's household survey programs to improve coverage as a supplement to the current field operations. The coverage improvement results of 0.9% and 2.2% for the LFS and census tests, respectively, are very encouraging. However, the overall address register coverage rates in the 95-96% range are not high enough that one would like to advocate use of a register on a national scale as a substitute for field listing.

Address register coverage for multiple dwellings was not as good as single family dwelling coverage. Booth (1976) reports similar results. Much of the undercoverage and overcoverage for multiples was due to problems with the address standardization software used to construct the pilot registers. The software rarely recognized words like "basement", "upper" and "lower" when they were used to identify multiple households at the same address. In addition, apartment numbers were often eliminated and minor variations in address syntax between administrative files sometimes resulted in two different sets of search keys for the same address. Since these search keys included different apartment numbers, they could not be eliminated during unduplication of the address register. A small study involving manual verification of unstandardized addresses from administrative files suggests that use of the new address standardization software currently under development would improve address register coverage by between one and two percent for the LFS test area. There is more potential for improving address register coverage using this software for the Ottawa census test area, which includes a higher proportion of multiple dwellings. One should note that, even if the new address standardization software leads to a coverage improvement of only 1%, the coverage of an address register constructed using this software will be better than 1986 Census coverage for the Ottawa test area. [Later drafts of the paper will include results using the new software.]

The use of more administrative files might also improve address register coverage. Depending on the coverage impact of improved address standardization, future testing work may involve experimentation with address registers constructed using more than five administrative files.

The Area Master File (AMF) is a file maintained by Statistics Canada containing detailed street network information for urban areas. Approximately 60% of Canada dwellings are in areas included in the AMF, which is used during the delineation of enumeration areas and the creation of maps for use by census enumerators. Excluding time spent by professional personnel, there would be three main cost components associated with construction of an address register for all Canadian urban areas covered by the AMF:

(i) computing costs,

(ii) file acquistion costs, and

(iii) clerical costs for manual verification.

On the basis of the pilot study experience, our estimate of the computing cost of constructing an address register on a national scale using the pilot project methodology is $100,000. For this level of expenditure a sequential tape file could be produced. Integration of an address register with a database management system would involve additional costs.

Administrative file prices vary substantially and a file acquisition strategy suitable for construction of a national register has not yet been completely worked out. On the basis of information available thus far, we estimate that the data acquisition cost of constructing a national register using five files in all areas would be between $100,000 and $150,000. The pilot study experience clearly indicates that a register constructed using only the three national files currently made available to Statistics Canada on a regular basis - Revenue Canada, Family Allowance and Old Age Security — would not provide adequate coverage. Coverage of address registers constructed using only these national files would be less than 90% for both Ottawa tests. The possibility that use of a single additional file may be sufficient to produce an address register with adequate coverage prior to any field work for certain areas of the country will be investigated in future testing work.

Any address register constructed on a national scale would be partially verified by clerical staff before it is used in the field. The difference between overcoverage rates for the LFS and the census pilot registers indicates that a manual verification operation would significantly improve the quality of an address register. We estimate that an operation involving the computerized identification of problem cases and their manual resolution would require approximately 10 person-years of clerical time. Problem cases that could be checked in such an operation include block faces with only street number, street numbers with only one apartment, and addresses that are missing important standardized components (such as street number).

4. Potential Uses of an Address Register in the Census

In this section we enumerate possible uses of an address register in the conduct of the census. We do so for the most part at this juncture without reference to strategy or recommendations regarding implementation. As these uses are examined in detail later this year, we will be assessing factors such as cost and quality implications, requirements for development and testing, and the degree to which individual uses can stand alone or are interdependent. Recommendations will be made on what uses (if any) to make of an AR in 1991 on a production basis, what testing (if any) of uses of an AR should be imbedded into the 1991 Census, and for which uses (if any) development and testing should be deferred until after the 1991 Census. To be recommended for the 1991 Census, uses will have to satisfy criteria of low risk, adequate time to develop and test, and better or equivalent dwelling coverage at reduced or equivalent costs with a net overall gain from a combined cost and quality perspective.

Potential census uses are described below as they relate to delivery of questionnaires, follow-up of nonresponse and edit failures, control of census documents in processing, and geography operations both in preparation for the census and during processing of the census. More details may be found in Dibbs et al (1987).

4.1 Use of Address Register In Delivery of Census Questionnaires

On the previous occasions when AR research was carried out at Statistics Canada, the primary use envisioned was as a vehicle for conducting a mail-out census. We decided this time to investigate as well other options for use of a register in collection that could be integrated more easily into existing processes, and hence whose implementation would pose less risk to already functioning systems. Given constraints on both time and resources, a priori such options would appear more attractive for the 1991 Census. If warranted, uses involving more fundamental changes to the census could then be extensively tested and developed for implementation at a later census.

The alternatives we have identified for use of an AR in delivery of questionnaires are:

(i) To permit a mail-out census

A cost comparison for urban areas of the current list/drop-off methodology versus mail-out from an AR was conducted by Gamache-O'Leary, Nieman, and Dibbs (1987). The intent was to examine under a conservative set of assumptions whether there was evidence of a clear cost advantage for a mail-out census, in order to identify at as early a juncture as possible in the research whether mailing-out should fit prominently in plans for 1991.

In comparing costs, the largest unknown at this time is the cost of creating a national AR with dwelling coverage comparable to that obtained with the traditional method. The analysis revealed that if the AR created directly from administrative sources achieved comparable coverage, then there would be a clear cut cost advantage with a mail-out census. On the other hand, if one field check of the register were required, this would represent a rough break-even point in terms of costs. If two field checks were required, say one by Statistics Canada personnel and another by Canada Post letter carriers, then a mail-out census would be more costly.

The requirement for two field checks may be less conservative than realistic; for example, the U.S. Bureau of the Census carries out two field checks of their initial address list before mail-out in conducting the U.S. Census.

Hence it was decided at an early juncture to rule out a general mail-out for 1991 and to concentrate on other uses of an AR in collection for 1991. Nevertheless if an AR can be created and maintained at a high level of accuracy intercensally at a reasonable cost, then a mail-out may become an attractive option worth reconsidering for 1996.

Indeed, if it can be argued that the role of the census is to estimate the population, as opposed strictly to count it, a dual frame approach with a nearly complete AR and an inexpensive mail-out methodology, supplemented by an area frame for coverage improvement might be feasible and more cost effective either than attempting to create a complete list across the board or than conducting a traditional census.

On a more limited scale a mail-out has still not been ruled out for 1991 in high rise apartment buildings, where it could circumvent the increasing problem of gaining physical access to specially secured buildings.

(ii) Pre-list with automated VR

Under this alternative, an address list would be pre-printed from the AR for each Enumeration Area. The Census Representative would update the list at drop-off. Any additions or deletions (whether due to demolition or other reasons) would be recorded by the CR for data capture. Timing of the data capture would depend on other factors such as the extent of centralization of later operations.

As elaborated on later, an automated VR with capture of CR changes would fit best with scenarios for greater centralization of follow-up activities and an automated control file. It would also permit closer integration of AR based household surveys and the census either in the sample design, or for the purposes of data quality and coverage comparisons, etc. Also the capture of changes made by the CR's would improve the coverage and accuracy of the AR, which could benefit other applications such as for household surveys or for the next census.

A concern with giving the CR a nearly complete list in advance, is that a poor job of finding additions on the CR's part could translate into worse coverage than that with the current methodology. This issue will be investigated in testing planned for the fall of 1987 (see section 5).

Four fifths of enumerated Canadian dwellings receive a short questionnaire in the census. Other dwellings receive a long version. Currently, the dwellings that receive a long form are selected by the CR using systematic sampling from the dwellings listed in the VR. If the CR determines that a dwelling is vacant at drop-off it is excluded from the sampling process. Under the automated pre-list alternative, further consideration needs to be given to the relative merits and demerits of: (1) automated pre-sampling of dwellings on the AR, with sampling of additions by the CR; versus (2) having all the sampling done by the CR.

Under (1), the CR would be required to: enter codes for various categories of unoccupied or erroneous dwellings on the list; add additional dwellings to the bottom of the list; and apply a systematic sampling rule to sample the additions.

Under (2), a way of proceeding would be to have two columns set aside on the VR for the CR to assign household numbers separately to occupied versus un-occupied dwellings. The CR would assign dwelling numbers as the EA is canvassed (with a random starting point between one and the sampling interval as is currently done). Additional dwellings would be added to the bottom of the list, but assigned a number corresponding to where they were found. The CR would follow the same procedures as are currently used to select every kth occupied dwelling; i.e. with the current sampling fraction every occupied dwelling whose number ends in 0 or 5 would be sampled.

Advantages of (1) are that it would reduce the scope for CR errors in sampling, and it would permit a better sample design, for example, gearing the sampling rate in an area to factors such as the size of the area.

Disadvantages of (1) are that it would require development of new systems modules for sampling from the AR and for verification of sampling of additions by the CR, it would impact on existing processing systems which expect dwelling numbers in different ranges for occupied versus unoccupied dwellings, and it would lead to some increase in sampling variance due to the inclusion of more vacants as well as non-existent dwellings in the sample.

Advantages of (2) are that changes to census processing systems would be minimal, and the sampling variance would not be inflated as under (1).

A disadvantage of (2) is that the procedures for the CR to follow would be more complicated than current procedures or those under (1), which could translate into errors in sampling.

(iii) Pre-List with manual VR

A low risk application in collection would be to issue the CR's a computer generated address list for their EA, which would be used to facilitate the task of creating a manual VR. Changes to existing census procedures and systems would be minimal, and the pre-list would help to standardize the quality of resulting VR's, and permit use of labels to facilitate later processing steps (see sections 4.2 - 4.4). Also labels distinguishing between pre-listed dwellings and CR additions would facilitate updating of the AR to reflect census additions, which could be done as an activity outside the normal stream of processing the census.

A disadvantage of this alternative is that it would not be conducive to centralized nonresponse and edit follow-up, or to an automated control file.

Nevertheless due to its low risk and ease of implementation, it would seem to be a good candidate for use of an AR in 1991 provided the fall 1987 field testing demonstrates a beneficial impact on coverage.

(iv) Post drop-off coverage check

An alternative use of an AR in collection with even lower risk would be only to use the AR after the fact as an independent check on the quality of the manual VR's.

Three procedures which are not mutually exclusive are possible. First, there could be a macro check of drop-off totals against the number of dwellings on the AR to determine if coverage at drop-off is acceptable. Second, a sample of addresses from the register could be matched against addresses on the VR as acceptance quality control of the CR's drop-off work. Third, there could be a detailed comparison of the AR against the VR to identify missed dwellings. A small scale test of the third option in the 1986 Census reported by McCormick (1986) indicated that this could be a more cost effective method of coverage improvement than a similar check performed in the 1981 Census using post office lists and staff (Copp 1982). This use would have the benefit of providing an independent check of the quality of CR's work, and its impact on coverage can only be a positive one. However clearly such a use would be more costly and time consuming than using the AR as a pre-list. The matching of VR's and the AR's would be a manual operation since the VR addresses are not data captured and resolution of discrepancies would generally require additional field work. At issue is whether this approach would yield better coverage than use of an AR as a pre-list, and whether coverage gains (if any) would be worth the additional cost. The field testing scheduled for the fall of 1987 will address this issue.

4.2 Follow-up for Nonresponse and Edit Failures

Currently follow-up for nonresponse and certain edit failures is carried out locally by CR's. In 1986, 13% of returns required nonresponse follow-up. In nonresponse follow-up the CR is encouraged to use the telephone if a name was obtained at drop-off which can be found in the local telephone directory. However, since contact rates at drop-off have fallen to as low as 40% in urban areas, increasingly this information is not available and personal visits are required.

The availability of telephone numbers on the AR for the roughly 80% of households with published numbers would greatly facilitate nonresponse follow-up by telephone. Even though for many of the edit failure cases telephone number would appear on the questionnaire, the AR would be of some residual benefit. The greater use of telephoning could be achieved while continuing to do the follow-up decentrally under either of the pre-list delivery alternatives.

However the pre-list automated VR delivery alternative would lend itself to centralization of the nonresponse and edit follow-up, for those cases which can be handled by telephone. Potential benefits of such centralization would appear to be improved quality due to the use of experienced staff with more direct supervision, and possibilities for automation of processes; for example, in the longer term the use of Computer Assisted Telephone Interviewing.

As described by Royce and Pryor (1987), centralization of nonresponse and edit failure follow-up has been identified as another of the research and testing topics for the 1991 Census, with plans for development and testing closely linked to the AR research.

4.3 Automated Control File

An automated control file (ACF) would be a file containing information on the status of each dwelling for which a census questionnaire is delivered. An ACF would go hand-in-hand with mail-back of census questionnaires to central sites, and centralization of nonresponse and edit follow-up. Such a scenario would permit greater control and monitoring of those early steps from drop-off up to the data capture of the questionnaires - steps which are now decentralized and less amenable to close control.

Under the pre-list automated VR delivery alternative, the starting point for the ACF would be the data capture of additions and modifications to the VR carried out by the CR at the time of drop-off. This activity would take place right on the heels of drop-off, to ready the ACF for the check-in of questionnaires as they are mailed back.

With an ACF, check-in of the questionnaires would be streamlined by the use of bar-coded labels. These labels would be generated from the AR for pre-listed dwellings, and affixed to census questionnaires at the time of drop-off. Extra labels could be pre-printed within each EA to handle additional dwellings.

The ACF would be updated to reflect the current status as individual census documents pass through the processing steps of check-in, edit, and follow-up (where necessary). To minimize costs of recording status changes, at each step in the process after documents are batched by outcome in readiness for the next step, they could be re-passed through the bar-code readers.

The ACF would then be used to generate, on a demand basis, up-to-date status reports on numbers of cases at various stages of processing. More advanced functions would be possible, such as building in a scheduler and using the ACF to generate workloads for follow-up activities. An early accounting of number of dwelling additions and questionnaire return rates by EA would permit trouble shooting of potential problem cases.

Given that there is currently good control of processing at the EA level following data capture, the ACF would probably be best restricted to the steps up to and including data capture. Captured records could be stockpiled, with completed EA's being sent through to later steps.

While the pre-list manual VR delivery option is not well suited to a full scale ACF, even with this delivery alternative and decentralized follow-up, there is still potential for savings in head-office check-in through use of bar coded labels as described previously. This would greatly reduce the current manual task of matching questionnaires against the VR to ensure all dwellings are properly accounted for.

4.4 Applications in Geography Processes

An address register would impact on ongoing geography maintenance activities, on pre-censal activities such as delineation of Census Tracts and Enumeration Areas, and on processing of census data to add geographic coding.

One of the major ongoing geography maintenance activities is the updating of the Area Master File (AMF). The AMF is a machine readable file which covers urban areas of Canada, and contains the geographic identification down to the block face level required to generate automated maps showing street networks and standard geography boundaries. Updating consists of capturing new and changed street patterns. Frequency of updating varies from city to city depending on factors such as availability of source data from municipalities and growth rates, with updating being most intense prior to the census to facilitate accurate delineation, and after the census so that the AMF reflects conditions at the time of the census for linkage to the census data base. A regularly updated AR (with annual or more frequent updates) could increase the cost efficiency of AMF updating efforts by signalling areas of change.

The delineation of EA's and CT's prior to the census would benefit from up-to-date counts, since number of dwellings is one of the principal criteria used in the delineation. Currently, for many areas nothing more recent than counts from the previous census are available. However for areas of known large growth field counts are obtained intercensally which are used both in updating the area sampling frame for household surveys and in preparation for delineation. A regularly updated AR providing up-to-date dwelling counts for all AR covered areas would eliminate the need for such special field checks whose coverage is only partial. The improved delineation that would result would greatly reduce the instance of EA splits occuring in the field at census time, which are both costly and disruptive.

Gains would also result in geocoding of census data, provided the AR is used to generate labels which are affixed to questionnaires at drop-off. Geocoding is currently a manual operation which links household numbers to their block face centroids (x and y co-ordinates), to permit retrieval of census data for non-standard geographical areas. The capture of a unique address key appearing on the label coincident with capture of the census questionnaire would link census data to the AR, where geocodes would already be present for the majority of addresses, greatly reducing the amount of clerical work required.

4.5 Source of Information on Structural Type of Dwelling

Advantages of municipal assessment records as a source for AR construction are that their coverage is good, and the files are updated on an annual basis. Additionally based on experiences in Ontario and British Columbia, it seems there is some potential for equating property or other codes on these files to information on dwelling structural type that could be used in evaluating the structural type information collected in the census.

Alternatively, under either pre-list delivery option for use of an AR, dwelling structural type where available on the AR could be pre-printed. In such cases, the CR's responsibility would be to verify this information and record any corrections on the VR. In cases where the information is not pre-printed, the CR as now would be responsible for determining and recording the structural type.

It is planned to evaluate the use of dwelling structural information during the field testing planned for the fall of 1987.

5. Directions of Future Work

Decisions on what uses if any to make of an AR in the 1991 Census will have to be made towards the end of 1987 or early in 1988, when development and implementation plans will be formalized.

In order that supportable recommendations can be made, the major thrust of the project in the coming months will be in the design, conduct and analysis of a field study to evaluate the cost and quality (coverage) implications of use of an AR in census collection as a pre-list to be updated by CR's at drop-off (delivery alternatives (2) and (3) from the previous section), versus use of the AR in a post drop-off coverage check (alternative 4). The test is planned for five cities representative of Canada's various regions, the 5 Regional Office cities, in September 1987, and will also include evaluation of dwelling structural information on an AR.

Other components of the work plan include:

(i) Systems prototype featuring integration with existing geography files

Plans are to develop a systems prototype on personal computers using a relational DBMS, that could later be portable to the mainframe. This system will ideally integrate the AR and other major geography files and systems, such as the Area Master File, and the Canada Conversion File, which links postal codes to standard geography.

(ii) Pilot register for pre-censal geography applications

One of the first production activities associated with the 1991 Census will be the districting of EA's and CT's, which will begin early in 1988. To test the benefits of an AR to pre-censal Geography operations, and other Geography operations such as AMF updating, it is proposed to construct a pilot register for a sample of areas within each of the five Regional Office cities, and to maintain the registers over the period leading up to the 1991 Census.

(iii) Proposal to test household survey use of an AR

This test would use the pilot AR constructed in (ii) above. For a collection of Labour Force Survey strata in each of the Regional Office cities, the LFS sample would be converted to an AR based design featuring separate strata and data collection methodologies for households with and without telephones on the AR. The study would also include investigation of approaches to dealing with undercoverage of the AR.

(iv) Non-Statistics Canada uses of an AR

During the current research and developmental stage, Statistics Canada can declare any pilot registers to be confidential, thereby exempting them from requests for release outside the agency. However, in the event of implementation of an AR, considerations relating to confidentiality would be different. The confidentiality of information pertaining to individual addresses would hinge on factors such as whether the information is already in the public domain, and the conditions under which the information is acquired by Statistics Canada. Clearly it would be in Statistics Canada's interest to set a price for any portion deemed to be non-confidential, since otherwise under freedom of information legislation, the agency would be obliged to supply it at nominal cost to any party requesting it.

Analysis is needed of the legal, public relations and political implications of different interpretations of what would or wouldn't be confidential. The benefits in terms of unduplication of effort that could result from Statistics Canada's sharing both information from an AR and costs of its creation and maintenance, would have to be weighted against the harm that might result from any public perception of invasion of privacy.

In addition to potential non-Statistics Canada uses of information on individual addresses, summary data from a register which by its nature would be non-confidential, may also be of potential benefit to others. Dwelling counts by postal code are an example of such summary data.

While the benefits of an AR to Statistics Canada programs should be of primary importance in decisions on implementation, consideration should also be given to the desirability of and demand for non-Statistics Canada uses.

ACKNOWLEDGEMENTS

A team of Statistics Canada specialists in geography, geocartographics, informatics, survey operations and methodology is involved in the AR research described in this paper. Team members include Y. Deguire, S. Earwaker, T. Hill, P. Lozano, R. Nadwodny, L. Nieman, J.P. Parker, P. Tallon, A. van Baaren, J. Yan and the authors.

Table 1 : Pilot Address Registers - Matching Results

	LFS Test		Census Test	
	Number	Percentage	Number	Percentage
Total Dwellings	3540		5942	
Matches				
Exact	3115	88.0	4460	75.1
Record Linkage	79	2.2	200	3.4
Total	3194	90.2	4660	78.4

Table 2: Coverage of LFS and Pilot Address Register - Ottawa Test

	Single Family		Multiple		Overall Coverage	
	No.	%	No.	%	No.	%
Total Dwellings	2228		1312		3540	
Total LFS Dwellings	2231	100.1	1424	108.5	3655	103.2
LFS Overcoverage	14	0.6	181	13.8	195	5.5
LFS Undercoverage	11	0.5	69	5.3	80	2.3
Total AR Dwellings	2269	101.8	1401	106.8	3670	103.7
AR Overcoverage	64	2.9	206	15.7	270	7.6
AR Undercoverage	23	1.0	117	8.9	140	4.0

Table 3: Coverage of 1986 Census and Pilot Address Register - Ottawa Test

	Single Family		Multiple		Overall Coverage	
	No.	%	No.	%	No.	%
Total Dwellings	1898		4044		5942	
Total Census Dwellings	1906	100.4	4048	100.1	5954	100.2
Census Overcoverage	43	2.3	178	4.4	229	3.9
Census Undercoverage	51	2.7	174	4.3	217	3.7
Total AR Dwellings	2092	110.2	4551	112.4	6643	111.8
AR Overcoverage	276	14.5	690	17.1	966	16.3
AR Undercoverage	82	4.3	183	4.5	265	4.5

Table 4 : Census and Address Register Overcoverage - Census Test

Type of Overcoverage	Census	AR
Too Many Dwellings associated with Street Name and Number	94	525
Incorrect EA/Invalid Address	72	258
Nonresidential	56	162
Demolished	7	12
Duplicate	0	9
Total	229	966

Table 5 : Pilot Address Registers - Coverage Improvement

	LFS Test		Census Test	
	Number	Percentage	Number	Percentage
Single Family	11	0.5	43	2.3
Multiple (Number of Dwellings)	20	1.5	87	2.2
Total	31	0.9	130	2.2

REFERENCES

Armstrong, J., Deguire, Y., Hill, T. and Lozano, P. (1987). Research into the parsing and standardization of free format addresses at Statistics Canada. Internal Report, Statistics Canada.

Booth, J.K. (1976). A summary report of all address register studies to date. Internal Report, E-414E, Statistics Canada.

Burgess, R. (1983). Reverse record check tabulations. Internal Report, Statistics Canada.

Copp, J. (1982). Postal check evaluation. Internal Report, Census and Household Survey Methods Division, Statistics Canada.

Cyr, A. (1984). Analysis of dwelling coverage based on re-interview reports. Internal Report, Statistics Canada.

Dibbs, R., Tallon, P., Proulx, P., and Yan, J. (1987). Implications of an address register on census processes. Internal Report, 1991 Census, Statistics Canada.

Fellegi, I.P. and Krotki, K.J. (1967). The testing programme for the 1971 census in Canada. Proceedings of the Social Statistics Section, American Statistical Association, 29-38.

Fellegi, I.P. and Sunter, A.B. (1969). A theory for record linkage. Journal of the American Statistical Association, 64, 1183-1210.

Gamache-O'Leary, V., Nieman, L., and Dibbs, R. (1987). Cost implications of mail-out of census questionnaires using an address register. Internal Report, 1991 Census, Statistics Canada.

Hill, T. and Pring-Mill, F. (1978). Generalized iterative record linkage system. Proceedings of the Workshop on Exact Matching Methodologies, Arlington, Virginia, 327-333.

Jaro, M. (1986). "Matching the 1985 test census of Tampa, Florida", Proceedings of the Survey Research Section of the American Statistical Association (forthcoming).

McCririck, D. (1986). Coverage check report. Internal Report, 1986 Census, Statistics Canada.

Royce, D. (1986). Address register research for the 1991 Census of Canada. Journal of Official Statistics, 2, 447-456.

Royce, D. and Pryor E.T. (1987). Canada's Census of Population: 1986 experience and 1991 plans. Internal Report, 1991 Census, Statistics Canada.