ADDRESS REGISTER RESEARCH AT STATISTICS CANADA

J.D. Drew, D.A. Royce and A. van Baaren

1. Introduction

The next Census of Population in Canada is due to be held in June 1991. As part of its overall programme of research and testing for this Census, Statistics Canada has been conducting research into methods for creating a database of residential addresses (an Address Register) for urban areas of Canada. Such a register has many potential uses in conducting a Census. For 1991, the emphasis is on using the Address Register for improving the coverage of dwelling units.

This paper describes a programme of Address Register research which has been conducted during the past three years. Section 2 provides some background, including a brief description of research that dates back to the 1960s and the reasons for renewed interest in the topic at Statistics Canada. Section 3 describes the methodology used for creating the Address Register and presents results on the quality of the register. Section 4 describes the methodology and presents the results of a major test of the Address Register to evaluate its potential for improving dwelling coverage in the Census. Sections 5 and 6 describe plans for future uses of the Address Register in both the Census and in other Statistics Canada programmes, including use as a frame for household surveys.

In keeping with the theme of the session, the paper emphasizes applications of the Address Register in data collection. It should be pointed out, however, that the Address Register has uses throughout the survey-taking process, including data processing, data quality evaluation and even as a substitute for direct data collection. A description of these other applications can be found in other papers by the present authors listed in the References.

2. Background to the Project

Address Register research is not a new topic at Statistics Canada. Research was carried out as part of the testing programme for the 1971 Census (Fellegi and Krotki, 1967) and continued well into the mid-1970s (Booth, 1976). These early research efforts were primarily aimed at using the Address Register as a frame for a mail-out Census, with cost savings being the major benefit.

Experimental ARs were created and tested for several cities, using a variety of sources of addresses. Among these sources were the records from the previous Census, electricity billing lists, municipal tax assessment rolls, records from the Post Office and field listing by Labour Force Survey interviewers. In some cases the initial ARs were updated with a postal check done by the local letter carriers.

Field tests of these ARs demonstrated that the quality was generally comparable to and in some cases better than that obtained by field listing, with the possible exception of sub-addresses within the larger urban centres. Despite the promising quality, however, the research did not continue. The creation and maintenance of these Address Registers was very labour-intensive, with a high capital cost at the front end. Instead of mail-out, the method adopted (and still used) for the Census is for an enumerator to create the address list manually (the so-called Visitation Record) while dropping off the Census questionnaire just prior to Census Day.

The current research project was initiated at the International 1991 Census Planning Conference, held in Ottawa in October 1985 (Royce, 1986, 1987). It was proposed that a number of developments since the mid-1970s had made it possible to consider creating and maintaining Address Registers by automated, and hopefully much less expensive, methods:

(a) the availability or potential availability of an increasing number of up-to-date administrative files, such as records from telephone and utility companies, tax files and social security files, with addresses already in machine-readable form;

(b) the universal use of the postal code on these files, which, combined with recently-developed postal code - census geography conversion files, facilitated the automated geo-coding of these addresses;

(c) improved address standardization and record linkage methodologies and systems which could be used to unduplicate multiple files; and

(d) the declining costs of storing and working with large databases.

Although the major benefit was again seen as cost reduction through conversion to a mail-out Census, a number of other potential applications were identified. A five-year research plan was developed in early 1986 (Royce, 1986) and a team was formed to begin work.

One of the first activities of the project was to compare the costs of an Address Register/Mail-out Census to the traditional Visitation Record/Drop-off Methodology (Gamache-O'Leary et al, 1987). The conclusion was that the new approach would be less expensive only if the initial Address Register were of sufficient quality to require little or no updating in the field prior to being used. In early 1987, two experimental ARs for Ottawa and Vancouver were created using the new methodology and were tested by comparing them to address lists created by Census enumerators or Labour Force Survey interviewers. The tests put the coverage of the Address Register at approximately 90-95%, not good enough to be used without field updates (Drew et al, 1987a), and as a result it was decided not to pursue the mail-out option any further for 1991.

At about the same time, however, information about the level of undercoverage in the 1986 Census was coming to light. The Reverse Record Check (Statistics Canada, 1988) estimated that the rate of population undercoverage had increased to 3.2%, compared to a level of about 2% in the previous three censuses. Methods for reducing undercoverage in the Census suddenly became of high priority.

Although the Ottawa and Vancouver pretests had shown that the Address Register itself had relatively low overall coverage when compared to field listing by enumerators, the tests had also noted that the Address Registers had identified a surprisingly high number of dwellings that enumerators had missed. Under the circumstances, it was decided that the research project should concentrate on the development of the Address Register as a coverage improvement method for the 1991 Census. To explore this possibility more thoroughly, it was decided to mount a major test in November 1987 that would test two different coverage

improvement methods. Before describing the test and its results, however, we first give a description of the methodology for creating the Address Register.

3. Methodology for Creation of the Address Register

The basic source of addresses for the Address Register is a set of administrative files. Table 3.1 presents the administrative files used as sources in constructing Address Registers for the November 1987 Test. It is anticipated that for the 1991 Census, appropriate subsets of the same administrative files will be used for AR creation. Three national files already available to Statistics Canada were used for all cities: the Revenue Canada personal taxation file (TAX), and Health and Welfare Canada files of Family Allowance (FAM) and Old Age Security (OAS) recipients. In addition, for each site, two lists were purchased from among municipal assessment rolls (MUN), telephone billing lists (TEL) and electricity billing lists (ELE). Edmonton was an exception in that due to a delay in obtaining one of the extra files, the Address Register was constructed using only four files.

Table 3.1: November 1987 Test: Source Files by Test Site

Test Site	Source File					
	TAX	FAM	OAS	MUN	TEL	ELE
Vancouver	x	x	x	x		x
Edmonton	x	x	x	x		
Toronto	x	x	x	x	x	
Montreal	x	x	x		x	x
Halifax	x	x	x	x	x	

The four principal steps involved in AR creation are discussed below.

Address Standardization

Address information on administrative files is typically in free format, by which we mean there is no fixed position or even order of appearance for the components of the address such as street name, street number, apartment number, and so forth. It is necessary to analyse the address information to identify the components, in order that the address can be re-written in a standard form to facilitate matching.

An automated system has been specially developed for this purpose (DeGuire, 1988). The system breaks the free format address into tokens, which are strings of consecutive letters or numbers, separated by blanks or delimiters such as commas. Some tokens are recognized by the system as keywords. Examples of keywords include "Street", "Rue", "Apt", "App" and so forth. Based on the pattern of numeric and alphabetic tokens, and known keywords, we have found that it is possible to uniquely decode over 95% of addresses unambiguously into components. While as few as eight patterns account for 52% of addresses, the number of variations is large and over 1600 patterns are needed to handle 95% of the addresses. The remaining 5% of addresses for the test were reviewed and, where possible, deciphered manually. We have found, however, that manual address standardization is a time consuming and error prone operation; it is therefore very likely to be cost effective to spend more resources on automation in an effort to reduce the amount of manual intervention.

Merging and Unduplication

After merging the standardized addresses from all the source files, the next step is to eliminate duplicates -that is, addresses referring to the same address. This is broken into two parts - exact matching to get rid of exact duplicates, and record linkage to identify duplicates where there is disagreement or only partial agreement on one or more of the standardized components. Such discrepancies occur for numerous reasons, such as variations in spelling, use of non-standard abbreviations, and so forth. The record linkage is carried out using Statistics Canada's record linkage software GIRLS (Hill and Pring-Mill, 1978) which is based on the Fellegi and Sunter (1969) methodology.

Geographic Coding

Since we want to ultimately produce lists of addresses by Census Enumeration Area from the Address Register, the linkage of the Address Register to standard Census geographic coding to the level of the block or block face is crucial. This linkage bears directly on the coverage of the Address Register at the Enumeration Area level, as well as on the cost of enumeration.

A number of possibilities exist for establishing this link. To date, no single method has shown to be the best. The November 1987 Test used only the Postal Code portion of the address to match to a Postal Code - EA conversion file that had been created by matching the Postal Code Master File (from Canada Post) to Statistics Canada's automated street network file (the Area Master File). Another possibility is to use the other components of the address to match directly to the Area Master File. A third option is to use a linkage that was established by the capture of Postal Codes for a one-fifth sample of dwellings in the 1986 Census. Plans exist to update and maintain this link, as well as to evaluate its accuracy. However we are mindful of the fact that to use it in linking an Address Register to Census Enumeration Area would impose requirements for accuracy and updatedness well beyond what has been needed to support current uses.

Edit and Imputation

The final step in address register construction consists of fine tuning. For instance, logical gaps in apartment numbers can be imputed. Some clearly erroneous addresses which escaped detection at earlier steps in address register construction may be spotted clerically and deleted.

Further details on Address Register construction are contained in Drew et al (1987b).

How good is the Address Register created with this method? Table 3.2 gives some results from the November 1987 test which reflect on the quality of the AR creation process, especially the steps of unduplication and imputation. All entries in columns two through eight are expressed as percentages of the figures in column one, which represents the true number of valid addresses as established at the end of the test (see Section 4 for details of the test methodology).

Table 3.2: Address Register Effectiveness by Test Site (%)

	Tot. valid addr.		Total addr. AR		AR-undercoverage		AR-overcoverage	
	Test (1)	1986 Census (2)	before imp. (3)	after imp. (4)	before imp. (5)	after imp. (6)	Total (7)	dup. addr. (8)
Vancouver	8824	97.1	96.1	97.6	9.9	9.8	7.3	3.0
Edmonton	7517	89.3	90.3	111.3	26.9	23.7	35.0	23.8
Toronto	7459	90.1	100.0	108.2	14.6	14.4	22.6	8.3
Montreal	9556	91.1	100.1	103.1	18.6	17.5	20.7	3.4
Halifax	9201	90.4	102.0	104.1	11.9	11.5	15.7	3.7
Total	42557	91.7	97.9	104.5	16.1	15.2	19.7	7.8

From the last column of the table it may be seen that the unduplication module (and with it the address standardization module) still leaves considerable room for improvement. This was particularly true in

Edmonton, where lack of unduplication contributes 23.8% to the total 35.0% overcoverage of the Address Register. We suspect that much of this may be caused by the common practice of using numbers as street names in Edmonton. The high undercoverage of the AR in Edmonton is felt to be due to the fact that neither the telephone nor the electricity billing lists were available.

The table also shows the relatively limited usefulness of imputation of addresses. While imputation increased the total number of addresses from 97.9 percent of the true total to 104.5 percent of the true total (bottom row, columns three and four), it only reduced the undercoverage rate of the Address Register by 0.9% (bottom row, columns five and six).

The fact that both undercoverage and overcoverage of the AR are relatively high (generally 10-20%) also points to the fact that further improvements to the geographic coding are still needed. Much of the undercoverage and overcoverage simply reflects the assignment of the address to the wrong geographic area.

What files were the best? Table 3.3 shows the contribution to net coverage by source file and test site, with figures in parentheses giving the contribution which is unique to that source file.

Table 3.3: Net Coverage (and Unique Contribution) of Source Files (% of valid addresses)

Site	TAX	FAM	OAS	MUN	TEL	ELE
Vancouver	70 (2)	29 (†)	23 (†)	51 (0)	-	84 (9)
Edmonton	62 (23)	23 (4)	14 (3)	33 (2)	-	-
Toronto	52 (1)	19 (0)	17 (0)	73 (5)	67 (8)	-
Montreal	42 (1)	19 (0)	12 (1)	-	61 (7)	67 (8)
Halifax	73 (5)	29 (†)	19 (†)	54 (1)	69 (7)	-

The table confirms the low contributions of the Family Allowance and Old Age Security files (as might be expected from their nature) and the relatively high contributions of the telephone and electricity billing lists. The high contribution of the Municipal Assessment file for Toronto is surprising since this file only lists homeowners.

4. The November 1987 Address Register Test

The objective of the November 1987 Test was to evaluate the potential improvement in dwelling coverage that could be achieved through use of an Address Register. The test was carried out in the five cities in which Regional Offices of Statistics Canada are located, specifically Halifax, Montreal, Toronto, Edmonton and Vancouver. The test was designed so that rates of both undercoverage and overcoverage of dwellings could be estimated for the standard Census method and two experimental methods of using the Address Register. A flowchart illustrating the test methodology is shown as Figure 4.1.

In the first experimental method, known as the Pre-list method, the enumerator was given the Address Register for the Enumeration Area (EA) and was asked to update it (adding and deleting dwellings) during a canvass of the EA. In the other method, known as the Post-list, a different enumerator first canvassed the EA and compiled a list of dwellings. The list at this point thus gave the standard Census method. The enumerator was then given the Address Register for the EA, identified dwellings on the Address Register but not his/her list, and followed up on these addresses in the field to determine whether or not they should have been listed.

Although each EA in the sample was enumerated twice, it was necessary to avoid contacting respondents

NOVEMBER 1987 ADDRESS REGISTER TEST METHODOLOGY

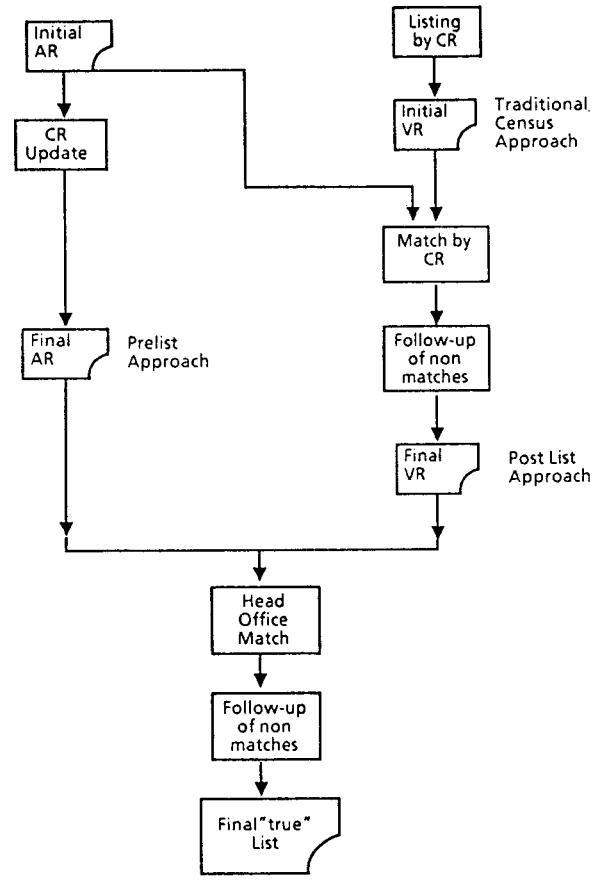

Figure 4.1: Flowchart.

more than once. Accordingly, half of the EAs were randomly designated as "contact" for the Pre-list but "no contact" for the Post-list, while the other EAs were designated "no contact" for the Pre-list and "contact" for the Post-list method.

Following the completion of this stage of fieldwork, the two address lists compiled under the Pre-list and Post-list methods were matched at Head Office by computer, with manual resolution where necessary. All addresses which matched were assumed to represent valid addresses. All non-matches, i.e., addresses that appeared on one of the lists but not the other, were further checked in the field to determine whether or not the address was valid. This match-and-reconcile operation resulted in a final list of valid addresses to which the other three lists (standard Census method, Pre-list method and Post-list method) could be compared. Comparisons of the original Address Register to the final list of valid dwellings have already been shown in Section 3.

The basic results, in terms of rates of dwelling undercoverage and overcoverage, are shown in Tables 4.1 and 4.2. Table 4.1 shows the rates of gross overcoverage, gross undercoverage and net undercoverage (or in a few cases overcoverage, indicated by an asterisk) for the three methods. Table 4.2 presents the same information, re-arranged to show the percent reduction (or increase) in coverage error of

the two experimental methods relative to the Census method. For example, from Table 4.1 we see that for Halifax the Post-list method had an overcoverage rate of 2.6%, compared to 2.1% for the Census method. From Table 4.2 we see that this represents a 21.3% increase in the overcoverage rate. Similarly, the undercoverage rates (Table 4.1) for Halifax were 4.0% and 2.0% for the Census and Post-list methods respectively, which from Table 4.2 is seen to represent a 49.0% decrease in the undercoverage rate.

Because the procedure used in the 1991 Census would involve contact, the rates shown are based on the portion of the sample where the method involved was designated as "contact". As an important side finding, it was confirmed that attempting to make contact was an important factor in achieving good dwelling coverage.

Table 4.1: Rates of Coverage Error for November 1987 Address Register Test ("contact" cases)

Test Site	Percent Overcoverage			Percent Undercoverage			Percent Net Undercoverage		
	Census	Post-list	Pre-list	Census	Post-list	Pre-list	Census	Post-list	Pre-list
Halifax	2.1	2.6	1.4	4.0	2.0	2.4	1.8	0.6*	0.9
Montreal	1.9	2.1	1.7	4.3	2.7	2.5	2.5	0.6	0.8
Toronto	1.8	2.8	0.8	7.2	2.9	4.6	5.3	0.2	3.8
Edmonton	1.0	1.1	12.3	2.8	2.3	6.3	1.7	1.2	6.0*
Vancouver	1.9	2.4	1.2	3.6	1.6	4.5	1.7	0.8*	3.4

Note: * denotes net overcoverage

Table 4.2: Percent Reduction (Increase) in Coverage Error Relative to Census Method

Test Site	Overcoverage		Undercoverage		Net Coverage Error	
	Post-list	Pre-list	Post-list	Pre-list	Post-list	Pre-list
Halifax	21.3*	31.7	49.0	40.5	69.6	50.8
Montreal	10.8*	10.1	37.8	42.2	74.2	66.3
Toronto	51.5*	56.0	59.2	35.8	97.0	28.9
Edmonton	4.4*	1079.0*	18.3	127.2*	32.0	243.9*
Vancouver	22.6*	38.4	56.0	24.6*	56.0	95.2*

Note: * denotes increase in coverage error

In terms of undercoverage, the Post-list method was clearly the best in all sites except Montreal, where the Pre-list method came out slightly better. The Post-list method achieved substantial reductions in undercoverage compared to the Census method, although at the expense of some additional overcoverage. The Pre-list method was better than Census in three of the five sites, but worse in Edmonton and Vancouver.

For overcoverage, the Pre-list method was the best in all sites except Edmonton, where it performed poorly (overcoverage rate of 12.3%). As mentioned earlier, this problem is probably because the frequent use of numbers for street names resulted in the Address Register having a high rate of duplicates (Table 3.2). Although enumerators reduced the rate considerably (from 35.0% to 12.3%), the rate is still unacceptably high.

Overall, the Post-list method would appear to be preferable, since it achieves a greater reduction in undercoverage while adding only a small amount of overcoverage, as reflected in the percent net coverage error in Table 4.1. Because it is also less risky than the Pre-list option, the Post-list method has been chosen for the 1991 Census.

It should be noted that the levels of dwelling undercoverage observed in the test are higher than those usually achieved in the Census itself. Part of this is undoubtedly due to the fact that an actual Census will pick up dwellings through other means which were not part of the test, for example advertising or enumerators finding dwellings during follow-up. However part of the explanation for the higher rates is the test methodology itself. Because of imperfections in the Head Office match operation, there will be some duplicate addresses remaining in the final list of valid dwellings. These duplicates will inflate the number of "valid dwellings" and therefore result in higher apparent undercoverage rates. All efforts possible were taken to identify and eliminate duplicates, but some undoubtedly remain. While they will inflate the estimates of undercoverage, however, they will have relatively little effect on the comparisons between methods. This problem would not, of course, occur in the Census itself.

Although of secondary importance, the field costs associated with the various methods were also of interest. Table 4.3 shows the speed of enumeration, in terms of person days per 100 addresses, by enumeration method, contact/no contact and stratum. The strata were used in selecting the test sample and refer to the majority type of dwelling in an Enumeration Area. The numbers in parentheses denote the actual rate of contact made by enumerators, for those cases where contact was attempted. The figures for the Post-list method reflect the combined effort of the Census method and the additions obtained by doing the post-list check, but excluding the time for the enumerator to match the two lists. As a result the figures for the Post-list method depend on the coverage rates of both the initial Census method and the original Address Register.

Table 4.3: Enumeration Speed in No. of Person days/100 Addresses

Method / Contact / Stratum	Pre-list			Census			Post-list		
	No	Yes	(Rate)	No	Yes	(Rate)	No	Yes	(Rate)
Single detached	.43	.75	(.39)	.52	.69	(.45)	.73	.89	(.42)
Single attached	.40	.55	(.26)	.49	.61	(.30)	.70	.80	(.28)
Apartment	.42	.53	(.29)	.41	.69	(.34)	.62	.90	(.31)
Duplex	.37	.78	(.31)	.50	.39	(.36)	.91	.74	(.36)
Total	.41	.60	(.29)	.49	.64	(.34)	.70	.84	(.32)

Not surprisingly, the Pre-list method is the fastest, followed by the Census method and the Post-list method. The question of interest is whether the increase in coverage is worth the extra time required, however, and for the 1991 Census the answer is "Yes". Table 4.3 also suggests, however, that if the coverage of the original Address Register could be improved to the point where the coverage of the Pre-list method came close to that of the Post-list method, then some cost savings could be achieved while at the same time improving coverage compared to our current Census method.

Also from Table 4.3, enumeration is generally faster when contact is not attempted, except for the Census and Post-list methods in the "Duplex" stratum. The data in these areas are based on a small number of observations, however, and the differences between contact and no contact are not statistically significant.

5. Future Census Applications

For the 1991 Census, the current plan (subject to funding approval for the 1991 Census) is to create and use the Address Register for coverage improvement using the Post-list method. The Post-list method was chosen both because of its better performance in the November 1987 Test and its low level of risk. If, for example, the Address Register cannot be created on time for some area, then the procedure would simply revert to our current Census method. With the Pre-list method, lack of availability of the AR at the last moment would be a major problem.

Use of the AR will be limited to those urban areas of the country covered by our Area Master File. It is felt that the geographic coding of addresses is not of sufficiently high quality outside of these areas. The Area Master File generally cover urban centres of 50,000 population or more, accounting for approximately 55 percent of the Canadian population. In future years, with extension of the Area Master File to smaller centres, the coverage of the Address Register could be extended as well.

In 1991, it is estimated that use of the AR will add about 68,000 persons or more to the Census count. This represents a reduction of a quarter percentage point in the undercoverage rate. The estimated cost is $3.9 million. While expensive, the method is relatively cost effective compared to other coverage improvement methods. For example, a post-drop-off Postal Check is estimated to cost $3.1 million, but would add less than half as many persons. The Address Register also tends, because of the nature of the files used to create it, to improve coverage more in certain types of dwellings, such as duplexes, basement apartments, etc. Thus differential undercoverage (at least by type of dwelling) should be reduced.

The current plans also call for using the 1991 Census to update the Address Register itself. As part of the operation of matching the Address Register to the Visitation Register, the Census identifier from the Visitation Record will be coded onto the Address Register. Addresses on the AR but not the VR will be listed on a separate form for follow-up to verify if they represent valid addresses; if so they will be added to the VR and enumerated. In addition, addresses on the VR but not the AR will be listed on another form, for eventual addition to the AR. The relevant information (Census identifiers for addresses that match, Census identifiers and address information for addresses that do not match) will be key entered and used to update the Address Register. The updated Address Register will then continue to be maintained using administrative files. The updating costs are estimated to be an additional $1.6 million through to March 1994.

A second test is now being planned for September 1989. The test is intended as a dress rehearsal of the procedures to be used in 1991, and so only the Post-list method is being used. As well as testing the procedures to ensure they will function as intended, the test will provide better information on costs which will be used to refine the cost estimates for the 1991 Census. Also of interest is an assessment of several improvements that have been made to the methods and software used for creating the Address Register, in particular the geocoding of addresses.

If successful in 1991, the Address Register may eventually find a number of other uses in conducting the Census. These are described more fully in the paper by Royce (1986), and we list only a few of the collection-oriented ones here:

(i) Inclusion of telephone numbers on the AR could allow enumerators to conduct follow-up by telephone in cases where the respondent did not mail back the questionnaire, reducing the cost of follow-up.

(ii) Availability of the list of addresses in machine readable form could facilitate automation of much of the collection operation, for example automation of questionnaire check-in, generation of assignments for follow-up, automated generation of status reports, and control of subsequent processing steps.

(iii) More up-to-date dwelling counts derived from the Address Register could provide better information for planning the logistics of Census collection, for example delineating geographic areas into assignments, identifying the number of staff needed, etc.

(iv) A high quality Address Register could eventually be used for a mail-out or a Pre-list Census, resulting in cost savings to the Census.

Finally, even if none of the above applications are developed, maintaining the Address Register up to date would greatly reduce the cost of using it again in the next Census, scheduled for 1996.

6. Applications to Household Surveys and Other Statistics Canada Programs

Although the Address Register research is being sponsored by the Census, there are a number of applications to other Statistics Canada programmes. Among potential uses, for example, are:

(i) New businesses could be identified from some of the administrative files for potential addition to Statistics Canada's Business Register.

(ii) It could become the foundation for a national database on housing. The municipal assessment rolls, one of the source files of the AR, often contain data related to the dwelling. For example, the Ontario file contains type of structure, tenure, assessed value, date of construction, date of renovation, number of bedrooms, number of bathrooms, square footage, presence or absence of finished basement, and central air conditioning. Much of this information is now collected through the Census or in household surveys.

(iii) Dwelling counts at small geographic areas could be a useful piece of information for improving and expanding the demographic programme of post-censal population estimates for sub-provincial areas.

(iv) The Address Register could serve as a file for automatically coding address information such as collected in questions on previous place of residence and place of work.

The major non-Census application, however, lies in using the Address Register as a frame for conducting household surveys, in particular the monthly Labour Force Survey used for the production of labour force data and for numerous other household surveys on specialized topics. On balance, it would appear to be preferable to both the current area sampling frame and the alternatives of either a telephone frame or a dual frame combining area and telephone frames.

A strength of the current area frame methodology is its conceptual completeness, since all land area receive a probability of selection. However it has a major drawback - the need for clustering of the sample. The need for clustering stems from both the high cost of creating and maintaining, through fieldwork, lists of dwellings in selected land areas and from the need to conduct the first month interview in person. Clustering the sample is the only effective method for reducing these costs to a reasonable level. While new methods, involving matching of LFS dwelling lists to telephone company lists to obtain telephone numbers, will permit up to 50% of initial interviews in urban areas to be done

by telephone, clustering of the sample is still necessary for the remaining 50%.

The Address Register as a frame would have several advantages. Dwellings on the Address Register could be stratified into those for which telephone numbers are available, and those without telephone numbers. The stratum with telephone numbers would include those dwellings with published residential numbers, and for those dwellings an efficient, unclustered sample could be selected. Data collection could be by telephone, with the flexibility to globally or locally revert to face-to-face follow-up of non-response cases in order to maintain acceptable response rates.

The non-telephone stratum would contain those dwellings without telephones and those with unpublished numbers. For this stratum, the sample would be clustered. For example, "clumps" of 4-5 consecutive dwellings on the list would be selected. Clustering is called for in such cases since initial interviews with these dwellings would be face-to-face. This methodology would also respect the privacy concerns of those with unpublished numbers, unlike such telephone frame techniques as Random Digit Dialing.

One unresolved question is whether there would be a need for a coverage improvement sample, i.e., a sample selected independently of the Address Register to represent dwellings missed by the Address Register. Such a sample would be required if updating based on the 1991 Census and post-censal updates from administrative files still resulted in dwelling coverage below an acceptable threshold relative to Census dwelling coverage.

Determination of an acceptable threshold would depend on several factors, such as characteristics of the population in missed dwellings and the costs of a coverage improvement sample. Currently a 4% population undercoverage relative to the Census is the norm for household surveys, and it has been estimated that about half of this stems from undercoverage of dwellings. Hence, if coverage comparable to current levels were to be maintained, the threshold would be about 2%.

The design for a coverage improvement sample would follow the approach that, for a suitably determined areal sampling unit, the coverage of the Address Register would be rigorously checked in the field, and any additional dwellings found would constitute a frame of "dwellings missed by the Address Register". This frame would be sampled from, and replenished over time as required.

A cost-benefit analysis of replacing the current area frame and sample design with a new sample design based on the Address Register has been conducted, and concluded that the savings from efficiencies in the new design would be enough to offset the costs of maintaining the Address Register, with the possibility of some net reduction in costs. While the overall savings would not be large, the fact that the Address Register would be maintained inter-censally would lead to major cost savings for the next Census, scheduled for 1996.

7. Summary and Conclusions

The research described in this paper has demonstrated the technical feasibility of creating Address Registers from multiple administrative files. The quality of Address Registers created with this methodology is not yet comparable to that obtained by field listing of addresses, with rates of undercoverage and overcoverage typically in the 10-20% range. Considerable additional research is needed into several aspects of the methodology for creating Address Registers, in particular methods for standardizing addresses, the effective use of edit and imputation of addresses, and the geographic coding of addresses.

Despite these weaknesses, however, field tests have shown that the Address Register has considerable potential as a method of improving dwelling unit coverage over field listing alone. The current plan is to use the Address Register for this purpose in urban areas of Canada in the 1991 Census. How successful the Address Register is in 1991 will determine to a very large extent the long-term future of the research. If successful, the existence of a continuing, up-to-date database of residential addresses will have a profound effect on much of the methodology of the data collection carried out by Statistics Canada.

8. References

Booth, J.K. (1976) "A Summary Report of All Address Register Studies Completed to Date". Statistics Canada report E-414E.

DeGuire, Y. (1988) "Postal Address Analysis". Survey Methodology, Vol. 14, No. 2, pp. 317-325.

Drew, J.D., Armstrong, J.B. and Dibbs, R. (1987a) "Research into a Register of Residential Addresses for Urban Areas of Canada". American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 300-305.

Drew, J.D., Armstrong, J.B., van Baaren, A. and DeGuire, Y. (1987b) "Methodology for Construction of Address Registers Using Several Administrative Sources". International Symposium on Statistical Uses of Administrative Data, Ottawa.

Fellegi, I.P. and Krotki, K.J. (1967) "The Testing Programme for the 1971 Census in Canada". American Statistical Association, Proceedings of the Social Statistics Section, pp. 29-38.

Fellegi, I.P. and Sunter, A.B. (1969) "A Theory for Record Linkage". Journal of the American Statistical Association, 64, pp. 1183-1210.

Gamache-O'Leary, V., Nieman, L., and Dibbs, R. (1987) "Cost Implications of Mail-out of Census Questionnaires Using an Address Register". Statistics Canada internal report.

Hill, T. and Pring-Mill, F. (1978) "Generalized Iterative Record Linkage System". Proceedings of the Workshop on Exact Matching Methodologies, Arlington, Virginia, pp. 327-333.

Royce, D.A. (1986) "Address Register Research for the 1991 Census of Canada". Journal of Official Statistics, Vol.2, No. 4, pp. 447-455.

Royce, D.A. (1987) "Applications of an Address Register in the Canadian Census". Proceedings of the International 1991 Census Planning Conference, Statistics Canada, pp. 207-215.

Statistics Canada (1988) "Undercoverage Rates From The 1986 Reverse Record Check". Census User Information Bulletin Number 2.