DWELLING COVERAGE IN THE CANADIAN CENSUS - THE 1996 DWELLING COVERAGE STUDY

Craig Brown, Peter Dick, Statistics Canada
Craig Brown, Statistics Canada, 15-RH Coats Building, Tunney's Pasture, Ottawa, Ontario, K1A 0T6

Key Words: Address List, Coverage, Census

1. Introduction

The purpose of the 1996 Dwelling Coverage Study (DCS) is to estimate dwelling coverage in the Census, specifically in a Centralized Edit (CE) environment. A large scale test of the CE methodology is being conducted in the Ottawa area for the 1996 Census of Canada. Centralized Edit collection methodology relies extensively on the use of an Address Register (AR) to conduct a mailout/mailback Census as opposed to the traditional list-leave/mailback methodology. If the CE test is successful, it is anticipated that the CE methodology will be extended to all urban centres across Canada for the 2001 Census. Before this methodology is implemented across Canada, it would be useful to understand the nature of dwelling undercoverage in this environment, as it is expected that this change in collection methodology will result in a change in the undercoverage patterns of dwellings. As well, the Visitation Record used in the traditional list-leave/mailback Census will not be created in a CE environment and will therefore be unavailable for the AR to use as an update source. Thus, it is important to have an independent mechanism to evaluate the coverage of the AR.

The DCS will estimate dwelling coverage and obtain information about the dwellings missed (ie. type of dwelling, number of persons in the dwelling, etc.). This will be done by producing an independent list of dwellings in selected areas and matching this list to the Census list of dwellings for these same areas. Discrepancies between the two lists will be reconciled and estimates of dwelling undercoverage will be produced. Investigations will also be conducted to determine why these discrepancies occurred. As well, this study will provide the basic design parameters for a national DCS in 2001 if it is decided to extend CE to all urban areas in Canada.

This paper outlines the methodology for the 1996 DCS. Section 2 contains background information on CE methodology. Section 3 lists the objectives of the study. The sample design is explained in Section 4. Section 5 describes the data collection procedures. Section 6 outlines the post-processing activities as well as the weighting and estimation to be conducted. Finally, Section 7 outlines the evaluation and final report to be prepared at the conclusion of this study.

2. Background

Centralized Edit collection methodology is being tested in Eastern Ontario during the 1996 Census of Canada. With this new methodology, Census questionnaires are mailed out and then returned directly to a District Office. Collection operations, such as clerical edits and telephone follow-up, are centralized at this District Office (Choudhry, 1994). The test area covers 10 Federal Electoral Districts and includes more than 400,000 dwellings. Three distinct collection areas combine to make up the CE test area: The Precanvass area, the Prelist area and the List/Leave area.

The Precanvass area consists of the urban component of the Census Metropolitan Area of Ottawa-Carleton and accounts for 65% of all dwellings in the CE test area. An Address Register (AR) was constructed from administrative records in the Precanvass area. Enumerators precanvassed all areas to update this list. Prior to mailout, this address list was updated further using an address file from Canada Post.

The Prelist area is composed of seven smaller cities in Eastern Ontario and accounts for 10% of all dwellings in the CE test area. Enumerators canvassed these areas and created an address list of all dwellings in the Prelist area. Prior to mailout, this list was updated using an address file from Canada Post.

The third area is the List/Leave area. This area covered all rural areas and accounts for 25% of all dwellings in the CE test area. Census questionnaires were not mailed out in these areas. The traditional list-leave methodology was used in these areas. Enumerators canvassed their area, listed each dwelling and dropped off the Census questionnaire.

3. Objectives and Definitions

If it is recommended to extend Centralized Edit collection methodology to include all of Canada during the 2001 Census, it would be useful to understand the nature of dwelling coverage in this environment. The objectives of the DCS are to:

1) Estimate the number of private dwellings missed by the Census in the CE test area. Dual System Estimation will be used to estimate the true dwelling population in the CE test area. Patterns of undercoverage will be identified.

2) Evaluate the components used to compile the Census lists of dwellings. Since an accurate and timely address list is essential to ensure that coverage is as complete as

possible, the DCS will identify any components used to create the address file that resulted in coverage error.

3) Evaluate the geographical assignment of addresses. This will identify areas of weakness with the geocoding system. In particular, the occurrence of addresses getting assigned to an incorrect Census block or Enumeration Area will be examined.

4) Establish the design parameters for a national DCS if it is decided to extend CE to the national level in 2001.

Central to the DCS is the concept of a dwelling. The DCS defines a dwelling as a set of living quarters in which a person resides or could potentially reside. The dwelling must have its own entrance either from outside or from a common hall and must not pass through the living quarters of someone else. This is the same definition used by the Census. The DCS excludes unoccupied dwellings that are either marginal or under construction/renovation as well as collective dwellings and Indian Reserves.

4. Sample Design

A total of 3,500 dwellings were sampled in the CE area: 2,000 in the Precanvass area, 1,000 in the Prelist area and 500 in the List/Leave area. The following sections outline the sample design in each of these three areas.

4.1 Precanvass Area Sample Design

The sampling unit used in the Precanvass area was a Census block. A Census block is defined as a geographic area bounded on all sides by streets, roads, rivers or enumeration area limits. An average block in the Precanvass area (excluding blocks containing large apartment buildings) contained forty dwellings. In total, there were over 4,800 Census blocks covering 275,000 dwellings in the entire Precanvass area. These blocks were stratified into six strata.

The first stratum consisted of all blocks that contained an apartment building with thirty or more units. It is anticipated that dwellings will be missed less frequently in large apartment buildings. The second stratum was formed by grouping all blocks that were split in the field during CE enumerator canvassing. Since a block gets split when a new boundary, such as a new street, breaks the existing block into two or more blocks, split blocks are considered a greater risk for dwelling undercoverage. For example, the presence of a new street could indicate that the area is undergoing expansion; that, in turn, means an up-to-date address list may be difficult to maintain. Thus dwelling undercoverage may be more prevalent in these blocks.

The other four strata were formed by stratifying all remaining Census blocks into either a Highrisk stratum or a Lowrisk stratum and then stratifying further by size of block. Using 1991 Census data, several summary statistics were calculated for each block:

- the percentage of dwellings in the block that were small apartments (ie. dwellings in lowrise apartment buildings or multi-unit buildings);
- the percentage of dwellings in the block that were new dwellings;
- the percentage of dwellings in the block that were rented in 1991; and
- the percentage of dwellings in the block that were vacant in 1991.

These four variables, particularly small apartments and new dwellings, are generally associated with dwelling undercoverage. For example, basement apartments are difficult to locate, while a new dwelling may not have been constructed at the time the address list was compiled. The levels of these variables within a Census block were then used to determine the degree of coverage risk expected in the block.

The algorithm eventually used for stratification was modelled after the 'targeting database' approach used in the 1995 U.S. Census Test for identifying hard-to-count areas (Robinson and Kobilarcik, 1995). With this approach, geographic areas were profiled and assigned a 'hard-to-count' composite score. Areas that exceeded a predetermined cutoff were then targeted as the 'hardest to count'.

First, all Census blocks were ordered in descending order of percentage small apartments. Blocks in the 96th to 100th percentile received a small apartment score of 10, those in the 91st to 95th percentile received a small apartment score of 9 and so on. All blocks below the 50th percentile received a small apartment score of zero. This process was then repeated for the remaining three variables. The individual scores were then summed to form a Coverage Risk Factor (CRF) for each block. All blocks were then sorted in descending order of CRF and any block surpassing the 75th percentile was assigned to the Highrisk stratum (ie. blocks identified as posing the greatest coverage risk).

Due to unique circumstances within some blocks, the algorithm needed to be modified. For example, blocks composed primarily of new dwellings were under-represented in the Highrisk stratum. This was expected due to the dwelling composition of these blocks. Since newly constructed dwellings did not exist during the 1991 Census, there was no 1991 Census information available on the occupancy status or tenure of these dwellings. Although these blocks scored high on the 'new dwelling' variable, they received scores of zero for the 'unoccupied dwelling' variable and the 'rented' variable. This resulted in a CRF that was too

low to be categorized into the Highrisk stratum. It was therefore decided to increase the weight given to the new dwelling score by a factor of two, creating a revised Coverage Risk Factor (CRF2). As before, the 75th percentile was used as the cutoff between the Highrisk and Lowrisk strata.

An analysis of these strata revealed that although CRF2 helped shift some 'new dwelling' blocks into the Highrisk stratum, blocks that should have been classified as Highrisk were still being excluded from the Highrisk stratum. For example, larger than average sized blocks (60-80 dwellings in size) with a moderate number of new dwellings or small apartments (15-25) were assigned to the Lowrisk stratum. Such blocks should be assigned to the Highrisk stratum as these 15-25 dwellings may represent a new housing development or a cluster of basement apartments, signifying areas where higher undercoverage may be expected.

Based on these observations, the algorithm was modified to accommodate the percentage of dwellings belonging to a certain category while also taking into consideration the size of the block. For example, a block containing 60 dwellings, of which 12 are new dwellings should be considered as great a coverage risk as a block containing five dwellings, all of which are new dwellings. Therefore it was decided to use the original CRF in conjunction with additional criteria to account for both large and small blocks as well as to add more importance to the small apartment and new dwelling variables. Thus, if a block met any of the following conditions, it was considered a coverage risk and assigned to the Highrisk stratum:

1) Coverage Risk Factor $\geq$ 20; or
2) Percentage of Small Apartments $\geq$ 25%; or
3) Number of Small Apartments $\geq$ 10; or
4) Percentage of New Dwellings $\geq$ 50%; or
5) Number of New Dwellings $\geq$ 10.

The two strata that resulted from this algorithm were then stratified further by blocksize, with blocks having 100 or more dwellings forming their own sub-stratum within each of the Highrisk and Lowrisk strata. The strata created in the Precanvass area are given in Table 1.

The allotted sample size of 2,000 dwellings in the Precanvass area was then allocated among the strata. The average block size in the Large Apartment stratum was 190 dwellings, so 380 dwellings were allocated to the Large Apartment stratum to allow for two blocks to be selected from this stratum. This left 1,620 dwellings to be allocated over the five remaining strata. An allocation proportional to the size of the strata (ie. number of dwellings) and the expected undercoverage rate in each strata was used to determine the initial sample allocation of these 1,620 dwellings between the Highrisk and the Lowrisk strata:

$$n_h = n * \frac{N_h\sqrt{(P_hQ_h)}}{\sum N_h\sqrt{(P_hQ_h)}}$$

where P_h = the assumed undercoverage rate in the strata ($P_{Highrisk}$ = 2% and $P_{Lowrisk}$ = 1%)

The allocation assumed that the undercoverage rate in the Highrisk strata will be twice the undercoverage rate in the Lowrisk stratum. Had it been assumed that the undercoverage rate in the Highrisk areas would be three times greater than in the Lowrisk areas, over half the Precanvass sample would have been allocated to the Highrisk strata. Although the creation of the strata was based on sound assumptions (ie. small apartments and new dwellings are prone to higher undercoverage), these strata are for the most part experimental. Thus, since their reliability is somewhat unknown, a conservative approach was taken and undercoverage rates of 2% in the Highrisk strata and 1% in the Lowrisk strata were assumed. These assumptions will be evaluated at the conclusion of the study.

Table 1: Strata in the Precanvass Area

	STRATA						
	LARGE APTS.	SPLIT BLOCKS	HIGHRISK		LOWRISK		Total
			< 100 Dwellings	≥ 100 Dwellings	< 100 Dwellings	≥ 100 Dwellings	
# of Blocks	517	93	1,043	163	2,870	148	4,834
# of Dwellings	98,624	4,309	38,754	26,440	83,035	22,565	273,727
# Small Apts.	1,378	12	6,228	1,099	1,953	153	10,823
# New Dwellings	2,744	2,693	12,221	7,884	3,821	450	29,813

The above allocation resulted in allocating 779 dwellings to the Highrisk sample and 841 dwellings to the Lowrisk sample. The 779 dwellings in the Highrisk sample were then allocated among the three Highrisk strata. An attempt was made to allocate the sample proportional to the number of dwellings in each stratum. The average number of dwellings per block was used to determine the number of blocks that would be selected from each stratum. Some sample was shifted from the Highrisk non-large block stratum to the other two Highrisk strata to allow for a sample size of at least two blocks in each stratum. A similar procedure was followed in the Lowrisk strata where some Lowrisk non-large block sample was shifted to the Lowrisk large-block sample in order to ensure that two blocks could be selected from the Lowrisk large-block stratum.

4.2 Prelist Area Sample Design

As in the Precanvass area, the sampling unit was a Census block. Here the average block contained thirty dwellings. There were over 1,400 blocks covering 47,000 dwellings in the Prelist area.

Prelist blocks were stratified in a manner similar to that used in the Precanvass area: blocks containing large apartment buildings; blocks that were split in the field; blocks considered a greater coverage risk and blocks considered less of a coverage risk all formed their own strata. A simplified version of the algorithm used in the Precanvass area was then used to differentiate between Highrisk and Lowrisk blocks. However, 1991 Census data could not be used to obtain further dwelling information (ie. dwelling age, occupancy status and tenure) since the 1991 Census household number - address relation was not known. This is because in 1991 an Address Register did not exist for the Prelist area as it did for the Precanvass area. Therefore, only the percentage of dwellings in each remaining block that were small apartments was calculated. This variable was obtained from the address list created during the field listing in these areas.

A Prelist block was categorized as Highrisk if it met either of the following two conditions:

1) Percentage of Small Apartments $\geq$ 25%; or
2) Number of Small Apartments $\geq$ 10.

The strata created in the Prelist areas are given in Table 2. Further sub-stratification according to block size was not done as there were very few blocks with more than 100 dwellings.

The allotted sample size of 1,000 dwellings in the Prelist area was then allocated among the strata. This was done in the same manner as the allocation of the Precanvass sample. Since an average block in the Large Apartment stratum contained 120 dwellings, 240 dwellings were allocated to the Large Apartment stratum to allow for the selection of two blocks. The remaining 760 dwellings in the sample were then allocated between the Highrisk strata and the Lowrisk stratum. As in the Precanvass area, this was done proportional to the number of dwellings in the strata and the expected undercoverage rate in the strata. As before, the assumed undercoverage rate was 2% in the Highrisk strata and 1% in the Lowrisk stratum.

4.3 List/Leave Area Sample Design

As the traditional drop-off method is used for the delivery of Census questionnaires in these areas, an address list was not created in advance for this area. There are approximately 100,000 dwellings in the List/Leave area encompassing approximately 500 Enumeration Areas (EAs). These EAs were stratified into two strata: a town/village stratum and a non-town/village stratum. Five EAs were then randomly selected from each stratum with each selected EA getting broken into segments of about 50 dwellings each. One segment from each selected EA was then randomly selected.

Table 2: Strata in the Prelist Area

	STRATA				
	LARGE APTS.	SPLIT BLOCKS	HIGHRISK	LOWRISK	Total
# of Blocks	68	93	227	1,019	1,407
# of Dwellings	8,171	2,028	8,509	28,202	46,910
# Small Apts.	161	92	2,481	1,387	4,121

5. Data Collection

Due to the time needed for Census to complete non-response follow-up, the DCS will not move into the field until the first week of September, three and a half months after Census Day. As dwelling coverage is the focus of this study, this time lag is not expected to have a major impact on the study since it should be relatively straightforward to determine if a dwelling existed on Census Day.

Since a reliable dwelling list is crucial for this survey, enumerators with Census experience will conduct the DCS listing and interviewing. However, to avoid biasing the results of the DCS, persons who listed for the Census are not allowed to list the same area for the DCS. Each enumerator will list between 100 and 150 dwellings (ie. about three blocks in the mailout areas or two EA segments in the List/Leave area).

5.1 Field Listing

The first phase of data collection is to conduct the field listing operation. The DCS enumerators will visit their assigned areas and record each dwelling that meets the definition of a dwelling. As contact is extremely important for discovering dwellings that are difficult to locate, the enumerators will attempt contact at each dwelling they list. The initial listing phase should take approximately one week to complete. The DCS listing records compiled by the enumerators will be returned to Head Office where they will undergo a matching operation with the Census.

5.2 Address Matching

Upon receipt of the DCS listing records in Head Office, a clerical match between the DCS and the Census will be conducted. In the Precanvass and Prelist areas, the address lists compiled by the Census (ie. the original mailout list plus any updates made during the Census) will be matched to the address lists compiled by the DCS enumerator for the same area. In the List/Leave area, the entire Census Visitation Record for the EA containing the selected EA segment will be used for matching.

Each dwelling listed by the DCS will ultimately be assigned to one of three categories:

1) Matched to the Census (same block):
A matching dwelling was found on the Census list and the block number was in agreement between the DCS and the Census. These dwellings will be considered as listed correctly by the Census and will not require follow-up.

2) Matched to the Census (different block within the EA):
A matching dwelling was found on the Census within the same EA as the selected block but the block number differed between the DCS and the Census. A field follow-up will be conducted to determine the correct block number for the dwelling. These dwellings will be used to evaluate the geocoding process of addresses.

3) Not matched to the Census:
The dwelling was not listed on the Census list for the EA that the selected block belonged to. A field follow-up consisting of a household interview will be conducted for these dwellings. These dwellings represent undercoverage in the Census. A second match to the Census database will be done at a later stage to determine if the dwelling may have been enumerated by the Census in a different EA.

Field follow-up will also be conducted for dwellings suspected as missed by the DCS. These are dwellings that were listed on the Census mailout list for the selected block but were not listed by the DCS. These dwellings will be assigned to one of two categories:

1) Error in the Census:
The dwelling is verified as non-existent and represents overcoverage on the Census mailout file.

2) Missed by the DCS:
The existence of the dwelling is confirmed and represents undercoverage in the DCS.

5.3 Field Follow-Up and Interviews

Two types of field follow-up will be conducted immediately following the address matching operation. The first type of follow-up is for dwellings that were not matched to the Census. These dwellings will be verified to ensure that they were not listed in error by the DCS. If the dwelling is a valid dwelling, a short questionnaire will be administered to the occupants of the dwelling. This questionnaire will obtain basic information on the Census Day occupants as well as information about the dwelling.

The second type of follow-up requires only a return to the dwelling and does not involve a household interview. There are two situations when such a follow-up will be conducted. One is to establish the correct Census block that a dwelling belongs to in cases where the block number recorded by the DCS differs from that on the Census mailout list. The second instance is for dwellings suspected as missed by the DCS. The existence of these dwellings will be verified to determine if they were indeed missed by the DCS or represent overcoverage on the Census mailout list. Once all follow-

up work is complete, all field materials will be returned to Head Office for data capture.

6. Post-Processing Reconciliation

Upon the completion of the field work, all DCS listing booklets and questionnaires will be data captured. The file containing the captured data will be used to conduct further matching to the Census database. A search will be done to determine if dwellings classified as missed by the Census were enumerated in an EA outside the original search area. This could occur if the Census assigned the dwelling to the wrong EA through a geocoding error. Past vintages of the address file will also be searched to determine if dwellings classified as missed by the Census were at one time on the address file only to be deleted by one of the updating components. As well, any field follow-up that was inconclusive will be reconciled and dwellings identified as erroneous Census enumerations will be verified to ensure that they were noted as such by the Census.

Weighting and estimation will be done once all dwellings have been categorized. A ratio adjustment will be done by post-stratifying according to known Census dwelling totals. The initial weights will be multiplied by the ratio of the observed total of Census dwellings in the stratum divided by the sum of the weighted Census dwellings over the stratum.

Once the weighting and post-stratification is done, estimates of the number of dwellings missed by the Census, and the DCS, will be produced. Estimates will be produced by dwelling age group, occupancy status and type of dwelling. Estimates of the number of persons missed will also be produced and all sampling variances will be calculated.

7. Evaluation and Final Report

The evaluation of the DCS will examine all aspects of the study. This evaluation will take place in early 1997. It will discuss any problems encountered during the process and make recommendations for the study if it is decided to conduct a national DCS in 2001.

The final report will summarize the results of the study. All objectives will be revisited and conclusions will be made about the impact of CE on dwelling coverage and the implications of extending CE collection methodology to the national level in 2001.

Acknowledgements

The authors would like to thank B. Allard, R. Carter, H. Choudhry and J. Tourigny for helpful comments.

References

CHOUDHRY, G.H. (1994). A Data Collection Methodology for the 2001 Canadian Census. Statistics Canada, Methodology Branch, Working Paper No. SSMD 94-007 E.

ROBINSON, J.G. and KOBILARCIK, E.L. (1995). Identifying Differential Undercounts at Local Geographic Levels: A Targeting Database Approach. Paper presented at the Annual Meetings of the Population Association of America, April 6-8, 1995.