

THE CENSUS OF CANADA: THE DWELLING CLASSIFICATION STUDY

Peter Dick

Census Research and Development Section, Social Survey Method Division
Statistics Canada, 15B RH Coats Building, Ottawa, Ontario, Canada
K1A 0T6, dickpet@statcan.ca

KEY WORDS: Census, coverage study, non-response adjustment, whole household imputation

Abstract: Since 1981, the Canadian Census has conducted a sample survey, the Dwelling Classification Survey (DCS), to study private dwellings listed as vacant on Census Day. One error of classification that affects the Census counts is the misclassification of dwellings that were occupied on Census Day, as vacant. This error results in an undercount of persons and of occupied dwellings. One objective of the DCS is to include the estimate of persons living in “vacant” dwellings into the final Census counts. In 2001, the DCS was expanded to include private dwellings that did not return a Census form but were listed as occupied. This enlargement of the DCS affects two major programs. First, the DCS now estimates the number of persons living in Census “non-response” dwellings beyond the estimate of persons living in “vacant” dwellings. These estimates were produced in time to be included in the initial Census release. Secondly the population estimates will use a new methodology for estimating the number of persons living in Census “non-response” dwellings. The results and the impact of this survey are described.

1. Introduction

The Census of Canada is conducted every five years, the most recent being May 15, 2001. Since 1971, the Canadian Census collection has used a list/leave methodology to collect the required information from about 98% of the population. This methodology consists of a Census Representative (CR) listing all private dwellings in a pre-selected, well delineated geographic area, called an Enumeration Area (EA), containing between 300 and 600 private dwellings, and leaving a Census questionnaire at all the listed dwellings. Overall there were approximately 35,000 EAs needed for Census collection. The respondent is responsible for completing and returning the questionnaire. A problem frequently faced by the CR after Census Day is that of questionnaires that have not been returned. This occurs due to one of two reasons: either the dwelling is vacant and has no residents in it or the respondents in the dwelling have not returned the questionnaire. The way in which the CR deals with the classification of these dwellings can have a profound impact on the quality of the final Census counts.

A CR typically can make two errors of dwelling classification that impact the Census counts. First, some dwellings, initially classified as non-respondents by the CR, are actually vacant thus creating an overcount of persons and occupied dwellings in the Census. Secondly, some dwellings, actually occupied on Census Day, are incorrectly classified as vacant. This error - known as an occupied vacant - results in an undercount of persons and of occupied dwellings. There is some compensation to these errors since they move in opposite directions with respect to coverage of persons. There is no reason, however, to expect these errors to cancel out.

Prior to 2001, the Census handled these classification errors in different ways. The general approach to vacant dwellings, since 1981, has been to use a coverage improvement survey, the **Dwelling Classification Survey (DCS)**, to classify the “true” status of dwellings listed as vacant in the Census. The DCS is conducted in a sample of EAs after the Census non-response field follow up operation had been completed in a sample of EAs. All dwellings in the selected EA which were classified by the CR as vacant receive a questionnaire designed to determine the true occupancy status on Census Day. The questionnaires are then processed and the dwellings are classified as to their “true” occupancy status on Census Day. The number of “occupied” vacants are estimated from this sample and the Census database is then adjusted to account for the persons living in these “occupied” vacants. The results of the DCS for occupied vacants are required for the initial Census release.

The traditional Census approach to non-response dwellings was completely different. In 1996, a special collection form - the “Form 4” - was generated in the field for non-response dwellings. These “Forms 4” were only to be used when it was clear in the field that a completed Census questionnaire would not be returned from the dwelling. The CR attempted to determine the household size and if this was not possible then they recorded that the household size was ‘unknown’. The number of non-response dwellings was strictly controlled by Census collection: an EA with more than 2% “Forms 4” was rejected and returned to the field for further collection until the 2% threshold had been reached. Late in the collection period, this 2% rule was lifted but generally very few EAs exceeded this value. After the Form 4 was created, it was data

captured and placed on the Census database. If the household size was unknown - about 75% of the time - an algorithm in Census processing was used to generate a household size by borrowing the household size from a nearby neighbour in the processing stream. Once all non-response dwellings had a household size, the complete responses for these households were imputed.

In 1996, as an experiment to evaluate the Census approach to non-response dwellings, the DCS was expanded to include non-response dwellings (Dick, 2001). The general procedure was to re-visit all non-response dwellings in the selected DCS sample EAs and administer the DCS questionnaire to a knowledgeable respondent. These questionnaires were edited for consistency and then the non-response dwellings were classified as to whether they were vacant, not a private dwelling, not a habitable private dwelling or occupied. If they were occupied (this is the situation the Census is most concerned about) then the number of usual residents was determined. The results of this research indicated that the Census methodology of assuming all non-response dwellings were occupied and, when the household size was unknown, imputing the household size from a nearby neighbour was generating significant overcoverage in the Census. Consequently, it was decided in 2001 to expand the DCS to estimate the persons living in the non-response dwellings in addition to the vacant dwellings and to adjust the household size distribution through imputation for the non-response dwellings.

This paper summarizes the results of the 2001 DCS. Section 2 outlines the DCS sample design, shows some of the estimates and presents an evaluation of these estimates. Finally, Section 3 discusses the impact of the DCS on the 2001 Census counts and the 2001 population estimates. More details on the results of the 2001 DCS can be found in Dick (2002).

2. Design and Estimation

2.(a) Design

The sample design that the DCS used in 2001 is a slight modification of the previous designs used in the 1981 to 1996 Censuses¹ (Statistics Canada, 1999). Previously, only vacant dwellings in the list/leave areas²

¹ Prior to 2001 the DCS was known as the Vacancy Check Study.

² Excluded from the DCS sample are collective, canvasser and Indian Reserve EAs which implies the entire Territory of Nunavut is out of scope. Overall, approximately 98% of the population is in the list/leave EAs.

were included in the target population. For 2001, however, the non-response dwellings have to be accounted for in the design. The basic design was to select a sample of 1,400 list/leave EAs and, within a selected EA, include all the private vacant and non-response dwellings.

In 2001 a slight modification was proposed in an effort to include a higher portion of the non-response dwellings. When the DCS interviewer created their assignment, all dwellings for which a questionnaire had not been returned were to be included. However, the results from the field seemed to indicate this procedure was not followed exactly. It appears, in some locations, the 1996 procedure, of listing only the vacant and non-response dwellings identified at the start of DCS collection, was followed. This created a problem with constructing the frame since many non-response dwellings were not identified until after DCS collection had finished.

The sample allocation of the 1,400 EAs was carried out in a two step procedure. First, two take all strata were created for the urban parts (Whitehorse and Yellowknife) of the Territories. Then, as in 1996, 1,304 EAs were allocated to the provinces using Bankier's power allocation scheme (Bankier, 1988). This procedure is a compromise approach that balances equal Coefficient of Variation for each province with a minimum national CV: details on this stage of the allocation can be found in Ramsay (2001). The final step was to allocate an extra 38 EAs to Montreal (15), Toronto (15) and Vancouver (8) on the assumption that these centers would have more non-response dwellings than other areas.

2.(b) Estimation

The purpose of the DCS processing is to use the results of the questionnaires to classify every private dwelling selected in the sample into one of four major categories: occupied dwelling, vacant dwelling, a dwelling that is not habitable and not a private dwelling. There are some other minor categories but over 99% of the dwellings in the sample will eventually be classified into one of these four groups. The general approach in the DCS was to process the entire sample as a whole irrespective of the initial classification of the dwelling. After the processing and classification were finished, the DCS sample was matched to two separate Census files: one, a listing of all non-response dwellings in the list/leave EAs and the other, the listing of all vacant dwellings in the list/leave EAs. These matches were done on a dwelling to dwelling basis. Any dwelling that did not match to either of these files was considered to be a regular Census respondent and, consequently, out-of-scope for the DCS.

After processing has been completed, the

DCS results are then weighted separately³ for non-response dwellings and vacant dwellings using the following three step procedure:

- (a) an initial weight equal to the inverse probability of selecting the EA from the list of eligible EAs was assigned. An EA that was selected and later split in the field would have the same weight applied to the resulting split EA. This is identical to the approach used in 1991 and 1996;
- (b) a factor to account for non-response to the DCS (at the selected EA level) was calculated by inflating the initial weight by the ratio of dwellings in the selected EA over the actual number of valid in-scope dwellings in the DCS. This approach is new to 2001; and
- (c) a post-stratification adjustment factor was calculated in order to ensure the DCS estimates are calibrated to the known total of dwellings in the Census for 25 sub-regions. This part of the weighting methodology is identical to the approach used for the 1991 and 1996 VCS estimates.

The post-stratification split each province (other than PEI) into Urban/Rural parts. In addition, the 3 largest CMAs - Toronto, Montreal and Vancouver - were each a separate stratum in Ontario, Quebec and British Columbia respectively. The three possible strata within a province are:

CMA: those urban areas with over 1,000,000 persons in one province (i.e. Toronto, Montreal and Vancouver);

Urban: those urban areas with between 50,000 and 1,000,000 persons; and

Rural: rural and urban areas with under 50,000 persons.

The results of the 2001 DCS are summarized in Table 1. The table is split into two major parts: the Census estimates and the DCS estimates. The Census estimates are simply the number of persons living in non-response dwellings using the old Census methodology - the sum of the known number of persons identified at Census collection and Census nearest neighbour imputation when the household size is not known. The DCS estimates of the number of persons are split into two components corresponding to non-response and vacant dwellings. The standard errors for the DCS estimates are also indicated along with the totals.

The main item of interest in Table 1 is the

³ Since the estimates were required at different times during Census processing, the non-response and the vacant portions are weighted independently. In 2006 it is likely this distinction will not be used.

difference between the Census count and DCS estimates of persons living in non-response dwellings. Overall the difference is almost 100,000 (i.e. 414,905 - 317,587 = 97,318) persons and in every province the Census methodology has counted more persons. Even if the sampling variability of the DCS is taken into account, these differences would be declared to be significant. Clearly a large part of this difference must be due to a bias underlying one of the approaches. In order to understand how this difference arises in the estimates, Table 2 shows a breakdown of the two estimates of persons living in non-response dwellings.

The construction of Table 2 is straightforward. The Census observed 179,785 dwellings that were classified as non-response with 414,905 persons living in them, hence giving an average of 2.31 persons per household. The DCS classified 36,106 of the non-response dwellings as vacant, hence the DCS had only 143,681 occupied dwellings with 317,587 persons. Overall the difference between the Census and the DCS estimate is 97,318 persons. However, the Census would have put, on average 2.31 persons into the 36,106 vacant dwellings, or about 82,788 persons. Thus we can see that of the difference of 97,318 persons, 82,788 are due to the error of the Census assuming all non-response dwellings are occupied and the remainder of the difference (14,530) is due to the difference between the two methods for the household sizes in occupied dwellings.

Examining Table 2 more closely, we can see that the methodology used by the Census generates household sizes that are quite close to the DCS household size. However, as in the results from the 1996 research, the Census assumption of at least one usual resident in every non-response household seems to be generating an overcoverage of persons. This result is fairly constant across all provinces. The vacant rate (vacant dwellings over total non-response dwellings) per province varies from about 15% to 25% with the national rate being about 20%. The other item of interest in Table 2 is the closeness of the household sizes in occupied dwellings. However, it appears that if the dwelling was occupied, then both approaches - the Census and the DCS - produce household sizes that are fairly comparable, especially in the larger provinces of Ontario, Quebec and Alberta.

3. Results and Conclusions

The impact of the new DCS methodology for the Census counts is summarized in Table 3 below. This table is split into two major components: the unadjusted Census counts and the adjusted Census counts. The unadjusted Census counts are essentially the number of persons counted on Census questionnaires that clearly indicated the number of persons living in the household.

The Census adjustments refer to the additions made to these base Census counts. The adjustments have two major components: the persons living in vacant dwellings and the persons living in non-response dwellings. Table 3 shows how these were combined to form the counts the Census released on March 12, 2002, and the old (1996) method for calculating the Census release.

To calculate the Census release for 2001 (new) or 1996 (old) the common elements are the unadjusted (base) Census counts, the estimated number of persons living in vacant dwellings and the number of persons living in (out of scope) non-response dwellings outside the area covered by the DCS. The only change is in the calculation of persons living in non-response dwellings in the DCS areas. The final release, then, was changed by almost 100,000 persons by using the DCS approach compared with the 1996 approach.

The table above can also be used to calculate the percentage of households in the Census database that had their responses imputed: this will be called the Whole Household Imputation Rate. Since persons living in vacant dwellings and persons living in non-response dwelling have no information, the Census - once the household size has been determined - imputes an entire household into these dwellings⁴. Assuming that outside of these two types, the Census counts did not require whole household imputations, then rates can be calculated from the ratio of vacant and non-response persons over the final Census counts. Note that the two rates can be calculated using either the DCS estimate (New Methodology in Table 3) or the Census estimate (Old Methodology in Table 3) for non-response persons. By using both methods it is possible to compare the rates to the 1996 Census. Note the DCS estimates for the 1996 Census are from the research conducted after the 1996 Census (see Dick, 2001). These rates are shown on the following figure:

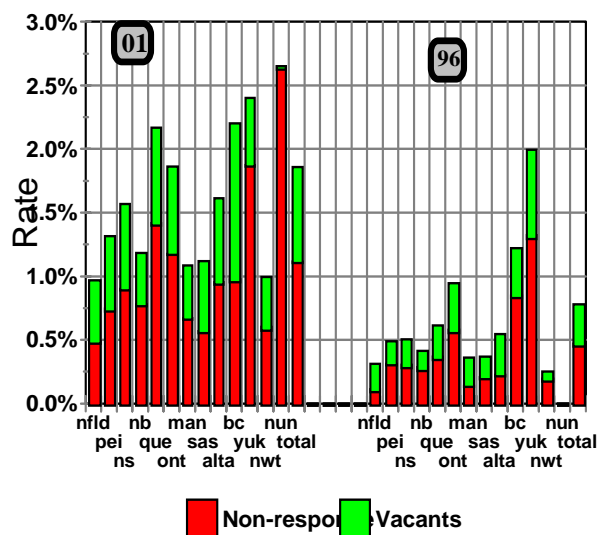
Note the massive jump in rates nationally, from under 1% in 1996 to almost 1.9% in 2001 when the DCS estimates are used. The provincial distribution shows some interesting variations. Quebec and the Atlantic provinces have had the largest increase between 1996 and 2001, while Ontario and British Columbia have increased but not on the scale of the others. If the whole household imputation is an indicator of Census

⁴ This isn't quite true. The Census handles the non-response dwellings with a whole household imputation method and the vacant dwellings through a weighting approach. However, it is true to note that all persons in this study that appear in the Census database appear through a methodological adjustment.

collection difficulties, then it would appear that Census collection was not only much more difficult in 2001 but also that these difficulties were much more widespread than in the past.

However it should be noted that the DCS methodology actually created fewer persons than the old Census methodology would have. From Table 3 we can see that in 2001 the new DCS methodology added 222,720 persons in vacant dwellings and 317,587 in non-response dwellings. Thus overall the DCS created 1.86% of persons on the Census database. By contrast, if the previous methodology for non-response dwellings had been used, 414,905 (plus the 222,720 in vacant dwellings) would have been added to the Census

Whole Household Imputation Rates



database, or 2.18%. Thus the new DCS methodology has reduced the imputed persons in 2001 by over 0.3%.

One of the results of the 1996 coverage studies was that the overcoverage of persons created by the Census methodology for non-response dwellings was not removed from the final population estimates. Thus if the 1996 methodology had been continued in 2001, there would be an overcount of persons. The number of persons living in non-response dwellings has been a difficult group for the population estimates program to account for prior to 2001.

In the past, the Census estimates of persons living in non-response dwelling were assumed to be valid and were entered unadjusted directly into the population estimates. In 2001, a new procedure is being introduced in order to control this estimate. The coverage studies will now assume that any person with a Census Day address in a non-response dwelling is

missed. This is identical to how persons with a Census Day address in a vacant dwelling have been handled since 1981. When it comes time to generate the population estimates, the Census estimate of persons in non-response dwellings will be removed - or netted out - and the coverage study, or the DCS estimate, will be submitted in its place. Assuming the DCS provides an unbiased estimate of these people, then the difference between the Census and the DCS estimate would be an estimate of the Census overcoverage in these dwellings. This estimate - which was present in the final population estimates in 1996 - will then be removed from the 2001 estimate.

References:

Bankier, M. (1988). Power allocations: determining sample sizes for subnational areas. *The American Statistician* **42**: 174 -177.

Dick, J.P. (2001) The Dwelling Classification Study. *Presentation at the Federal - Provincial - Territorial Meeting on Demography, September 21, 2001.* Ottawa, Ontario.

Dick, J.P. (2002) The Dwelling Classification Survey: Results and Methods. *Presentation at the Federal - Provincial - Territorial Meeting on Demography, September 12, 2002.* Ottawa, Ontario

Ramsay, T. (2001). Methodology for the 2001 Dwelling Classification Study. *Internal Statistics Canada report: January 8, 2001.*

Statistics Canada (1999). Coverage: 1996 Census Technical Reports. *Catalogue No. 92-370-XPD* Ottawa, Ontario.

Acknowledgments:

The author would like to thank Dave Dolson, Réjean Lachapelle, Mike Bankier and Karen Switzer for many comments that improved this paper.

Table 1: DCS Estimates of Persons Living in Non-Response and Vacant Dwellings

Province/Territory	Census Estimates Non-Response Dwellings	Dwelling Classification Study Estimates					
		Non-Response		Vacant		Total	
		Persons	St Error	Persons	St Error	Persons	St Error
Newfoundland	3,496	2,268	179	2,438	434	4,707	469
Prince Edward Island	1,161	978	206	768	184	1,746	276
Nova Scotia	11,284	8,042	511	6,058	756	14,099	913
New Brunswick	7,665	5,501	66	2,954	593	8,458	597
Quebec	139,541	100,741	7,789	54,268	6,610	155,009	10,215
Ontario	148,604	124,825	12,282	78,393	10,056	203,219	15,873
Manitoba	8,790	6,602	161	4,663	692	11,260	711
Saskatchewan	7,359	5,142	172	5,383	936	10,517	951
Alberta	34,133	26,982	423	19,564	3,078	46,544	3,107
British Columbia	52,448	36,269	2,862	47,933	4,389	84,202	5,240
Yukon	248	177	12	151	24	326	27
Northwest Territories	176	60	29	147	50	206	58
Total	414,905	317,587	14,841	222,720	13,269	540,293	19,908

Table 2: Decomposition of the Estimates of Non-response Dwelling

Province/Territory	Census			Dwelling Classification Study					Difference in Persons		
	Dwellings D	Persons P	P/D	Classified as Occupied			Classified as Vacant		Total	Persons in Vacant Dwellings	Imputed Person
				Dwellings D	Persons P	P/D	Estimate	Rate			
Newfoundland	1,431	3,496	2.44	1,185	2,268	1.91	246	17%	1,228	601	627
PEI	508	1,161	2.29	392	978	2.49	116	23%	183	265	(82)
Nova Scotia	5,062	11,284	2.23	3,980	8,042	2.02	1,082	21%	3,242	2,412	830
New Brunswick	3,304	7,665	2.32	2,676	5,501	2.06	628	19%	2,164	1,457	707
Quebec	65,787	139,541	2.12	50,834	100,741	1.98	14,953	23%	38,800	31,717	7,083
Ontario	59,782	148,604	2.49	48,396	124,825	2.58	11,386	19%	23,779	28,303	(4,524)
Manitoba	3,798	8,790	2.31	3,254	6,602	2.03	544	14%	2,188	1,259	929
Saskatchewan	3,246	7,359	2.27	2,313	5,142	2.22	933	29%	2,217	2,115	102
Alberta	14,197	34,133	2.40	11,834	26,982	2.28	2,363	17%	7,151	5,681	1,470
British Columbia	22,471	52,448	2.33	18,697	36,269	1.94	3,774	17%	16,179	8,809	7,370
Yukon	118	248	2.10	90	177	1.97	28	24%	71	59	12
NWT	81	176	2.17	30	60	2.00	51	63%	116	111	5
Canada	179,785	414,905	2.31	143,681	317,587	2.21	36,104	20%	97,318	82,788	14,530

Table 3: Decomposition of the Census Estimates

Province/Territory	Unadjusted Census Counts (Base)	Census Adjustments				Census Counts	
		Vacant Dwellings	Non-response Dwellings			New Methodology	Old Methodology
			Out of Scope EAs	List/leave EAs			
				DCS Estimates	Census Estimates		
Persons							
Newfoundland	507,958	2,438	266	2,268	3,496	512,930	514,158
PEI	133,516	768	32	978	1,161	135,294	135,477
Nova Scotia	893,779	6,058	128	8,042	11,284	908,007	911,249
New Brunswick	720,868	2,954	175	5,501	7,665	729,498	731,662
Quebec	7,080,522	54,268	1,948	100,741	139,541	7,237,479	7,276,279
Ontario	11,197,397	78,393	9,431	124,825	148,604	11,410,046	11,433,825
Manitoba	1,107,434	4,663	884	6,602	8,790	1,119,583	1,121,771
Saskatchewan	967,969	5,383	439	5,142	7,359	978,933	981,150
Alberta	2,926,844	19,564	1,417	26,982	34,133	2,974,807	2,981,958
British Columbia	3,821,652	47,933	1,884	36,269	52,448	3,907,738	3,923,917
Yukon	27,986	151	360	177	248	28,674	28,745
NWT	36,988	147	165	60	176	37,360	37,476
Nunavet	26,036	0	709	0	0	26,745	26,745
Total	29,448,949	222,720	17,838	317,587	414,905	30,007,094	30,104,412