Census Population Coverage Error: Results and Methods

Colleen Clark

Statistics Canada, Social Survey Methods Division,
Census Coverage Studies Section, 120 Parkdale Avenue, Ottawa, Ontraio, Canada, K1S 2A3,
Colleen.Clark@statcan.gc.ca

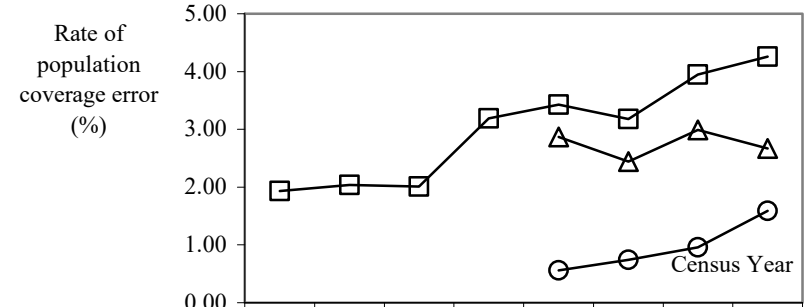
Keywords: Coverage error; census data quality; census coverage error; population estimates; undercoverage; undercount; census undercoverage; census undercount; overcoverage, overcount; census overcoverage; census overcount; coverage rates, record linkage; data quality

1. Introduction

The 2006 Canadian Census of Population required the participation of the entire population of Canada, some 32.5 million people distributed over a territory of 9 million square kilometres. Although there are high quality standards governing the collection and processing of the data, it is not possible to eliminate all errors. Coverage errors are one of the most important types of error since they affect not only the accuracy of the counts of the various census universes, but also the accuracy of all of the census data describing the characteristics of these universes. There are two types of population coverage error. *Population undercoverage* refers to the error of the census excluding someone who should have been enumerated. *Population overcoverage* refers to the error of the census either enumerating someone more than once or including someone who should not have been enumerated. The latter is negligible in the Canadian context. The net impact of undercoverage and overcoverage on the size of a population of interest is known as *population net undercoverage*. Undercoverage is usually larger than overcoverage.

Estimates of coverage error are produced only for the census population universe. Chart 1 gives the population undercoverage rate for the 1971 Census through to the 2006 Census, the population overcoverage rate and the population net undercoverage rate for the 1991 Census through to the 2006 Census. As the chart shows, population coverage error is a growing data quality concern. Undercoverage, now at 4.26% for the 2006 Census, has more than doubled since 1981when it was 2.01%. Population overcoverage was first measured for the 1991 Census. Although overcoverage is the smaller coverage error, it has doubled since 1996.

Chart 1: 1971 Census to 2006 Census rates of population coverage error for Canada



	1971	1976	1981	1986	1991	1996	2001	2006
Undercoverage	1.93	2.04	2.01	3.19	3.43	3.18	3.95	4.26
Overcoverage					0.56	0.74	0.96	1.59
Net Undercoverage					2.87	2.44	2.99	2.67

Notes:
1. Blank cells indicate data not available.

2. Overcoverage was first measured for the 1991 Census. The increase from 1991 to 1996 is due in part to a change in methodology.
Source: Statistics Canada, 1971 to 2006 Census Coverage Studies

There are two post-censal studies that measured 2006 Census population coverage error. The Reverse Record Check (RRC) estimated population undercoverage while the Census Overcoverage Study (COS) estimated population overcoverage. The Dwelling Classification Survey (DCS), carried out during census processing, addresses coverage error from non-response dwellings and occupied dwellings classified in error as unoccupied. Census data are adjusted for this type of coverage error. Census data are not adjusted for population net undercoverage. Rather, estimates of net undercoverage are used in the production of the base population for Statistics Canada's demographic estimates of population. Population estimates are used for calibration in many surveys such as the Canadian Labour Force Survey. Preliminary estimates of 2006 Census population coverage error were released March 27, 2008. Because estimates of population net undercoverage directly impact transfer payments from the federal government to the provincial and territorial governments, a lengthy and detailed certification exercise with the provincial and territorial statistical focal points was carried out over the following four months. Some changes were subsequently made to the methodology for non-response adjustment in the RRC to arrive at the final estimates released on September 29, 2008. .

This paper presents selected results and methodologies from the 2006 Census coverage studies. Section 2 gives an overview of the 2006 Census methodology, lists some of the sources of coverage error and defines the measures used to quantify coverage error. An overview of the methodology of the RRC and the COS is in section 4. Section 5 presents census collation undercoverage, a new measure introduced for the 2006 Census. Section 6 gives some concluding remarks.

2. Coverage Error

2.1 2006 Census Methodology

Overall, the 2006 Census of Population moved from a decentralized, manual operation to a more centralized and automated one: Questionnaires were mailed by Canada Post Corporation in a majority of urban areas; the Address Register, which is updated by listing operations, provided the mailing addresses; and item non-response follow-up became centralized. Further, in some regions it was difficult to recruit enough staff.

In the 2006 Census, about 98% of households were enumerated via self-enumeration. For the 70% of dwellings located in Mail Out areas Canada Post delivered a census questionnaire. In the List Leave areas, covering 28% of the dwellings, households received their questionnaire from an enumerator. The canvasser method, whereby a census enumerator visited a household and completed a questionnaire, was used for remote and northern areas of the country, on most Indian reserves, and in large urban downtown areas where there was a concentration of transient residents. For the first time, the 2006 Census offered all households in Canada the option of completing their questionnaire online. Approximately 18% of households responded via the Internet.

As part of census processing, hot deck imputation was carried out to account for persons living in non-response dwellings and occupied dwellings classified in error as unoccupied. This procedure increased the number of occupied dwellings by 3.6% and decreased the number of unoccupied dwellings by 5.2%. Further, 70.9% of all dwellings classified as non-response were actually occupied. Imputation added 571,521 persons to the census database, 1.81% of the census count of population. Compared to the 2001 Census, a large increase in both the number of non-response dwellings and the number of misclassified dwellings resulted in doubling the number of persons imputed.

2.2 Sources

Coverage error reflects error on the part of the respondent and on the part of census operations. Undercoverage can occur in the first stage of the census if the list of dwellings constructed to cover the census dwelling universe is incomplete. Dwellings that have just been built are particularly difficult to list. Depending on the date that they are completed and the timing of the census frame update, there is a risk

that a new dwelling may not be included in the census frame thereby contributing to undercoverage. Coverage errors can also be introduced during the processing stage when records for persons or households are erroneously created or cancelled.

Coverage error is most likely to have occurred during the field collection stage. A respondent can erroneously omit someone whose usual place of residence, according to the census rules, is the dwelling. This is undercoverage. Conversely, a respondent can erroneously include someone whose usual place of residence is elsewhere. Overcoverage occurs if that person is listed at their usual residence or elsewhere. Examples of overcoverage include: (i) children whose parents have separate residences and each parent includes the children on their census form, (ii) persons who need to reside away from their family for reasons of work who are listed on their family's form and also on the form for the dwelling they live in while working, and (iii) students away at school who are listed both by their roommates and their parents. These living arrangements may also result in undercoverage. Although efforts are made to enumerate homeless persons, the risk of undercoverage is higher for this population. As for many surveys, reduced respondent participation continues to be an issue for the census. This is evidenced by both increased non-response and increased undercoverage.

An individual's usual place of residence refers to the dwelling in which, as of May 16, 2006, a person lives most of the time. In most cases, it is easy to determine someone's usual place of residence. However, there are a number of situations where the process is not elementary and special rules have been created in order to define an individual's usual place of residence. Although the rules are set out in the census form, the list is long and there may be comprehension challenges for some respondents.

2.3 Definitions

If T is the true number of persons in the census target population and C is the published census count of the number of persons in the census target population, then the error in using C instead of T is $N = T - C$. This error, denoted as N, is net population coverage error. If U is population undercoverage denoting the number of persons not included in C who should have been, and O is population overcoverage denoting is the number of persons included in C who should not have been, then an estimate of T is given by $\hat{T} = C + \hat{N} = C + \hat{U} - \hat{O}$. Since undercoverage is more common than overcoverage, net population coverage error is stated as net undercoverage. Let us assume that overcoverage from persons included in C that are not in the census target population is zero (Empirical evidence from past coverage studies have shown this number to be very small.). Therefore, $\hat{O}$ is restricted to an estimate of the number of excess enumerations, usually duplicates. Census population coverage error can be usefully expressed as rates relative to the true population: $\hat{R}_U = 100 * \frac{\hat{U}}{\hat{T}} = 100 * \frac{\hat{U}}{C + \hat{N}}$, $\hat{R}_O = 100 * \frac{\hat{O}}{\hat{T}} = 100 * \frac{\hat{O}}{C + \hat{N}}$, and $\hat{R}_N = 100 * \frac{\hat{N}}{\hat{T}} = 100 * \left(\frac{\hat{U} - \hat{O}}{C + \hat{N}} \right)$. A positive net undercoverage rate indicates than undercoverage is larger than overcoverage; there are more people not included in the published census count C than there are excess enumerations. This has been, and continues to be, the experience of the Canadian census. For some domains of interest, however, there is net overcoverage where the number of persons missed is smaller than the number of excess enumerations. It is the goal of the census coverage studies to estimate U, O and N, the corresponding rates, and the standard errors of these estimates for a variety of geographic and demographic characteristics; most notably for each province and territory.

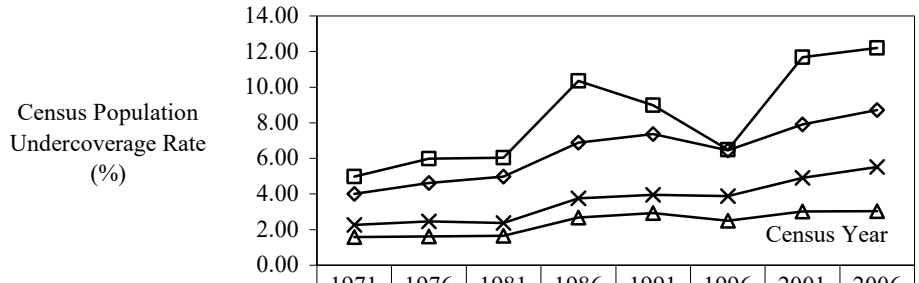
3. Estimates

Compared to the 2001 Census, coverage error has increased. The rate of undercoverage increased from 3.95% (with a standard error (s.e.) of 0.13) to 4.26% (s.e. 0.17), and the rate of overcoverage increased

notably from 0.96% (s.e. 0.05) to 1.59 (s.e. 0.01). Since the overcoverage rate increased more than the undercoverage rate, the estimated net undercoverage rate decreased from 2.99% (s.e. 0.14) in 2001 to 2.67% (s.e. 0.17). It has been important to stress with users that even though net undercoverage is down, coverage error has increased because both undercoverage and overcoverage has increased. Even though there are high quality standards governing all census operations, changes in census methodology from 2001 to 2006 may have a role in these increases.

Chart 2 shows rates of population undercoverage by age and sex for Canada for the 1971 Census through to the 2006 Census. The demographic trends observed in 2006 continue historical trends. A profile of the person most likely to have been counted as undercoverage in the 2006 Census emerges from cross-tabulations as male, between 18 and 34 years of age, and single. Mother tongue other than English or French (immigrants from non-English/non-French countries) is also important. The rate of undercoverage for men was almost twice the rate for women, 5.51% (s.e. 0.26) versus 3.04% (s.e. 0.23). For both men and women, undercoverage was highest for young adults aged 18 to 34. Among young adult males, undercoverage was over 10% for the 18-19 age group at 10.06% (s.e. 2.45), the 20 to 24 age group at 12.21% (s.e. 1.12), and the 25 to 34 age group at 11.42% (s.e. 0.86). High undercoverage is also related to marital status, highest for single persons and separated persons.[1].

Chart 2: 1971 Census to 2006 Census rate of population undercoverage by age and sex for Canada



	1971	1976	1981	1986	1991	1996	2001	2006
Males	2.27	2.46	2.37	3.75	3.95	3.89	4.90	5.51
Females	1.59	1.61	1.65	2.68	2.93	2.49	3.02	3.04
Males 20-24	4.97	5.99	6.03	10.36	8.98	6.48	11.68	12.21
Females 20 - 24	4.01	4.62	4.98	6.89	7.36	6.45	7.91	8.71

Source: Statistics Canada, 1971 Census to 2006 Census coverage studies.

There are some demographic trends in overcoverage. Across the provinces and territories, overcoverage varied much less than undercoverage did. Overcoverage was only slightly higher for males than females, 1.62% (s.e. 0.02) versus 1.56% (s.e. 0.01). Overcoverage is highest for children and young adults from age 5 to 34 with the exception of the youngest children. The rate for those 0-4, 1.35% (s.e. 0.07), is similar to the rates for the 35+ group. Rates are highest for young adults aged 20-24 at 2.88% (s.e. 0.11). Considering marital status, overcoverage was high for single persons

In order to route clerical flow during the 2006 COS, a code was developed classifying the type of overcoverage. Households who moved around Census Day and were enumerated at two addresses accounted for 20% of the persons who were involved in multiple enumerations. Another 20% was from identical households who were either at the same address or very close such as next door. This outcome may indicate a problem in census operations. The next largest group was children whose parents do not live in the same household who are listed once with each parent. These persons accounted for 16.9% of overcoverage. Another 12% is for students or young adults newly away from home, and 11% is for

[1] Single persons are those who report that they are not legally married and are not in a common-law relationship.

persons in two dwellings where the members of the first are a subset of the members of the second household. This is the first time that such data were available for analyzing overcoverage. Since classification of overcoverage cases was done on a trial basis and designed for work flow, the data only provide some trend information. However, it is reassuring to see that the data, except for a small number of cases where there was not enough data, easily fit the coding scheme, and, that overcoverage is coming from the types of living situations we had expected.

4. Methodology

The 2006 Census coverage studies differ from the 2001 Census coverage studies. This was the first time that the names of persons listed on all of the census forms were available electronically on the census response database (RDB). This greatly increased the efficiency of coverage studies since matching could include the name and not be restricted to just demographic characteristics. Indeed, the measurement of overcoverage was undertaken by a new study, the Census Overcoverage Study, specifically designed to exploit the use of an individual's name for identifying overcoverage.

4.1 Reverse Record Check

Canada is divided into 10 provinces and 3 territories. The territories are the sparsely-populated northern regions. In the Revere Record Check (RRC), a random sample of individuals representing the 2006 Census target population was taken from frames independent of the census. For the provinces, this consisted of a list of persons enumerated in the 2001 Census, a sample of persons missed in the 2001 Census, a list of intercensal births from provincial birth registries, and, from to federal administrative data, a list of intercensal immigrants and a list of persons who are not Canadian citizens but have a permit to temporarily reside in Canada such as students, workers, and refugee claimants. The frame for each territory is the territorial health care file. The 2006 RRC sample consisted of 67,813 persons in the provinces and 86,460 persons in the territories. [2]

In 2001, the RRC measured both undercoverage and a component of overcoverage. The introduction of the COS meant that the RRC no longer needed to measure any overcoverage. The RRC methodology was consequently changed so that not all cases were sent for field collection thereby significantly reducing field costs. The first step after selecting the sample was to search the RDB data from the survey frame and various update sources such as taxation data. If the search resulted in locating the selected person on the RDB, collection was not required. The exception was a sample of those that had been found. This sample was used to estimate some parameters of the non-response adjustment model.

When the selected person was not found, a telephone interview via computer-assisted telephone interviewing (CATI) out of the regional offices (ROs) was conducted to collect further information to declare the individual as in scope or not in scope for the census, and when in scope, to provide further data for searching. Including the sample for the non-response adjustment model, an interview was achieved for 84.2% of the 20,114 cases sent to the ROs.

Estimates of population undercoverage are based on persons in the RRC sample who were classified 'Missed'. These persons have been found to be in scope for the 2006 Census but no evidence of enumeration in the 2006 Census could be found in the Response Database. Nationally, there were 5,431 Missed sampled persons in the provinces and 676 Missed sampled persons in the territories On a weighted basis, the Missed represented 8.28% of the national sample Following classification of each person in the sample as Enumerated, Missed, or Out of Scope[3], design weights were adjusted to account for non-response whereby the total design weight of the non-respondents was shared among a group of respondents

[2] The large sample size in the territories is because a different methodology is used. The sample frames were first matched to the entire census database. Matches were classified as Enumerated if they were found in the same territory or out of scope if they were found elsewhere. All of the matched persons from the sample frames were included in the RRC sample with a weight of one. An additional sample was selected from the non-matches.

[3] A detailed classification of 12 codes was used to code both the results of searching and the groupings required for the non-response adjustment model.

most like the non-respondents in their propensity to respond, a characteristic highly correlated with the probability of being classified Missed.

It should be noted that the RRC used an early version of the Response Database. Consequently, there are some persons classified as Missed who are included in the final census database. Some enumerations were deemed too incomplete, usually invalid name date, to be used by the RRC (RRC Incomplete Enumerations). The sampled person's Census Day Address may point to a dwelling that contains imputed enumerations. Last, there were some enumerations added to the census database after the data were extracted to create the RRC database. The size of these groups is shown in Table 4.

Table 1 shows the number of persons sent for CATI collection and the number for whom collection was completed. Overall, 84.2% of the sample sent for collection resulted in an interview. Table 2 gives the unweighted distribution of the RRC sample by classification. 'Listed' and 'Mobile' are process metadata used for forming the weight adjustment groups for non-response adjustment. Listed means that the sampled person could be classified without resort to collection data. Mobile means that the sampled person's Census Day Address was an address uniquely supplied by collection. Note that non-response refers to having sufficient date from all sources, of which one may be collection, to classify the sampled person as either Enumerated, Missed, or Out of Scope. Despite the field completion rate of 84.2%, data from other sources meant that there was sufficient information to assign one of these classifications to 98.0% of the sample.

Table 1: 2006 RRC Completion counts by sampling frame for Canada

Frame	Sent	Completed	% completed
Provinces			
2001 Census	12,680	10,943	86.3
Missed	1,456	1,204	82.7
Births	802	708	88.3
Immigrants	1,473	1,171	79.5
NPR	1,485	957	64.4
Territories			
Health Care Files	2,218	1,961	88.4
Total	**20,114**	**16,944**	**84.2**

Source: Statistics Canada, 2006 Reverse Record Check.

Table 2: 2006 RRC sample classification for Canada

Classification	Number of persons	% of sample
Enumerated	**141,782**	**91.90**
Listed[1]	141,333	91.61
Not listed	449	0.29
Missed	**6,107**	**3.96**
Listed	745	0.48
Not listed mobile	1,847	1.20
Not listed not mobile	3,515	2.28
Out of Scope	**6,384**	**4.14**
Listed	2,485	1.61
Not listed	783	0.51
Non-response	3,116	2.02
Total	**154,273**	**100**

[1]Includes 84,980 for the territories sample.

Source: Statistics Canada, 2006 Reverse Record Check

4.2 Census Overcoverage Study (COS)

Table 3 gives the components of the 2006 Census estimate of population overcoverage for Canada Overcoverage is measured in two steps. Step 1 used exact matching to identify overcoverage. Step 2 used probabilistic record linkage with the aid of Statistics Canada's Generalized Record Linkage System (GRLS). In Step 1 the 2006 Census database is matched to a partial list of persons who should have been enumerated constructed from administrative data sources, mostly taxation files. This list had high coverage and had been, where possible, unduplicated. Exact matching was done using name (treated), sex, and birth date: 64.68% of the RDB resulted in a 1:1 match, 1.76% M:1. 0.05% M:M, and 33.25% resulted in no match indicating undercoverage (known) in the admin files. Records from the RDB who matched to more than one record on the admin files were automatically declared as overcoverage with a weight of one. Clerical review of a sample of matches revealed that a modest number of them were false matches from persons who were clearly different. Estimates were subsequently adjusted.

Step 2 matched the RDB Step 1 non-matches to the entire RDB using probabilistic record linkage. GRLS identified matches that were close but not exact. Each match was assigned a weight indicating the 'closeness' of the pair. Matches above a specified threshold were, as for Step 1, automatically declared overcoverage. Matches below a second specified threshold were automatically declared as not being

Table 3: Components of 2006 Census estimate of population overcoverage for Canada		
Step 1 overcoverage	Estimated number of persons with multiple enumerations	270,824
	Standard error	955
Step 2 overcoverage above upper	Number of persons with multiple enumerations	55,423
Step 2 overcoverage between thresholds	Estimated number of persons with multiple enumerations	180,523
	Standard error	3,001
Total COS overcoverage	Estimated number of persons with multiple enumerations	506,770
	Standard error	3,001
Final estimate of overcoverage (applying AMS adjustment.)	Estimated number of persons with multiple enumerations	515,715
	Standard error	3,207
Source: Statistics Canada, 2006 Census Overcoverage Study,		

overcoverage. Between the two thresholds, a sample was taken. These pairs of potential duplicates were clerically reviewed using names and various demographic characteristics including household composition in order to determine whether or not overcoverage had occurred. As part of evaluating the COS, an independent study was conducted to measure overcoverage. The 2006 Automated Match Study repeats a methodology carried out since overcoverage was first measured for the 1991 Census. Comparison of COS cases and AMS cases revealed that that some types of overcoverage were not included in the AMS. COS estimates were adjusted using the AMS estimates for the overcoverage not covered by the COS. Table 3 gives the amount of overcoverage identified in Step 1, in Step 2 from pairs above the threshold and pairs between the thresholds, the total COS overcoverage, and the adjusted COS estimates.

Table 4: Components of Estimated Population Coverage Error for Canada	
Component	**Number of perons**
MHAT: Missed from RRC	2,846,000
X: Netted out factor	1,462,000
for imputed persons	933,000
for late enumerations	106,000
for RRC Incomplete Enumerations	423,000
UHAT = MHAT - X	1,384,000
OHAT: Overcoverage from COS	516,000
NHAT = UHAT – OHAT	869,000
C	31,613,000
C + NHAT	32,482,000
Source: Statistics Canada, 2006 Census Coverage Studies.	

4.3 Estimation

Table 4 gives the components of the estimated population coverage error for Canada. The estimate of population undercoverage, $\hat{U}$, is constructed from the results of the RRC and census data as follows Let $\hat{M}$ = estimate of the number of persons in the RRC target population determined to be in the 2006 Census target population who have not been enumerated. $\hat{M}$ is the sum of the final weights of persons classified as Missed. Let X = number of persons included in C that cannot be identified in the RRC as enumerated such as persons whose Census Day Address is a dwelling for which imputation was done during census processing. Then $\hat{U} = \hat{M} - X$. At the national level, X was about half of $\hat{M}$. This is a notable increase from 2001 when X, identified as the netted out factor in Table 5, was only about a third of $\hat{M}$. The increase is due to an increase in both the number of non-response dwellings and the number of misclassified dwellings that resulted in doubling the number of persons imputed.

5. Persons not enumerated

Each census count C is composed of two elements: $C = E + I$ where $E =$ the number of enumerations (This is the number of people who were listed on all census forms.[4]), and $I =$ the number of imputed persons. This is an estimate of the number of persons missed in non-response dwellings and/ or misclassified occupied dwellings. Undercoverage, therefore, is a subset of all persons who were not listed on a census form but should have been. It does not include those were not enumerated either because no completed census form was returned for the dwelling (non- response dwelling) or the dwelling did not receive a form because they were erroneously classified as unoccupied (misclassified occupied dwelling). An estimate of the true number of persons in the census target population T is given by $\hat{T} = C + \hat{N} = C + \hat{U} - \hat{O} = E + \left(I + \hat{U}\right) - \hat{O}$. This formulation of $\hat{T}$ has three components:

1. E = the number of number of persons who were listed on a census form;
2. $\hat{O}$ = an estimate of the number of excess enumerations[5]; and
3. $\left(I + \hat{U}\right)$ = an estimate of the number of persons who were not listed on a census form who should have been.

The last component, $\left(I + \hat{U}\right)$, estimates *the number of persons missed in the census for **any** reason*.[6] Let us define $\left(I + \hat{U}\right)$ as *census population collection undercoverage*, denoted by A. The estimate of census population collection undercoverage is $\hat{A} = \left(I + \hat{U}\right)$ and the corresponding estimate of the rate of census population collection undercoverage rate, $\hat{R}_A$ is: $\hat{R}_A = 100 * \frac{\hat{A}}{\hat{T}} = 100 * \left(\frac{I + \hat{U}}{C + \hat{N}}\right)$. Since net census collection undercoverage can be defined by subtracting overcoverage $\hat{O}$ from $\hat{A}$.,
$\hat{T} = C + \hat{N} = E + \left(I + \hat{U}\right) - \hat{O} = E + \hat{A} - \hat{O}$. Although net collection undercoverage cannot be applied to census counts to adjust for coverage error, $\hat{A}$ and $\hat{R}_A$ provides a broader picture of how well the census was able to enumerate its target population. Since there may be bias introduced whenever the data of one household is copied to represent the data of another household, as is done in WHI, collection undercoverage also provides analysts with a baseline and another measure of census data quality.

Collection undercoverage was introduced for the 2006 dissemination of estimates of census coverage error. It has been well received by those concerned with census operations as 'comprehensive' and 'clarifying' and by those using census data as 'quantifying a new element of data quality.' The national 2006 Census net collection undercoverage rate was estimated to be 7.13%. This means that the 2006 Census achieved an enumeration, consisting of being listed on a census form only once, for 92.87% of its target population. This compares to 95.74% of the target population included in the 2006 Census count of 31,612,897 persons. The difference between the two rates is that persons living in non- response dwellings and persons living in occupied dwellings erroneously classified as unoccupied are included in the former (100 - net collection undercoverage rate) but not the latter (100 - net undercoverage rate) . The 2001 Census achieved an enumeration for 94.17% of its target population compared to 97.57% of the target population included in the 2001 Census count of 30,007,094 persons. Since net undercoverage diminished from 2001

---

[4] It is possible that some of the persons listed on the form may not appear in the final census database. So, in the strictest sense, 'enumeration' is used here to represent persons listed on the form who appear in the final census database.

[5] Cases of overcoverage usually involve duplicate enumerations where one person is listed on firms for two different dwellings. There are a very small number of cases where someone is listed more than twice.

[6] The distinction is in the interpretation of the word 'missed'. When speaking of undercoverage, 'missed' specifically excludes those not listed on a form who should been who were imputed for.

to 2006, 2.99% to 2.67%, the decrease in the census net collection undercoverage rate is due to the increase in the number of non-response dwellings and misclassified occupied dwellings.

6. Conclusion

Census population coverage error increased from 2001 to 2006. Undercoverage, collection undercoverage and overcoverage all increased notably. There is evidence that some part of the overcoverage may be from census operations. The 2006 Census coverage studies implemented a number of important improvements resulting in measures of coverage error that are less biased and more precise. The new methodology for measuring overcoverage was successful while revealing a number of areas for improvement for 2011. Limiting the RRC to just the measurement of undercoverage reduced the need for collection and thereby the cost. Respondent and interviewer burden were reduced by focusing collection on cases more likely to have be classified Missed.

The methodology of the 2011 Census implements collection by waves, a new process for identifying unoccupied dwellings, and a new field management system. Online collection will be offered in more areas and the take up is expected to be larger. There is a risk of reducing net undercoverage in the 2011 Census, not from increasing overcoverage but from increasing non-response. Net undercoverage for one province, for example, in the 2006 Census was, at 0.8%, close to zero because of high non-response and misclassified occupied dwellings. Steps are being taken in the 2011 Census to address the misclassification of occupied dwellings as unoccupied. Work has already begun on the 2011 coverage studies. Major changes in methodology are not planned. One area of consideration is to include household characteristics in COS matching. In the RRC, there is potential to increase the use of administrative data to further reduce the need for collection.