

Evaluation of Reverse Record Check Estimates of Undercoverage in the Canadian Census of Population

R.D. BURGESS¹

ABSTRACT

Estimates of undercoverage in the Canadian Census of Population have been produced for each Census since 1961, using a Reverse Record Check method. The reliability of the estimates is important to how they are used to assess the quality of the Census data and to identify significant causes of coverage error. It is also critical to the development of methods and procedures to improve coverage for future Censuses. The purpose of this paper is to identify potential sources of error in the Reverse Record Check, which should be understood and addressed, where possible, in using this method to estimate coverage error.

KEY WORDS: Matching; Mobility; Nonresponse bias; Response error; Reverse record check; Sampling error; Tracing.

1. INTRODUCTION

The Census of Canada is conducted every five years; the most recent was in 1986. Starting with the 1971 Census, the main data collection methodology has been self-enumeration: less than 4% of the population are enumerated using the canvasser method. In geographic areas where self-enumeration is used, each dwelling is listed and a questionnaire dropped off by an enumerator just prior to Census Day (June 3 in 1981 and 1986). In larger urban areas the respondent household is asked to return the completed questionnaire by mail to the local supervisor of the enumeration. In rural areas and smaller urban areas the questionnaires are picked up by the enumerator.

The enumerator is to perform basic checks of coverage and response quality for his/her assignment and follow up on missing and incomplete questionnaires. Supervisory checks and quality control of the enumerator's work are also carried out. However, there is no independent and rigorous check of the listing of dwellings. Further, there is only limited opportunity to verify the number of persons listed on the questionnaire by the respondent household.

Not unexpectedly there are overcoverage and undercoverage errors in the Census. Such errors are important because of the various uses of Census data; representation in the Parliament of Canada is determined using Census population counts; various federal-provincial government financial agreements incorporate formulae that have population count or distribution as a factor (Statistics Canada 1983b). In turn the quality of estimates of coverage error is an important issue: for the use of Census data; in considering adjustment of population and dwelling counts to compensate for the coverage error; and in attempting to improve coverage quality for future Censuses by identifying significant causes or areas of coverage error.

Since 1961, Statistics Canada has produced and published an estimate of undercoverage for each Census of Population. The method used to produce these estimates has been a Reverse Record Check (RRC) study which involves five general activities or stages:

¹ R.D. Burgess, Social Survey Methods Division, Statistics Canada, 4th Floor, Jean Talon Building, Tunney's Pasture, Ottawa, Ontario, K1A 0T6.

- (i) frame preparation – identification of a set of nonoverlapping lists that together are to cover the total population that should be enumerated in the Census;
- (ii) sample design and selection – selection of a random sample of persons from the lists;
- (iii) tracing – determination of the address of usual place of residence on Census Day for each selected person (or verification that he/she died or emigrated prior to the Census);
- (iv) searching – review of Census returns to determine whether the selected person had been enumerated or missed in the Census; and
- (v) weighting and estimation – weighting up of sample results to produce an estimate of the number of persons missed in the Census.

A more detailed description of this methodology can be found in Gosselin 1976 or Statistics Canada 1984.

Other methodologies – post Census re-enumeration, demographic analysis and administrative record checks – could also be used to estimate Census undercoverage. In the Canadian context, however, each of these methodologies would likely produce results less reliable than those of the RRC. Re-enumeration studies show a tendency to miss the same households or persons as the Census itself. Demographic methods are model-based and suffer from a lack of reliable emigration estimates, measure only change in net coverage between censuses, do not identify individual cases and causes of coverage error, and are weakened sub-nationally by error in internal migration estimates. Administrative record checks are limited by the absence of a national administrative system that either has more complete coverage than the Census or has coverage errors independent of Census coverage error – a condition that would allow an incomplete administrative file to be used. Even if such a complete system existed, its use would be another version of a reverse record check, unless it were completely up to date in coverage and addresses, as of Census Day.

For these reasons the reverse record check has been the preferred methodology in Canada, though demographic analysis methods have been used for corroborative analysis. However, the RRC itself has deficiencies. The purpose of this paper is to describe some of the sources of error or limitations in the RRC method, in the context of the Canadian Census of Population. In Section 2 aspects of the survey methodology of the RRC that can lead to error in the final results are reviewed. The results of some analysis of RRC estimates, in conjunction with data from other sources, have raised unresolved problems related to the use of RRC results in population estimation. These results are presented in Section 3. Some concluding remarks are given in Section 4.

2. LIMITATIONS OF THE REVERSE RECORD CHECK METHODOLOGY

A limitation, in the context of this paper, is anything that restricts the applicability of the Reverse Record Check estimates or the confidence with which they can be used. Limitations can arise because of: differences between what is conceptually required by users and what the RRC attempts to measure; shortfalls in the design of the Reverse Record Check in attempting to meet its objectives; or sampling, response and other errors. Some of these limitations might be eliminated or reduced through modification of specific aspects of the Reverse Record Check. Others will persist or, by their nature, cannot be addressed.

2.1 Applicability of Reverse Record Check Estimates

The objective of the Reverse Record Check is to provide estimates, for each of the ten provinces, of undercoverage in the Census of Population. Net coverage error is not estimated and the Yukon and Northwest Territories are excluded from the study.

The RRC estimates the proportion of the population missed in the Census – *i.e.*, the proportion of the population that was not enumerated but should have been. Overcoverage (persons enumerated more than once, and persons enumerated who should not have been or were fictitious) is not estimated by the RRC. Thus net coverage error, undercoverage minus overcoverage, is not estimated by this vehicle. Even if the amount is small, the potential importance of overcoverage lies in its size and distribution relative to undercoverage. For example, overcoverage of 0.2%, one tenth the level of undercoverage in 1976 and 1981, would be very important if the rate for a particular province is as high as 0.5%.

The two Canadian territories have not been included in the RRC because the size of their populations is small but they have exceptionally high rates of intercensal in and out migration. In terms of sampling error, to produce reliable estimates for the territories, a proportionally large sample of the territorial population would have to be selected – of the order of a 5% sample or 3,750 persons. The territories have in and out intercensal migration rates of a third or more. Therefore, 1,250 of the 3,750 persons (on average) in the minimum sample should be intercensal in-migrants, assuming a proportional sample is required. The RRC uses lists for which the address of residence for the majority of persons was obtained five years earlier and in-migrants to the territories can only be identified during the conduct of the study. This in itself is not a problem. However, the RRC uses only a 0.15% sample. The in-migrants to the territories, therefore, would be expected to be sampled at this latter rate and not at the required 5% rate. This would result in a sample of in-migrants to the territories of only 30 persons. Thus, within the current framework of the RRC, and without prohibitive additional expense, it is not possible to select a meaningful sample to represent that third or more of the territorial population who are intercensal in-migrants.

2.2 The Reverse Record Check Methodology

Each of the five stages of the Reverse Record Check is a known or potential source of error.

2.2.1 Frame

The sample for the RRC is selected from four lists or frames:

- (i) Census: persons enumerated in the previous Census – for example, the 1981 Census was used for the 1986 Reverse Record Check;
- (ii) Birth: intercensal births, obtained from vital statistics records;
- (iii) Immigrant: intercensal immigrants, obtained from records of Employment and Immigration Canada; and
- (iv) Missed: persons missed in the previous Census – which is available as a sample only from the previous Reverse Record Check (no complete list exists for this group).

These lists are intended to include or represent, without duplication of individuals on or between lists, all persons who should be enumerated (in one of the ten provinces) in the current Census.

Some people, however, are not represented on these lists. Included among these are: (a) intercensal and never enumerated illegal aliens; (b) certain classifications of refugee; (c) certain

Canadians “abroad” at the time of the previous Census who returned prior to the current Census; (d) persons who move from the territories to one of the provinces in the intercensal period; and (e) persons not enumerated in any Census covered by the application of the RRC, but who were usual residents of Canada prior to 1961.

It is assumed, without direct evidence, that the number of persons in category (e) has become small enough to be irrelevant. For the 1981 Census the size of category (d) was estimated to be of the order of 18,000 persons. Most of these persons were usual residents of the territories at the time of the previous (1976) Census. There were probably also a few of what would be Birth frame and Immigrant frame persons among the 18,000.

Category (c) includes some Canadians working, studying or travelling abroad who did not maintain a usual place of residence in Canada during their absence and may also include children born outside Canada to parents in this category. It does not include persons in the Canadian military, in External Affairs or other government service (and their families) living abroad. They are included in the Census frame and the Missed frame. For the 1981 Census, the size of this returning “abroad” group was estimated to be approximately 67,000 persons.

Refugee applicants and illegal aliens in Canada are to be enumerated in the Census, assuming they do not have a usual place of residence outside of Canada, and are not holders of work or student visas. For the 1981 and 1986 RRC studies, persons applying from abroad and entering Canada as refugees were included in the Immigrant frame. Persons applying within Canada were included in the Immigrant frame only if they had been granted refugee status. As of April 1985, there were 12,500 applications from within Canada under consideration PLAUT 1985. The number of illegal aliens in Canada is not known or reliably estimated. Some illegal aliens may be represented in the Census frame or even the Missed frame. Amnesty programmes in the 1970’s and 80’s will have resulted in some illegal aliens being entered in the Immigrant frame.

Under the current RRC methodology the exclusions to the frames are important to the extent that such persons are not counted in the current Census. Since the Immigrant frame tends to have a high undercoverage rate (8.5% compared to 2.0% overall in 1981), it is not unreasonable to expect a high undercoverage rate for the refugee status claimants. It is possible that the majority of illegal aliens were not counted in the Census. These elements of undercoverage could be significant relative to the estimated number of persons missed (approximately 500,000 in 1981). The refugee status claimants and the illegal aliens may have been clustered in a few urban centres within only certain provinces. This would increase the impact of such exclusions on the reliability of estimates.

The lists can also be expected to include some amount of overcoverage; *e.g.*, persons enumerated in the previous Census who should not have been or who were enumerated more than once, fictitious persons and processing errors. Some overcoverage is detected during the course of the RRC operations. In estimating undercoverage, however, the effect of overcoverage in the frames would be consequential only if it approaches or exceeds the undercoverage in the Census in size.

2.2.2 Sample Size and Design

Error due to sampling is a major limitation of the RRC results. While the potential size of this error is dependent upon sample size and design, the sample size is the more important element. It, along with the available lists, limits the design options.

The basic 1981 and 1976 RRC undercoverage estimates for provinces and their corresponding estimates of standard error are presented in Table 1. The coefficients of variation (standard error divided by estimated undercoverage) varied from 4.5% at the Canada (10 provinces)

Table 1
 Estimated Population Undercoverage in the 1981 and 1976 Census,
 by Province, showing Provinces with Significant Differences in
 Population Undercoverage (with 95% confidence)

Province	Population Undercoverage		Province with a Significantly Different Undercoverage Rate
	Rate	S.E.	
	(%)	(%)	
1981 Census			
Canada (10 Provinces)	2.01	0.09	
1. Newfoundland	1.74	0.45	10
2. Prince Edward Island	1.17	0.54	9 and 10
3. Nova Scotia	1.05	0.34	5, 6, 9 and 10
4. New Brunswick	1.81	0.30	10
5. Québec	1.91	0.21	3, 7, 8 and 10
6. Ontario	1.94	0.14	3, 7, 8 and 10
7. Manitoba	0.98	0.35	5, 6, 9 and 10
8. Saskatchewan	0.99	0.37	5, 6, 9 and 10
9. Alberta	2.54	0.36	2, 3, 7 and 8
10. British Columbia	3.16	0.33	all but 9
1976 Census			
Canada (10 Provinces)	2.04	0.10	
1. Newfoundland	1.10	0.39	5 and 10
2. Prince Edward Island	0.38	0.25	4, 5, 6, 8, 9 and 10
3. Nova Scotia	0.86	0.34	4, 5 and 10
4. New Brunswick	2.16	0.37	2, 3, 7 and 10
5. Québec	2.95	0.25	1, 2, 3, 6, 7, 8 and 9
6. Ontario	1.52	0.17	1, 2, 5, 7 and 10
7. Manitoba	1.07	0.33	4, 5 and 10
8. Saskatchewan	1.33	0.34	2, 5 and 10
9. Alberta	1.49	0.26	2, 5 and 10
10. British Columbia	3.13	0.31	all but 5

level, up to 13.6% at the regional (Atlantic, Québec, Ontario, Prairie and British Columbia) level and up to 46% at the provincial level. Sub-provincial coefficients of variation were typically higher. For an Electoral District of average size (86,323 persons in 1981) with an estimated 2% undercoverage, the coefficient of variation would be approximately 50%. For smaller geographic areas and small population groups the coefficient of variation could be much higher.

The sampling error, of course, has an effect on attempts to differentiate among provincial, and among other undercoverage rates. In turn this affects attempts to identify specific causes or areas of undercoverage, and undermines the validity of adjusting for coverage error as a means to improve Census counts. Those provinces with a significantly different undercoverage rate are also shown in Table 1. The undercoverage rates for the provinces appear to fall into six groupings, for 1981, based on both rate of undercoverage and provinces with which the rate is significantly different. For 1976, with eight groups, there was less similarity between provinces. No group in either Census, however, can be shown to be completely different from all others, and may not be.

This general situation is not dissimilar to that for applications of the Reverse Record Check for the 1966 and 1971 Censuses. From 1966 onward only the province of British Columbia has had an undercoverage rate significantly above the Canada level. The variation from Census to Census for most provinces, in large part, could be due to sampling error. Why it is not for British Columbia is a major concern for both the Reverse Record Check and the Census.

The need to use a sample of “missed” persons from the previous RRC also places a limitation on the design and sample size. There is no direct control of the size of this segment of the sample. Any limitations of the previous Reverse Record Check, to the degree that these were reflected in the estimate of “missed” persons, will be passed. (See Sections 2.2.4, 2.2.5 and 3).

2.2.3 Tracing

Given the nature of the lists or frames used for sample selection, addresses and other information may be up to five years out of date. Attempts are made to update addresses prior to Census Day using administrative files. (This was first carried out extensively for the 1986 RRC.) After Census Day, the Census questionnaire corresponding to the original address, or the update if available, is searched as a first attempt to determine whether the selected person was enumerated in the Census. Every selected person not found enumerated in the first search must be traced. The selected person, or a reliable source, must be contacted either to obtain an updated or confirmed address, or to determine the selected person’s status, *i.e.*, as deceased, emigrated, abroad.

Despite extensive tracing activities, not all selected persons can be traced. This may result in a form of nonresponse bias. In the 1981 RRC 3.4% of all selected persons were not traced. With overall undercoverage in the Census estimated to be 2.0% this “not traced” rate represents an important uncertainty in the RRC estimates.

A weight adjustment is carried out to account for these “not traced” cases. The effect of the weight adjustment for the 1981 Census was to impute an undercoverage rate of 3.27% for the “not traced” cases from the Census and Missed frames (jointly), 1.46% for the Birth frame and 11.94% for the Immigrant frame. Overall, the proportion of “not traced” weights “imputed” by the weight adjustment to “missed” was 1.6 times the initial (weighted) proportion represented by the “missed” cases among all traced selected persons. This suggests a relationship between “not traced” and “missed”. It is not known, of course, if the 1.6 rate was too high, too low or correct. To the extent that it is not correct, there may be some distortion in provincial estimates of undercoverage as well as a bias in overall estimates of undercoverage.

Since the rates of intercensal interprovincial in and out-migration vary from one province to another, there may be some distortion among provincial estimates. This will occur if the proportion of interprovincial movers within weighting groups is not the same among the cases traced and not traced.

Intercensal interprovincial movers (applicable for Census and Missed frames only) have a high undercoverage rate. This rate was estimated to be 6.13% for the 1981 Census, based upon mobility data from the 1981 RRC derived by comparing of the 1976 Census and 1981 Census addresses. The estimated undercoverage rate for intercensal migrants within a province (*i.e.*, between Census Subdivision (CSD) or municipality movers) was 3.83%. For intercensal non-migrant movers (within CSD or municipality) the undercoverage rate was estimated to be 2.83%. Given these rates and the distribution of mobility characteristics, the “imputed” undercoverage rate for the “not traced” cases from the Census and Missed frames put together would be expected to be at least 3.52% rather than the actual 3.27%. That is, given persons not traced almost always have moved. It is, in turn, assumed that these “not traced” cases included

proportionally at least as many migrants, within and between provinces, and had not less than the same undercoverage rates, by mobility status, as traced cases. (The distribution of mobility status of the enumerated population 5 years and older, estimated through the 1981 RRC was approximately: (i) Non-movers – 55%; (ii) Non-migrant Movers – 17%; (iii) Migrants Same Province – 21.7%; (iv) Migrants Different Province – 5%; and (v) Migrants From Outside Canada – 2%.)

Given the tracing methods used, it is not unreasonable to speculate that the proportion of migrants, and thus the undercoverage rate, was much higher for the “not traced” cases. If they were, then there could be a significant downward bias in the estimates of undercoverage. For example, if the “true” undercoverage rate among the cases not traced was close to 5.0%, then the bias in the undercoverage estimate at the Canada (10 provinces) level would exceed the sampling error.

2.2.4 Searching and Classification

After all tracing attempts have been made and any interviews conducted, each selected person is classified to one of six categories:

- (1) enumerated;
- (2) missed;
- (3) deceased;
- (4) emigrated or abroad;
- (5) overcoverage in a list or frame; and
- (6) not traced.

As outlined above, to determine whether a selected person has been enumerated or missed the Census questionnaire corresponding to the selected person’s address must be searched. For the search to result in the correct classification of the selected person, it is necessary that the address being searched be the correct address, and that the selected person be correctly identified on the Census questionnaire and in RRC documentation; *i.e.*, that there be no response error or nonresponse for the relevant items.

If the selected person is correctly identified (complete name, correct age and sex, *etc.*.) and there are no processing errors, then no selected person who was missed in the Census will be classified as “enumerated”. The converse is not true. If a selected person has been enumerated in the Census at some address other than that which is obtained from the list of selection, some other administrative source or a directory, then to be classified as “enumerated” that address must be provided by the selected person or some other contact. If the selected person does not or can not provide that address (for example, recall error or can not remember), then he or she will be classified as “missed” or “not traced”. Generally, when the selected person (or a parent, spouse or other reliable source) gives an address, or set of addresses, where he/she should have been or may have been enumerated, this address information is accepted as correct. Selected persons will be classified as “enumerated” or “missed” based on this address information. It is not known how accurate such address information actually is for persons classified as “missed”.

On the other hand there may be a higher probability of classifying a person missed as “not traced” than a person enumerated in the Census. Before a person can be classified as missed he/she (or a reliable source) must be interviewed to confirm the address and to obtain possible alternative addresses and certain Census data for him/her and the household. This procedure will eliminate some classification error. At the same time, if the information about a person missed is doubted this can only be resolved through the contact with him/her (or a parent,

spouse, *etc.*). If the doubt is not resolved the case will be classified as “not traced”. Conclusive information is not always necessary for a person who was enumerated. With exhaustive searching it may be possible to transform a selected person, who was enumerated, from “not traced” to “enumerated”, even if the address obtained is incomplete or incorrect. Such searching is much less likely to alter the outcome for persons missed in the Census.

The selected person is not always adequately identified. In accepting a selected person as matched; *i.e.*, found enumerated on a Census questionnaire – name is not always identical on the Census and RRC documents. Sometimes only the first person listed on a Census questionnaire has a complete name and in a few cases no names are given. If the identity of the selected person cannot be determined from the list or frame, then the case will be classified as “not traced” at the outset. Included among these will be persons “assigned” for absent households and refusals in the previous Census. Date of birth and other data are not always present, complete or found identical in matching. For the majority of cases the quality of matching is unquestioned, but a minority of cases raise doubts. Doubtful cases accepted as matched potentially are misclassified as “enumerated”. Those rejected as matched potentially are misclassified as “missed”, though most will be classified as “not traced”. Different rules for acceptance/rejection as matched, of course, may yield different estimates of undercoverage.

Some overcoverage in the frames can be detected. This will include: some foreign residents enumerated in the previous Census; persons “created” by processing error in the previous Census; immigrants who have not yet resided in Canada; births in Canada to non-resident parents; and fictitious or out of scope “persons” listed on the questionnaire from the previous Census. In 1981 these cases represented less than 0.1% of selected persons.

Overcoverage in the form of duplication in a frame will not be detected. Fictitious selected persons may go undetected and be classified among the “not traced” cases.

The final classifications of the selected persons from the 1981 RRC are presented in Table 2 (from Burgess 1986).

2.2.5 Weighting and Estimation

At the time of sample selection, a basic weight equal to the inverse of the sampling fraction is assigned to each selected person record. Two types of weight adjustment are made to this basic weight – one to account for “not traced” cases, the other to account for deviations in

Table 2
1981 RRC Final Classification of Selected Persons

Final Classification	Frame									
	Census		Birth		Immigrant		Missed		Total	
	Cases	%	Cases	%	Cases	%	Cases	%	Cases	%
Traced	29,761	97.1	3,211	92.3	1,392	96.1	807	96.1	35,171	96.6
Enumerated	27,541	89.8	3,096	89.0	1,113	76.8	696	82.9	32,446	89.1
Deceased	1,056	3.5	33	0.9	5	0.3	26	3.1	1,120	3.1
Emigrated/Abroad	299	1.0	34	1.0	111	7.7	24	2.8	468	1.3
Missed	865	2.8	48	1.4	163	11.3	61	7.3	1,137	3.1
Not Traced (incl. Overcoverage)	895	2.9	267	7.7	57	3.9	33	3.9	1,252	3.4
TOTAL	30,656	100.0	3,478	100.0	1,449	100.0	840	100.0	36,423	100.0

the representativeness of the sample, after elimination of "not traced" cases, relative to the lists of selection.

A "not traced" case represents a person enumerated or missed in the Census, a deceased person, an emigrant, a person abroad or overcoverage. The weights of the "not traced" cases, therefore, are redistributed among the "traced" cases. The adjustment is carried out within groups defined by various demographic and geographic characteristics, and frame.

The weight adjustment for the "not traced" cases is carried out in two stages. First, an adjustment is made for those cases for which no tracing was undertaken because there was inadequate information for matching and tracing. These cases are weighted into all other selected persons. Second, an adjustment is made for all other "not traced" cases. These are weighted into specific groups of the remaining selected persons. How the "not traced" adjustment is carried out is restricted by the information available on the "not traced" selected persons. Ideally, how a selected person was traced and whether he/she had moved and how far, as well as demographic characteristics, should be taken into consideration in defining weighting groups. To date only demographic characteristics and minimal mobility data have been used in the weight adjustment. (Persons selected in the Census frame who have not moved in the intercensal period and who were classified as "enumerated" are excluded from this weight adjustment.) By their nature it is difficult to categorize most "not traced" cases beyond the fact that they were not found enumerated at the address given on the list of selection.

For the second type of adjustment, totals for relevant sub-groups of the population are obtained from each frame (except for the Missed frame for which only a sample is available). Using these "known totals", an adjustment to the RRC weights is made within the corresponding subgroups of the sample. This is done to reduce the error in the estimates by ensuring that totals from the sample, for basic population characteristics for which undercoverage rates are published, correspond to the totals in the frames.

Neither adjustment deals at all with the various exclusions to the lists used for sample selection. In the calculation of any proportion of persons missed in the Census the published Census count of enumerated persons is used in the denominator in order to minimize sampling error. (The covariance of the estimate of "enumerated" persons and the estimate of "missed" persons tends to be negative.) Since the RRC does not represent all elements of the true population, the effect of using the Census count is to assume that the undercoverage rate for the exclusions is zero.

The estimator, which takes the general form defined as:

$$\text{Estimated proportion of persons missed} = \frac{\text{Estimated no. of missed persons}}{\text{no. of persons counted in the Census} + \text{Estimated no. of missed persons}}$$

is discussed further in Appendix 2.

2.3 Reducing Potential for Error and Methodological Limitations

Experimental work and evaluation of methods in the RRC may make it possible to eliminate or reduce the impact of some sources of error or limitations.

Overcoverage might be estimated by means of an independent study. Such a study is being conducted, on an experimental basis, for the 1986 Census. However, the cost to produce estimates of adequate quality at the province level may be very high.

The production of estimates for the Yukon and Northwest Territories requires a set of lists other than those used for the RRC. Such a set would have to be current and have no significant duplication that could not be removed or estimated. With such a set of lists, the basic

RRC methods could be applied. Some experimental work in this regard has been done and more is planned.

The lists used for the RRC could be augmented to eliminate some of the exclusions, for example, refugee status claimants and migrants from the territories to the provinces. These people, however, will be difficult to trace. Sampling these groups may do little more than change the nature of the problem.

A sample of "abroad" persons could be obtained by using the previous Reverse Record Check. Such a sample, however, would be very small, would not represent the entire group in question and the selected persons would be difficult to trace.

Other than illegal aliens the "never enumerated" group will become smaller and smaller over time. Intercensal illegal aliens, and other illegal aliens never enumerated in Canada, will remain excluded.

The impact of sampling error can be reduced by increasing the sample size. The question is to what size, at what cost, based upon what criteria? An increase in the RRC sample from its current 36,500 persons to 100,000 should be sufficient to bring the provincial standard error estimates, for the undercoverage rates, down below 0.2%. However, this may not be sufficient for purposes of adjusting the Census counts, depending upon the level and distribution of undercoverage estimates actually obtained. A reduction of the standard error to 0.1% for each province – the level yielded by the 1981 and 1976 RRC studies for the Canada (10 province) level estimate of undercoverage of 2% – would require a sample for Canada of approximately 350,000 persons, assuming the 1981 provincial levels of undercoverage, type of sample design and design effects. To conduct a high quality RRC operation for such a large sample, given the controls and quality checks required, would be much more costly than the mere increase in sample size suggests, and might be operationally unrealizable. Increasing the sample size, of course, would not reduce any bias in the estimates.

Tracing methods are examined before and after each RRC. Major changes were made for 1986 and changes and improvements are being contemplated for 1991. It must be expected, however, that there will again be a non-negligible percentage of "not traced" cases. These cases will continue to be dealt with by weighting or by imputation and weighting.

Evaluative studies can be conducted to assess the quality of matching and of address information provided by respondents or reliable sources. The potential impact of the matching algorithm or criteria can also be assessed to some extent. However, even if such studies identify a problem, solutions may not be readily forthcoming.

Modifications to the weighting procedures can be tested in an attempt to better deal with mobility and other characteristics when adjusting for "not traced" cases (Burgess 1986). Additional information for this purpose might be available from administrative sources. Some minor refinements using existing information can also be made. For example, the adjustment for "not traced" persons contacted, but from whom the necessary Census Day address information could not be obtained, might be different from that for "not traced" persons who potentially may be "deceased", "emigrated" or "abroad".

Adjustments using current Census totals of enumerated persons could be tested as well. For this to reduce any bias associated with "not traced" cases and persons not represented in the RRC sample, however, the basic classification of cases to "missed" must be without bias and there must be no interprovincial distortion of the proportion "missed". These types of modifications to the weighting would not in themselves eliminate bias.

3. ANALYSIS OF REVERSE RECORD CHECK RESULTS

The RRC not only provides estimates of the number of persons missed in the Census, but also independent estimates of the number of persons enumerated in the Census, and the number of intercensal deaths, emigration and persons who have moved abroad but who have not emigrated. These estimates are used in validating RRC estimates. Some of the results of this validation process serve to illustrate limitations discussed in Section 2.

Analysis has also been carried out to correlate geographic variation in undercoverage to variation in the distribution of Census population and household characteristics.

3.1 Independent Estimates

The Reverse Recoded Check estimates of persons enumerated in the Census, of intercensal deaths, and of persons leaving Canada in the intercensal period can be compared to estimates from other appropriately chosen sources – for example, estimates of enumerated persons to Census counts and estimates of deaths to Vital Statistics data. If there are no significant biases in the RRC estimates, then any differences between these estimates will usually be explainable by the corresponding sampling error of the RRC estimate. If there are significant differences, then these might be due to biases in the RRC estimates. The overall quality of these estimates, revealed by the comparisons, likely will be a reflection of the quality of the estimates of “missed” persons.

RRC estimates of emigrants (296,727) and of persons “abroad” (57,909) compared favourably with estimates based upon demographic analysis. The RRC estimate for emigrants, for example, is in the mid range of the five demographic analysis values examined – ranging from 197,000 to 372,000, with a mean value of 266,400. The RRC estimate of deceased persons (846,378) is very close to the value (840,689) published by Statistics Canada 1976 to 1981.

Comparisons of estimates for enumerated persons do indicate some problems. Some of these comparisons are presented in Table 3. For Canada (10 provinces) and for two of the ten provinces, the number of persons enumerated in the Census, as estimated by the RRC, is significantly different from the published Census count. The discrepancy of 209,911 at the aggregate level can be explained in part by exclusions from the lists or frames of the RRC. The discrepancies among provinces is difficult to explain. That in particular makes the discrepancy important. The 209,911 aggregate discrepancy must be considered in the context of the RRC estimate of 497,277 persons missed in the Census; similarly, the discrepancy for British Columbia of 80,304 in the context of an estimated 89,445 persons missed and the discrepancy for Alberta of 86,244 persons in the context of an estimated 58,335 persons missed.

An estimated 67,000 non-immigrants who had been “abroad” at the time of the previous Census arrived in Canada legally, and an estimated 18,000 persons moved from the territories to a province in the intercensal period. Assuming none of these people was missed in the Census, the discrepancy would be reduced to approximately 125,000 persons. This difference would remain at the outer limits of what would be reasonably accepted as due to sampling error only. Further, all of these 85,000 (67,000 + 18,000) persons would have had to have moved to Alberta and British Columbia to reduce the discrepancies for these provinces to within 95% confidence intervals – a clearly unreasonable supposition.

The remainder of the difference (125,000) could be made up of various (potential) errors in the RRC or the Census: (i) sampling error in the RRC estimate of enumerated persons; (ii) an increase in overcoverage in the 1981 Census – compared with the 1976 Census; (iii) RRC exclusion of illegal aliens and refugee claimants enumerated in the Census itself; (iv) underestimation of persons missed in the 1976 Census – these persons make up 1981 Missed

Table 3
Reverse Record Check Estimates of the Number of Persons
Enumerated in the 1981 Census by Province

Province	RRC Estimate of Persons Enumerated	S.E. of RRC Estimate	Census Published ¹ Count	Persons Enumerated RRC-Census	RRC Estimate of Persons Missed
Canada (10 provinces)	24,064,376	62,193	24,274,287	-209,911 ²	497,277
Newfoundland	568,696	8,256	567,681	1,015	10,039
Prince Edward Island	116,012	3,005	122,506	-6,494	1,456
Nova Scotia	837,045	11,185	847,442	-10,397	9,034
New Brunswick	685,332	8,167	696,403	-11,071	12,864
Québec	6,410,662	38,648	6,438,403	-27,736	125,180
Ontario	8,629,374	52,802	8,625,107	4,267	171,010
Manitoba	1,028,162	15,133	1,026,241	1,921	10,203
Saskatchewan	973,450	11,740	968,313	5,137	9,712
Alberta	2,151,480	24,238	2,237,724	-86,244 ²	58,335
British Columbia	2,664,163	19,798	2,744,467	-80,304 ²	89,445

¹ Statistics Canada 1982.

² Greater than 3 standard errors.

frame; and/or (v) over-estimation of persons missed in the 1981 Census. The extent to which each of these sources might have contributed to the difference is not known. The fact that a large part of the difference seems to be associated with British Columbia and Alberta is perhaps in some degree due to under-estimation of intercensal migrants. Migration to these provinces was particularly high between 1976 and 1981 (Statistics Canada 1979; 1983a).

There may also be some bias in the estimates of emigrated, abroad and/or deceased persons. If these are over-estimated for reason other than "not traced" bias, there should also be a tendency to under-estimate the persons missed, since the last address in Canada is sought and used in searching. Persons who emigrated, died or went abroad after Census Day may have been reported as such at the time of tracing, perhaps several months after Census Day. At the same time, the fact that deceased persons do not appear to have been under-estimated despite the exclusions to the RRC frames suggests a lower mortality rate for the exclusions (as is the case for immigrants – see Table 2) than for the entire population and/or over-estimation of this group.

The data in Table 4 show that intercensal migrants were under-estimated for all provinces except Saskatchewan. This may be in part associated with the "not traced" cases. The under-estimation for British Columbia may explain the discrepancy for this province shown in Table 3. On the other hand, the under-estimation for Alberta does not adequately explain the discrepancy for that province and, thus one or more of the factors (i) to (v) noted above must be contributing to this discrepancy.

Under-estimation of migrants may cause a distortion of undercoverage estimates among the provinces; *i.e.*, the large differences shown in Table 4 by province might be indicative of substantial biases in provincial under-enumeration rates. Further, as noted in Section 2.2.3, migrants have higher than average levels of undercoverage. If the enumerated persons within this group are under-estimated, while in general non-migrants are not under-estimated, relative to the Census, then estimates of undercoverage may be too low.

Table 4
Reverse Record Check Estimates of Migrants¹ Enumerated
in the 1981 Census, by Province

Province	Estimate of Migrants			Census Estimate of Inter-Provincial Migration		
	RRC	Census published estimate ²	Difference RRC-Census	In	Out	Out/In
Canada	4,670,311	5,046,500	-376,239	1,124,970	1,122,370	-
Newfoundland	61,499	72,100	-10,601	18,430	38,265	2.08
Prince Edward Island	13,257	20,530	- 7,273	9,945	9,950	1.00
Nova Scotia	125,949	137,865	-11,916	54,455	62,880	1.16
New Brunswick	96,607	109,955	-13,348	41,460	49,965	1.21
Québec	1,092,919	1,145,085	-52,166	61,310	203,035	3.31
Ontario	1,572,504	1,725,225	-152,721	250,570	328,640	1.31
Manitoba	143,391	165,105	-21,714	54,030	97,620	1.81
Saskatchewan	204,937	192,840	12,097	63,395	69,220	1.09
Alberta	669,995	691,970	-21,975	336,830	139,180	0.41
British Columbia	689,253	785,825	-96,622	234,545	123,615	0.53

¹ A migrant is a person who at the time of the previous Census was living outside Canada, in a different province or in a different municipality (or CSD). RRC mobility data used here are those given by the RRC sample person in the Census and not those derived within the RRC (based upon a comparison of addresses).

² Statistics Canada 1983a.

Discrepancies between the RRC estimate of enumerated persons and the Census count have also occurred for earlier Censuses. The value of the RRC estimate minus the Census count was 289,000 for 1971, and -324,000 for 1976. For both of these Censuses, the RRC estimates of persons deceased and emigrated/abroad were consistent with other sources. The large change from 1971 to 1976, coincident with the large negative values for two consecutive Censuses, cannot emanate from a single source. Changes in the size of overcoverage, larger than the size of the discrepancies, would be required between Censuses. This by itself, however, would not be consistent with the results of demographic analysis for these three Censuses (Statistics Canada 1987).

Remaining consistent with the demographic estimates, the differences would be explained in part by the presence of a large downward bias in the 1971 RRC estimate of persons missed. The 1971 unbiased estimate would have to be of the order of 3.8% rather than the estimated 1.9%. This would have to be accompanied by a not as large decrease in overcoverage between 1966 and 1971 followed by an increase in overcoverage for 1976 and a decrease for 1981. There would have to be also some under-estimation of missed persons for 1976.

Such a scenario is speculative, however, and no reason was found for such changes occurring. Other scenarios may also be possible. The occurrence of the discrepancies, however, does raise questions about the reliability of the RRC estimates and the potential effect of overcoverage on net coverage error.

The provincial distribution of the discrepancy between the RRC estimate of enumerated persons and the Census count differ among Censuses, further confounding its effects and potential sources. These results for the 1976 Census are given in Table 5.

Table 5
Difference Between Reverse Record Check Estimates of Persons
Enumerated and the 1976 Census Counts

Province	Difference in Population Enumerated (RRC-1976 Census)	Percent Difference
Canada (10 provinces)	-323,500	-1.4
Newfoundland	21,900	3.9
Prince Edward Island	-500	-0.4
Nova Scotia	- 4,500	-0.5
New Brunswick	-15,000	-2.3
Québec	-56,200	-0.9
Ontario	-207,000	-2.5
Manitoba	- 6,600	-0.6
Saskatchewan	1,400	0.1
Alberta	-43,400	-2.4
British Columbia	-12,800	-0.5

3.2 Variation in Geographic Distributions

The RRC estimates of undercoverage can be used as general indicators of the coverage quality of the Census. They are also intended to be used to direct the development and testing of coverage improvement procedures for future Censuses. Under ideal circumstances, they would be used to model undercoverage to produce estimates for small areas and as part of a coverage adjustment "correction" procedure. For these uses, geographic variation in coverage quality, indicated by the RRC results, is of particular concern. Variation in Census data distributions have been examined to determine whether they are correlated to the apparent variation in undercoverage among provinces. To date these investigations have not yielded satisfactory models or explanations.

A lack of success modelling undercoverage or explaining the variation between provinces may be due to, or confounded by: (i) bias and/or sampling error in the RRC estimates; (ii) undercoverage not strongly correlated to the Census characteristics of individuals, households and/or families; (iii) undercoverage correlated to a perhaps complex combination of Census and other characteristics; and/or (iv) a multitude of sources of undercoverage that must be considered separately; for example, undercoverage of individuals considered separately from undercoverage of entire households.

4. CONCLUSION

The RRC is thought to be the best vehicle developed to date for estimation of undercoverage in the Census in Canada. Its estimates provide basic measures to monitor and assess the quality of Census counts.

There are conceptual, theoretical and practical limitations to the RRC Check method as currently applied to the Canadian Census. The frames or lists used, while covering the large

majority of the population to be enumerated, are not comprehensive. Specific geographic areas are excluded as are certain segments of the population. The sample size is limited, but not necessarily to its present size, by constraints of tracing and matching, and by the demands for accuracy in operations. The "not traced" cases are a source of bias. The proportion of cases not traced, relative to the proportion of "missed" cases, in particular, adds an important uncertainty to the estimates, as does the inconsistency of RRC estimates of enumerated persons with corresponding Census counts.

In some instances the degree or impact of error, or limitations, could be evaluated in greater depth. Modifications and alternative procedures or methods that have a reasonable likelihood of improving the quality and applicability of the estimates can be applied. Potentially, alternatives can be developed. Such changes, however, would have varying costs and degrees of effectiveness associated with them. Also, it remains to be shown whether such changes would do more than enhance the status of the RRC estimates as general indicators of coverage quality in the Census.

ACKNOWLEDGEMENTS

The author would like to thank Gordon Brackstone, Geoff Hole, Judy Clarke and staff of Social Survey Methods for assistance and comments during preparation of the paper. The comments of the referees and editors were helpful and appreciated.

Appendix 1

Further Results From the Reverse Record Check

Results from the 1986 Census Reverse Record Check have been published (Statistics Canada 1988). The following extract displays the undercoverage rates for the 1981 and 1986 Censuses for demographic characteristics. Analysis of the 1986 undercoverage estimates by province, age, sex, marital status, mother tongue and other groupings is continuing.

1981 and 1986 Reverse Record Check Undercoverage Rates for Selected
Population Characteristics - 10 Provinces

Characteristic	1981 Estimated Population Undercoverage		1986 Estimated Population Undercoverage	
	Rate	S.E.	Rate	S.E.
	%	%	%	%
Sex				
Male	2.37	0.13	3.91	0.16
Female	1.65	0.12	2.87	0.16
Age Group				
0- 4	1.21	0.22	2.28	0.48
5-14	1.23	0.21	2.12	0.26
15-19	2.96	0.52	3.89	0.60
20-24	5.51	0.29	9.06	0.45
25-34	2.31	0.28	4.76	0.32
35-44	2.20	0.26	2.40	0.32
45-54	0.81	0.23	1.77	0.28
55-64	0.91	0.29	2.09	0.31
Marital Status				
Married/Separated	1.22	0.11	1.89	0.15
Divorced	5.10	1.03	7.07	1.07
Widowed	0.64	0.39	2.68	0.51
Single/Never Married	2.86	0.16	4.91	0.21
Mother Tongue				
English	1.86	0.11	3.12	0.13
French	1.80	0.20	3.10	0.33
Other	3.08	0.26	-	-
Urban/Rural Population Size Group				
Urban Areas	2.08	0.11	3.28	0.13
500,000 & over	2.29	0.17	3.58	0.15
100,000 to 499,999	1.86	0.31	2.94	0.33
Less than 100,000	1.80	0.23	-	-
Rural Areas	1.79	0.21	3.73	0.29

Appendix 2

Equations Used to Assess RRC Estimates and Estimator

The 1981 Reverse Record Check estimates have been assessed and discussed based upon four equations. The first simply defines the RRC population or frames. The second redefines the RRC sample in terms of the outcome or estimates of the study. The third defines the population enumerated in the Census in terms of the RRC estimate of enumerated persons. The fourth defines the error components for the estimate of missed persons.

Equation 1:

The RRC population size = $C_{76} + M_{76} - e(\hat{M}_{76}) + I_{76/81} + B_{76/81}$,

where

- C_{76} = number of persons counted, or enumerated, in one of the ten provinces in the 1976 Census,
- M_{76} = number of persons missed in one of the ten provinces in the 1976 Census,
- $e(\hat{M}_{76})$ = error (under or (-) over estimation of persons) associated with M_{76} , the Missed frame sample; *i.e.*, $M_{76} = \hat{M}_{76} + e(\hat{M}_{76})$,
- $I_{76/81}$ = number of registered 1976 to 1981 intercensal immigrants to one of the ten provinces,
- $B_{76/81}$ = number of registered 1976 to 1981 intercensal births in one of the ten provinces.

Equation 2:

The RRC estimates = $\hat{C}_{81} + \hat{C}_{fT81} + \hat{M}_{81} + \hat{M}_{fT81} + \hat{L}_{76/81} + \hat{A}_{81} + \hat{D}_{76/81} + \hat{O}_{f81}$

where

- \hat{C}_{81} = estimated number of persons in an RRC frame who were enumerated in one of the ten provinces in the 1981 Census,
- \hat{C}_{fT81} = estimated number of persons in an RRC frame who were enumerated in one of two territories in the 1981 Census,
- \hat{M}_{81} = estimated number of persons in an RRC frame who were missed in one of the ten provinces in the 1981 Census,
- \hat{M}_{fT81} = estimated number of persons in an RRC frame who were missed in one of the two territories in the 1981 Census,
- $\hat{L}_{76/81}$ = estimated number of persons in an RRC frame who were 1976 to 1981 intercensal emigrants,
- \hat{A}_{81} = estimated number of persons in an RRC frame who were abroad and had no usual place of residence in Canada at the time of the 1981 Census,
- $\hat{D}_{76/81}$ = estimated number of persons in an RRC frame who died in the 1976 to 1981 intercensal period,
- \hat{O}_{f81} = estimated overcoverage (number of "persons") in the Census, Birth and Immigrant frames which was detectable in the 1981 RRC operations.

Equation 3:

The estimate \hat{C}_{81} should = $C_{81} - \hat{C}_{81}[e(\hat{M}_{76})] - \hat{C}_{c/re81} - R_{81} - T_{76/81} - S_{76/81} + M_{n81} - O_{81} + \hat{C}(\hat{O}_{nf81})$,

where

- C_{81} = number of persons enumerated in one of the ten provinces in the 1981 Census,
 $\hat{C}_{81}[e(\hat{M}_{76})]$ = that component of $e(\hat{M}_{76})$ not or (-) over represented in \hat{C}_{81} ,
 $\hat{C}_{c/re81}$ = under or (-) over-estimation of “enumerated” persons in an RRC frame because of classification, response, sampling and “no trace” error in the 1981 RRC,
 R_{81} = number of persons abroad at the time of the 1976 Census who were in Canada at the time of the 1981 Census,
 $T_{76/81}$ = number of intercensal migrants from the two territories to a province,
 $S_{76/81}$ = net number of intercensal entries to the ten provinces, as of Census Day, not in an RRC frame and not accounted for above (e.g., illegal aliens),
 M_{n81} = number of persons not in a RRC frame who were missed in one of the ten provinces in the 1981 Census,
 O_{81} = overcoverage in the ten provinces in the 1981 Census,
 $\hat{C}(\hat{O}_{nf81})$ = estimated overcoverage (number of “persons”) in the Census, Birth and Immigrant frames which was *not detected* in the 1981 RRC operations and is represented in \hat{C}_{81} .

Thus,

$$\hat{C}_{81} - C_{81} = -\hat{C}_{81}[e(\hat{M}_{76})] - \hat{C}_{c/re81} - S_{76/81} + M_{n81} - O_{81} + \hat{C}(O_{nf81}) - R_{76/81} - T_{76/81},$$

assuming no error in \hat{O}_{nf81} .

Equation 4:

$$M_{81} - \hat{M}_{81} = \hat{M}_{81}[e(\hat{M}_{76})] - \hat{M}_{81}(\hat{O}_{nf81}) + \hat{M}_{c/re81} + M_{n81} = e(\hat{M}_{81}).$$

where

- $\hat{M}_{c/re81}$ = under or (-) over-estimation of “missed” persons in an RRC frame because of classification, response, sampling and “no trace” error in the 1981 RRC,
 $\hat{M}_{81}[e(\hat{M}_{76})]$ = that component of $e(\hat{M}_{76})$, represented in \hat{M}_{81} ,
 $\hat{M}_{81}(\hat{O}_{nf81})$ = estimated overcoverage (number of “persons”) in the Census, Birth and Immigrant frames which was not detected in the 1981 RRC operations and is represented in \hat{M}_{81} .

Note: There is a classification, response, sampling and “no trace” error component associated with each item of equation 2; e.g., $\hat{C}_{c/re81}$ and $\hat{M}_{c/re81}$. These taken in total sum to zero. In the above equations these error components exclude error caused by overcoverage and overcoverage which results in a “not traced”; e.g., non-existent persons enumerated in the previous Census. The effect of overcoverage is included, for example, in $\hat{C}(\hat{O}_{nf81})$ and $\hat{M}(\hat{O}_{nf81})$.

Similarly,

$$e(\hat{M}_{76}) = \hat{M}_{76}[e(\hat{M}_{71})] - \hat{M}_{76}(\hat{O}_{nf76}) + \hat{M}_{c/re76} + M_{n76}.$$

Error and part of the difference $\hat{C} - C$ can be passed from one RRC to another through the Missed frame and through overcoverage in the Census frame. This error could account for a large part of the difference $\hat{C}_{81} - C_{81}$. The effect on $\hat{C} - C$ may be much greater than on \hat{M} .

The rate of net coverage error in the 1981 Census, for the ten provinces, would be equal to:

$$\frac{M_{81} + M_{n81} - O_{81}}{C_{81} + M_{81} + M_{n81} - O_{81}};$$

and the rate of undercoverage would be:

$$\frac{M_{81} + M_{n81}}{C_{81} + M_{81} + M_{n81} - O_{81}}.$$

The estimator used in the RRC is

$$\frac{\hat{M}_{81}}{C_{81} + \hat{M}_{81}}.$$

Even a relatively small value of $M_{n81} - O_{81}$ could contribute significant bias to the results of the RRC, if these results are used as estimates of net coverage error. A relatively small value of $e(\hat{M}_{81})$ could contribute significant bias to the RRC undercoverage estimates: two potential elements of bias coming from the previous RRC; one from any misclassification within the RRC; and one from "missed" persons among those not included in an RRC frame. There may be, of course, some cancellation among these elements.

An alternative estimator would be to use \hat{C}_{81} instead of C_{81} in the denominator. There are specific and not unlikely circumstances under which the use of C_{81} would produce estimates with less bias at the national level. These circumstances, which involve the relative sizes of $\hat{C}_{81} - C_{81}$, O_{81} and M_{n81} do not hold, however, for provinces or estimates for which the Census count of enumerated is less than the RRC estimate of enumerated.

REFERENCES

- BURGESS, R.D. (1986). Major issues and implications of tracing survey respondents. International Symposium on Panel Surveys, Washington D.C.
- GOSSELIN, J.-F. (1976). The Methodology of the 1971 Reverse Record Check. *Survey Methodology*, 2, 180-193.
- PLAUT, G. (1985). Refugee determination in Canada. House of Commons Standing Committee on Labour, Manpower and Immigration.
- STATISTICS CANADA (1976 to 1981). *Vital Statistics Volume III: Death/Mortality* Catalogue 84-206, Statistics Canada.
- STATISTICS CANADA (1979). *Mobility Status and General Population Characteristics*. Catalogue 92-834, Statistics Canada.

- STATISTICS CANADA (1982). *Population Counts – 1976 and 1981 – Federal Electoral Districts*. Catalogue 99-908, Statistics Canada.
- STATISTICS CANADA (1983a). *Mobility Status*. Catalogue 92-907, Statistics Canada.
- STATISTICS CANADA (1983b). Reverse Record Check Tabulations, 1981 Data Quality Project, 1981 Census of Population, unpublished.
- STATISTICS CANADA (1984). Public Use Reverse Record Check Tape Users Guide. 1981 Census of Population, Statistics Canada.
- STATISTICS CANADA (1987). *Population Estimation Methods, Canada*. Catalogue 91-528E, Statistics Canada.
- STATISTICS CANADA (1988). Undercoverage Rates from the 1986 Reverse Record Check. User Information Bulletin, Number 2, Statistics Canada.