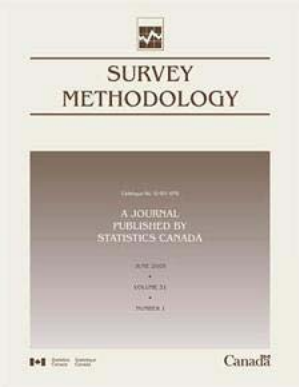
Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Census coverage error: A demographic evaluation

by Réjean Lachapelle and Don Kerr



June 2000

Statistics Canada Statistique Canada

Canada

Survey Methodology, June 2000
Vol. 26, No. 1, pp. 43-52
Statistics Canada, Catalogue No. 12-001

Census coverage error: A demographic evaluation

Réjean Lachapelle and Don Kerr [1]

Abstract

The 1996 Canadian Census is adjusted for coverage error as estimated primarily through the Reverse Record Check (RRC). In this paper, we will show how there is a wealth of additional information from the 1996 Reverse Record Check of direct value to population estimation. Beyond its ability to estimate coverage error, it is possible to extend the Reverse Record Check classification results to obtain an alternative estimate of demographic growth – potentially decomposed by component. This added feature of the Reverse Record Check provides promise in the evaluation of estimated census coverage error as well as insight as to possible problems in the estimation of selected components in the population estimates program.

Key Words: Census coverage error; Population estimates; Reverse record check.

1. Introduction

The Reverse Record Check (RRC), in various forms, has been used by Statistics Canada since the 1960's to estimate coverage error in the Canadian Census (Fellegi 1969; Brackstone and Gosselin 1973; Gosselin 1976; Burgess 1988; Carter 1990; Royce, Germain, Julien, Dick, Switzer and Allard 1994, Statistics Canada 1999). Using the Reverse Record Check, Statistics Canada has produced a long time series of population estimates, from 1971 through to the present, fully adjusted for census undercount. The current paper will demonstrate how there is additional information in the Reverse Record Check, which from a demographic perspective, can be exploited for the purposes of population estimation.

The demographic statistics program at Statistics Canada uses information from vital statistics, the most recent census, and various administrative sources in generating highly accurate and up to date population estimates. Information on births, deaths, immigration, emigration, among other demographic components, can be used to estimate population growth since the previous census. With each quinquennial census, a cycle ends and the accuracy of these estimates are put to the test (Romaniuc 1988). Systematic comparisons can be made between these estimates of growth and estimated growth as implied by comparing subsequent censuses (after adjustment for census coverage error).

The resultant difference (conventionally referred to as the error of closure of the intercensal population estimates) has a far from obvious interpretation. While a large error of closure is suggestive of problems in the population estimates, its specific nature is far from obvious (as to which demographic components are specifically responsible for the error). Furthermore, a honest appraisal of this closure error might suggest not only problems in the population estimates, but also potential problems in census coverage studies themselves (at the beginning and/or end of the intercensal period).

The current paper will demonstrate how an alternative estimate of demographic growth is possible, as based explicitly on the RRC classification results. Additional information is available, which assists greatly in the interpretation and decomposition of this closure error. Three alternative estimates of demographic growth for the intercensal period will be presented in the following section, including growth as estimated as part of the regular program of population estimates, implicit growth as obtained in comparing consecutive censuses, and growth as based explicitly on RRC classification results. Section 3 demonstrates how this RRC based estimate of growth can assist in the decomposition and interpretation of closure error, providing evidence of (i) bias in selected components of the population estimates, and (ii) possible problems in the RRC results Section 4 presents the results from this decomposition, followed by a brief discussion of its implications for both census coverage error measurement and the population estimates program.

2. Alternative estimates of demographic growth

2.1 Administrative record based estimates of growth: Post-censal estimates

Statistics Canada's regular program of population estimates involves the continuous registration and estimation of demographic events, as based on vital statistics and various administrative data sets. These events are added or subtracted from the population documented in the previous census (component method). In estimating a province's population on Census day 1996 (P_{est96}):

1. Réjean Lachapelle, Demography Division, Main Building, Tunney's Pasture, Statistics Canada, Ottawa, Ontario, K1A 0T6; Don Kerr, Department of Sociology, University of Western Ontario, London, Ontario, N6A 5C2.

$$P_{est96} = P_{91} + B_{91-96} - D_{91-96} + I_{91-96} - E_{91-96} + \Delta \text{NPR}_{91-96} + \text{NM}_{91-96}. \quad (1)$$

The baseline population (P_{91}) for this estimate builds on the 1991 Census after adjustment for all forms of coverage error, including net census undercount as measured through the 1991 RRC. The postcensal estimate can be obtained by adding or subtracting from this baseline the number of births between censuses (B_{91-96}), the number of deaths (D_{91-96}), immigrants (I_{91-96}), emigrants (E_{91-96}), net interprovincial migration (NM_{91-96}), and the net gain or loss of nonpermanent residents $(\Delta \text{NPR}_{91-96})$.

Non-permanent residents (NPRs) are persons with legal temporary status in Canada (*e.g.*, persons holding student or employment authorizations, minister's permits, refugee claimants, as well as their non-Canadian born dependents). Unlike with interprovincial migration, net gain or net loss of NPRs is not estimated through "flow" data on the ongoing in and out-flows of non-permanent residents, but rather estimated by comparing over time "stock" data on the total number of non-permanent residents living in the country. Further details of methodology, data sources and data quality issues can be obtained from the quarterly and annual releases of the population estimates program (Statistics Canada 1999; 2000).

2.2 Implicit estimate of growth

An implicit estimate of growth can be derived using the 1991 and 1996 Censuses, with both censuses adjusted for net undercount. With the exception of a small number of refusal Indian reserves, whose population figures are estimated independently, gross undercoverage was estimated entirely through the RRC in 1996, whereas gross overcoverage was a combined estimate from three studies (the RRC, the Collective Dwellings Study and the Automated Match Study). In 1991, the RRC was used only in the estimation of gross undercoverage, whereas gross overcoverage was estimated through a smaller study, the Private Dwelling Study, in combination with the 1991 Collective Dwelling and Automated Match studies. In addition, persons missed on refusal Indian reserves were estimated as part of the 1991 Reverse Record Check.

In the early evaluation of the 1996 coverage studies, the implicit growth obtained with the above adjustments was considered unrealistic. It has since been established that part of the 1991 estimate of net undercount was in error, and would have in reality been lower had selected methodological enhancements been introduced as in 1996 (Tourigny, Clark and Provost 1998). It has been shown that (i) a number of persons initially classified as missed in 1991 was too high due to misclassification, and (ii) the 1991 estimate of "overcount" was too low. As a result, 1991 estimates of undercount and overcount have been revised to reflect the impact of these methodological changes $({}^{\text{rev}}U_{91}, {}^{\text{rev}}O_{91})$. In addition, for reasons of consistency with 1996, separate modeled estimates of refusal Indian reserves (independent of the RRC) have been added to the Census in 1991.

More specifically, implicit growth (Δ^I) is obtained as:

$$\begin{aligned}\Delta^I &= P_{96} - P_{91} \\ &= \{P^c_{96} + U_{96} - O_{96} + \text{IR}_{96\text{M}} - \text{IR}_{\text{RRC96}}\} \\ &\quad - \{P^c_{91} + {}^{\text{rev}}U_{91} - {}^{\text{rev}}O_{91} + \text{IR}_{91\text{M}} - \text{IR}_{\text{RRC91}}\} \quad (2)\end{aligned}$$

where final population figures (P_{96}, P_{91}) are obtained using previously published census figures (P^c_{96}, P^c_{91}) adjusted for undercoverage $(U_{96}, {}^{\text{rev}}U_{91})$ and gross overcoverage $(O_{96}, {}^{\text{rev}}O_{91})$. In adding independently modeled estimates of refusal Indian reserves $(\text{IR}_{96\text{M}}, \text{IR}_{91\text{M}})$, it is necessary to remove that portion of the RRC estimate of gross undercoverage that corresponds to these reserves $(\text{IR}_{\text{RRC96}}, \text{IR}_{\text{RRC91}})$. The results presented in the current paper take these changes into consideration.

2.3.1 RRC based estimates of growth

The Reverse Record Check (RRC) is a record linkage and matching procedure that attempts to trace all persons in its sample, interview them to obtain a census day address, and match their records to individual census documents. This involves the construction of a sample intended to represent the same target population as the census being evaluated. This sampling frame, obtained in a manner that is totally independent of the census being evaluated, is constructed using the previous census, birth registrations over the intercensal period, administrative lists of intercensal immigrants, and an up-to-date listings of non-permanent residents. Persons missed in the previous census are represented by a sample of cases classified as "missed" in the previous RRC, in the absence of a complete list of such persons.

By working with this sample, the RRC targets all persons who could have potentially been part of the 1996 Census universe. Except for a very small sub-population of returning emigrants (Canadian citizens and landed immigrants who were abroad during the previous census), the RRC sample is complete and fully representative. The subsequent classification (missed, enumerated, emigrated, abroad, deceased or out of scope) is applied in the estimation of "missed" in the current census. At the same time, this classification also holds the potential for further inferences, *i.e.*, an additional estimate of demographic growth for the intercensal period.

To estimate demographic growth using the RRC, it is useful to consider the following two equations. In the first equation, the target population of the 1991 Census (P^T_{91}) is expressed in terms of all potential classification outcomes in 1996. In the second equation, it is possible to move in the opposite direction – by expressing the 1996 census target population (P^T_{96}) in terms of all possible statuses in 1991 (or in the case of births and immigrants, the intercensal period).

$$P_{91}^{T} = {}^{91}PP_{96} + {}^{91}NP_{96} + {}^{91\text{NP}}C_{96\text{PP}} + {}^{91\text{PP}}D_{96} + {}^{91\text{NP}}D_{96} + {}^{91\text{PP}}E_{96\text{FR}} + {}^{91\text{NP}}E_{96\text{FR}} + {}^{91\text{NP}}E_{96\text{EX}} \quad (3)$$

$$P_{96}^{T} = {}^{91}PP_{96} + {}^{91}NP_{96} + {}^{91\text{NP}}C_{96\text{PP}} + {}^{91-96}B_{96} + {}^{91\text{EX}}I_{96PP} + {}^{91\text{EX}}I_{96\text{NP}} + {}^{91\text{FR}}RE_{96\text{PP}} \quad (4)$$

where

- ${}^{91}PP_{96}$ - Canadian citizens and landed immigrants in Canada in 1991, also targeted by the 1996 census
- ${}^{91}NP_{96}$ - NPRs in Canada in 1991, also targeted by the 1996 census as NPRs
- ${}^{91\text{NP}}C_{96PP}$ - NPRs in Canada in 1991 who became landed immigrants over the intercensal period
- ${}^{91\text{PP}}D_{96}$ - Canadian citizens and landed immigrants in Canada in 1991, who died over the intercensal period
- ${}^{91\text{NP}}D_{96}$ - NPRs in Canada in 1991, who died over the intercensal period
- FR - persons with the right to live permanently in Canada (citizens and landed immigrants) that are not in the designated census target population
- ${}^{91\text{PP}}E_{96\text{FR}}$ - Canadian citizens and landed immigrants in Canada in 1991, who are outside the 1996 census target population
- ${}^{91\text{NP}}E_{96\text{FR}}$ - NPRs in Canada in 1991, who became landed immigrants or citizens, and are outside the 1996 census target population
- EX - persons who have never been citizens or landed immigrants, and are not in the designated census target population
- ${}^{91\text{NP}}E_{96\text{EX}}$ - NPRs in Canada in 1991, who did not become landed immigrants, and are outside the 1996 census target population
- ${}^{91-96}B_{96}$ - births over the 1991-1996 period, and in the 1996 census target population
- ${}^{91\text{EX}}I_{96\text{NP}}$ - persons not in Canada in 1991, who arrived over the intercensal period, and are NPRs in the 1996 census target population
- ${}^{91\text{EX}}I_{96\text{PP}}$ - immigrants who landed over the intercensal period, and are in the 1996 Census target population
- ${}^{91\text{FR}}RE_{96\text{PP}}$ - returning emigrants, *i.e.*, Canadian citizens and landed immigrants outside the census universe in 1991, and in the 1996 Census universe

An estimate of growth (Δ^{RRC}) can be obtained by subtracting the former equation from the latter:

$$\Delta^{\text{RRC}} = {}^{91-96}B_{96} + {}^{91\text{EX}}I_{96\text{PP}} + {}^{91\text{EX}}I_{96\text{NP}} - {}^{91\text{PP}}D_{96} - {}^{91\text{NP}}D_{96} - {}^{91\text{PP}}E_{96\text{FR}} - {}^{91\text{NP}}E_{96\text{FR}} - {}^{91\text{NP}}E_{96\text{EX}} + {}^{91\text{FR}}RE_{96\text{PP}}. \quad (5)$$

With the previously introduced sampling frames and classification outcomes, all terms (with the exception of the last term: returning emigrants) can be directly estimated from the 1996 RRC itself. The census target population in 1991 can be approximated through the sample drawn from the census and missed frames – with the identification of relevant classification outcomes. The census target population in 1996 can be approximated through all persons classified as either enumerated or missed in 1996. The final term (*i.e.*, returning emigrants) can be obtained independent of the RRC using the 1996 Census 5-year mobility variable, in identifying all persons outside the country five years ago (excluding recent immigrants and NPRs). It is possible to express this same RRC based estimate of demographic growth at the provincial level by incorporating an estimate of interprovincial migration. As the RRC relied on Health Care Files in Canada's two northern territories (the Yukon and NWT) with administrative lists of addresses current to census being evaluated, this estimate of growth is not possible for the relatively small populations living in Canada's far north.

A minor problem in the RRC design persists that potentially introduces a slight bias into its classification results. Unfortunately it is not possible to identify all NPRs in the RRC sample, with the potential for an unknown amount of frame overlap (*i.e.*, between the census, NPR and immigrant frames). As NPRs in the census can only be identified through the census long form (which is distributed to about 20% of all households), it is possible that some NPRs living in Canada in 1991, selected in the census frame, were also selected in either the immigrant or NPR frames (without being identified as such). While the RRC attempts to adjust for this overlap by identifying all such persons in the immigrant and NPR frames, an unknown bias exists to the extent that this is unsuccessful. This difficulty in identifying overlap leaves the potential of too many immigrants and/or NPRs in the sample, or too few, if too many persons are removed from the aforementioned frames. The latter outcome can subsequently deflate the estimate of demographic growth, gross undercoverage (among other classification outcomes), whereas the former has the opposite outcome.

2.3.2 RRC based estimate of growth: A more detailed decomposition

While both the postcensal and RRC based estimates of demographic growth should be highly comparable, the specific terms within each are not meant to be directly

equivalent. For example, births in the postcensal estimates denote all intercensal births occurring to a population – irrespective of whether such births move or die – whereas births in the discrete equation denote all births occurring yet still in Canada at the end of the intercensal period. With this in mind, it is possible to expand on the RRC based equation, to derive terms that are more comparable to those used in the postcensal estimates. The RRC based estimate of demographic growth can then be used in the evaluation of the components of demographic growth that enter into the component method.

To expand on this equation, it is useful to begin with births, again expressed in terms of possible RRC classification outcomes. As previously indicated, the birth term as included in equation (5) is only part of all births occurring over the intercensal period. More comprehensively, all births can be expressed as:

$$B^{91-96} = {}^{91-96}B_{96} + {}^{91-96B}D_{96} + {}^{91-96B}E_{96\text{FR}} \qquad (6)$$

where:

B^{91-96} = all intercensal births

${}^{91-96}B_{96}$ = all intercensal births ultimately classified as either enumerated or missed in 1996

${}^{91-96B}D_{96}$ = deaths of intercensal births

${}^{91-96B}E_{96\text{FR}}$ = Persons outside target population in 1996 yet born in Canada over the intercensal period

Similarly, all immigrants can be expressed as:

$$I^{91-96} = {}^{91\text{EX}}I_{96\text{PP}} + {}^{91\text{NP}}C_{96\text{PP}} + {}^{91-96\text{I}}D_{96} + {}^{91-96\text{I}}E_{96\text{FR}} \qquad (7)$$

where:

${}^{91\text{EX}}I_{96\text{PP}}$ = intercensal immigrants ultimately classified as either enumerated or missed in 1996

${}^{91\text{NP}}C_{96\text{PP}}$ = all NPRs in 1991 who obtained landed immigrant status and are ultimately classified as either enumerated or missed in 1996

${}^{91-96\text{I}}D_{96}$ = deaths occurring to landed immigrants over the intercensal period

${}^{91-96I}E_{96\text{FR}}$ = emigrants among intercensal immigrants (irrespective of whether or not they were living in Canada as NPRs in 1991)

In combining equations 5, 6 and 7, demographic growth can be requested as:

$$\begin{aligned} P_{96}^T - P_{91}^T = {} & B^{91-96} - {}^{91\text{PP}}D_{96} - {}^{91\text{NP}}D_{96} - {}^{91-96\text{B}}D_{96} \\ & - {}^{91-96I}D_{96} + I^{91-96} + {}^{91\text{FR}}RE_{96\text{PP}} - {}^{91\text{NP}}C_{96PP} \\ & - {}^{91\text{PP}}E_{96\text{FR}} - {}^{91\text{NP}}E_{96\text{FR}} - {}^{91-96\text{B}}E_{96\text{FR}} \\ & - {}^{91-96\text{I}}E_{96\text{FR}} - {}^{91\text{NP}}E_{96\text{EX}} - {}^{91\text{EX}}I_{96\text{NP}}. \end{aligned} \qquad (8)$$

Given that the final term of (8) is equivalent to:

$$\begin{aligned} {}^{91\text{EX}}I_{96\text{NP}} = {} & NP_{96} - NP_{91} + {}^{91\text{NP}}D_{96} + {}^{91\text{NP}}E_{96\text{EX}} \\ & + {}^{91\text{NP}}C_{96\text{PP}} + {}^{91\text{NP}}E_{96\text{FR}}. \end{aligned} \qquad (9)$$

If follows that:

$$\begin{aligned} & P_{96}^T - P_{91}^T = \\ & B^{91-96} - {}^{91\text{PP}}D_{96} - {}^{91\text{NP}}D_{96} - {}^{91-96\text{B}}D_{96} - {}^{91-96\text{I}}D_{96} \\ & + I^{91-96} + {}^{91\text{FR}}RE_{96\text{PP}} - {}^{91\text{NP}}C_{96\text{PP}} \\ & - {}^{91\text{PP}}E_{96\text{FR}} - {}^{91\text{NP}}E_{96\text{FR}} - {}^{91-96\text{B}}E_{96\text{FR}} \\ & - {}^{91-96\text{I}}E_{96\text{FR}} - {}^{91\text{NP}}E_{96\text{EX}} + NP_{96} \\ & - (NP_{91} - {}^{91\text{NP}}D_{96} - {}^{91\text{NP}}E_{96\text{EX}} - {}^{91\text{NP}}C_{96\text{PP}} - {}^{91\text{NP}}E_{96\text{FR}}) \end{aligned} \qquad (10)$$

or

$$\begin{aligned} & P_{96}^T - P_{91}^T = \\ & (B^{91-96}) - ({}^{91\text{PP}}D_{96} - {}^{91-96\text{B}}D_{96} - {}^{91-96\text{I}}D_{96}) + (I^{91-96}) \\ & - ({}^{91\text{PP}}E_{96\text{FR}} + {}^{91-96\text{B}}E_{96\text{FR}} + {}^{91-96\text{I}}E_{96\text{FR}} - {}^{91\text{FR}}RE_{96\text{PP}}) \\ & + (NP_{96} - NP_{91}). \end{aligned} \qquad (11)$$

This expanded version of equation (5) provides a breakdown of demographic growth at the national level, and allows for more meaningful comparisons with components estimated through administrative records. All terms, except for ${}^{91\text{FR}}RE_{96\text{PP}}$ and NP_{91} can be obtained directly from the 1996 RRC. The aforementioned hole in the RRC sampling frame requires an independent estimate of returning emigrants whereas the nature of the sample frame for NPRs explains the absence of the latter term. Rather than a listing of all NPRs to enter Canada over the intercensal period (as was the case with immigrants), the RRC relies on the most up to date administrative listing of NPRs in the establishment of its sampling frame (with no information on the number of NPRs living in Canada in 1991).

Postcensal estimates document demographic growth through the "continuous" registration and estimation of demographic events over time. The RRC estimates growth via information on the status of individuals as identified on at least two "discrete" dates (at the beginning and end of the intercensal period). Irrespective of this minor conceptual distinction between "continuous" versus "discrete" estimation, each term of equation 11 (within each set of parenthesis) roughly corresponds to a separate component as documented using administrative records. The first term identifies all intercensal births (*i.e.*, the weighted sum of the birth frame), the second term includes deaths (classification results across the birth frame, the missed frame, the census frame and immigrant frame), the third term includes all

immigrants (*i.e.*, the weighted sum of the immigrant frame), the fourth term includes emigrants (classification results across the birth frame, the immigrant frame, the missed frame census frames, as well as the returning emigrant component), and the fifth term corresponds to net gain or loss of NPRs. As the number of NPRs living in Canada in 1991 is not available in the 1996 RRC, for current purposes, this latter term is obtained using the 1991 census count, after adjustments for undercoverage. Again, it is possible to express this equation at the provincial level.

With equation (11), a detailed evaluation of the postcensal estimation program is possible. For example, if differences persist between RRC based estimates and postcensal estimates, it is possible to determine how much of the difference in estimated growth can be traced back to differences in migration (typically estimated with some difficulty in the postcensal estimates program) and how much can be traced to differences in natural increase. Briefly, Table 1 includes all of the aforementioned estimates of growth, including implicit growth, the growth as based on administrative records, and the two alternate estimates of growth as based on the RRC (simplified and expanded equations). Slight differences exist between the simplified and expanded equations – yet not nearly of the same size as with the other estimates (implicit, postcensal). In explanation of the differences between the two RRC based estimates, the simplified equation does not require the same detailed classification as with the expanded equation, is not biased to the same extend by the aforementioned problem of frame overlap, and does not rely on the 1991 census count of NPRs. The differences observed with the remaining estimates are the focus of the current decomposition.

Table 1
Alternate estimates of growth, 1991-1996, Canada and provinces/territories

	Implicit growth	Population estimates administrative records	RRC simplified	RRC expanded
NFLD.	-17,997	-9,263	-17,897	-17,751
P.E.I.	5,404	5,483	2,568	1,583
N.S.	15,781	24,271	17,075	16,860
N.B.	7,714	13,097	12,017	11,276
QUE.	206,307	300,849	261,357	252,014
ONT.	659,349	766,568	668,443	655,572
MAN.	23,682	24,981	7,377	6,288
SASK.	15,953	11,098	11,524	9,312
ALTA.	186,594	186,986	151,944	159,907
B.C.	505,025	466,611	465,864	472,342
YUKON	3,085	2,329	N/A	N/A
N.W.T.	6,837	5,864	N/A	N/A
Provinces (excl terr)	1,607,771	1,790,681	1,580,273	1,567,404
Canada	1,617,693	1,798,874	N/A	N/A

3. Decomposition of closure error

Implicit growth for the 1991-96 period is obtained only after all adjustments have been made to the censuses for coverage error, including revised 1991 figures on gross undercount and overcount and refinements for refusal Indian reserves. Alternatively, the RRC based estimate of growth (simplified version) is obtained in working with approximations of the 1991 and 1996 target populations, *i.e.*, the census and missed frames of the 1996 RRC and all persons classified as either missed or enumerated in this study. For this reason, there are minor differences between the two estimates that need to be more clearly identified in a full decomposition of closure error. In this context, it is useful to express implicit growth obtained with final population figures in terms of these approximations (sampling frames and classification outcomes). In a similar manner, as the error of closure is the difference between implicit growth and the growth associated with the postcensal estimates, the error of closure can also be expressed in terms of these approximations.

To simplify the presentation, let δ represent all possible negative growth terms in equation (5) and η as all possible positive growth terms:

$$\delta = ({}^{91\mathrm{PP}}D_{96} + {}^{91\mathrm{NP}}D_{96} + {}^{91\mathrm{PP}}E_{96\mathrm{FR}} + {}^{91\mathrm{NP}}E_{96\mathrm{FR}} + {}^{91\mathrm{NP}}E_{96\mathrm{EX}}) \quad (12)$$

$$\eta = ({}^{91-96}B_{96} + {}^{91\mathrm{EX}}I_{96\,\mathrm{PP}} + {}^{91\mathrm{EX}}I_{96\,\mathrm{NP}} + {}^{91\mathrm{FR}}RE_{96\,\mathrm{PP}}) \quad (13)$$

The population enumerated in both censuses can be represented as:

$${}^{91}P_{96} = ({}^{91}PP_{96} + {}^{91}NP_{96} + {}^{91\mathrm{NP}}C_{96\,\mathrm{PP}}). \quad (14)$$

Since the final population figures (P_{91}, P_{96}) used in the estimation of implicit growth involve separate modeled estimates for refusal Indian reserves, it is useful to restate the RRC based estimate of growth after specifically delineating such reserves. In designating persons living in refusal reserves in 1996 that were in the target population in 1991 as ${}^{91}IR_{96}$, the growth of these reserves through either migration or birth as estimated by the RRC by η_{IR}, and redefining ${}^{91}P_{96}$ to exclude all persons associated with these two terms, it is possible to return to equations (3) – (5) as:

$$P_{91}^{\mathrm{T}} = {}^{91}P_{96} + {}^{91}IR_{96} + \delta \quad (15)$$

$$P_{96}^{\mathrm{T}} = {}^{91}P_{96} + {}^{91}IR_{96} + \eta_{\mathrm{IR}} + \eta \quad (16)$$

$$\Delta^{\mathrm{RRC}} = P_{96} - P_{91} = \eta + \eta_{\mathrm{IR}} - \delta. \quad (17)$$

In expressing implicit growth in terms of the RRC sampling frames and classification outcomes, it is useful to build on the RRC estimate of growth (equation 17) in defining total growth beginning with P_{91} rather than P_{91}^{T}. In recognition that the final population estimate (P_{91}) is equivalent to the census and missed frames (P_{91}^{T}) minus overcoverage $({}^{rev}O_{91})$ plus refinements for refusal Indian reserves $(IR_{91M} - IR_{RRC91})$, it follows:

$$P_{96}^{T} - P_{91} = \eta + n_{IR} - \delta + {}^{rev}O_{91} + (IR_{RRC91} - IR_{91M}). \quad (18)$$

On the other hand, this target population (P_{96}^{T}) can also be expressed as:

$$P_{96}^{T} = EN_{96} + U_{96} + {}^{91FR}RE_{96PP} \quad (19)$$

where EN_{96} is an estimate of the number of persons enumerated in 1996. In recalling from equation 2 that:

$$P_{96} = P_{96}^{c} + U_{96} - O_{96} + (IR_{96M} - IR_{RRC96}) \quad (20)$$

implicit growth (Δ^{I}) can be expressed in terms of the RRC based estimates of growth, as:

$$\Delta^{I} = P_{96} - P_{91} = (P_{96} - P_{96}^{T}) + (P_{96}^{T} - P_{91}) = \{(\eta - \delta)\} + \{\eta_{IR} - (IR_{91M} - IR_{RRC91}) + (IR_{96M} - IR_{RRC96})\} + \{(P_{96}^{c} - EN_{96} - {}^{91FR}RE_{96PP} + {}^{rev}O_{91} - O_{96})\}. \quad (21)$$

Implicit growth (Δ^{I}) can be defined as the sum of (i) a RRC based estimate of growth (excluding refusal Indian reserves), (ii) a second term depending on the decision to estimate the refusal Indian reserves by an independent model, and (iii) a third term that involves a comparison of the RRC based estimate of enumerated and the number of persons actually enumerated in the 1996 census.

This latter term (the difference on enumerated) has an interesting interpretation, and is considered fundamental to the evaluation of the RRC (Tourigny, Bureau and Clark 1998; Royce 1993). Significant differences with this term can be read as implying either sampling errors and/or possible biases, as either classification error and/or problems in sample selection. To make this comparison meaningful, 1996 overcoverage and an estimate of returning emigrants are removed from the census counts – as neither can be included in the estimate of enumerated. Similarly, since the RRC selects part of its sample from the previous census, it inevitably carries forward some overcoverage inherent in its weights – which must subsequently be removed from its estimate of enumerated. These adjustments are included in the third term (the third set of brackets) in equation 21.

While the estimate of enumerated is inflated by the weights associated with overcoverage in the 1991 Census frame, only a portion is directly associated with this estimate – with the remainder spread across the other classification outcomes. Consequently, all classification results in the aforementioned equations are also slightly overstated. For the purposes of the current decomposition, this minor distinction is ignored. This is another reason, albeit of minor impact, why the RRC-based estimate of growth is different from the implicit estimate, as the latter is not biased by this overcoverage.

From the above, the error of closure is equivalent to:

$$\Delta_{91-96}^{D} - \Delta_{91-96}^{I} = [\Delta_{91-96}^{D} - \{(\eta - \delta)\} - \{\eta_{IR} - (IR_{91M} - IR_{RRC91}) + (IR_{96M} - IR_{RRC96})\}] - [\{(P_{96}^{c} - EN_{96} - {}^{91FR}RE_{96PP} + {}^{rev}O_{91} - O_{96})\}]. \quad (22)$$

In the decomposition of closure error, the first term inside brackets [ ] highlights the difference between the postcensal estimate of growth and the combined RRC estimate of growth (including refusal reserves, after refinements for modeled estimates). The second term (the difference on enumerated) provides evidence as to possible difficulties in the coverage studies. Theoretically, with the absence of sampling and non-sampling error in the RRC, this latter term should be negligible.

4.1 Decomposition results: Closure error

Table 2 presents closure error after finalizing both the 1991 and1996 estimates of population. By adding net undercount to the 1996 published Census figures, along with independent estimates of refusal Indian reserves, Canada's 1996 Census day population, adjusted for coverage error is estimated at 29,619,539. This figure is appreciably lower than the Census day estimate as generated through the postcensal estimates program of 29,800,720. The difference between the two figures – which is equivalent to the aforementioned difference between implicit growth and growth as based on administrative records – was higher than anticipated given past experience, at 181,181 (or 0.61% of the 1991 Census Day population).

Across provinces/territories, closure error is found to be particularly pronounced in Newfoundland (1.56%), in Canada's north (at -2.38% in Yukon and -1.44% in the NWT), and somewhat surprisingly, in its three largest provinces (as 1.30% in Quebec, 0.97% in Ontario and -0.99% in British Columbia). Regionally, closure errors larger than the national average are observed across eastern and central Canada (except for P.E.I.) while the western provinces have closure errors lower than the national one. It is specifically these errors that the current decomposition seek to evaluate and explain.

Table 3 presents the results from this decomposition, with closure error decomposed into (i) the difference

between the estimate of growth based on administrative records and the RRC based estimate (simplified version), and (ii) the difference on enumerated. Also included is the sampling error associated with the RRC estimates.

4.2 Comparisons between estimates of growth

Across all provinces (with the exception of Saskatchewan), growth estimated on the basis of administrative records is higher than the RRC based estimate. At the national level (excluding the territories), this discrepancy on growth (210,408) appears far more important in explaining closure error than the discrepancy on enumerated (-27,498). While for many provinces the difference on growth fell well within expectations in light of sampling error, selected provinces require further explanation. For example, the difference in growth in Ontario is large (98,125), which is almost one half the difference observed

Table 2
Coverage study results, relative to population estimate (1996 - Census day)

	{1}	{2}	{3}	{4 = 1 + 2 + 3}	{5}	{6 = 5 - 4}	{7 = 6 / 4 * 100}
	1996 census count with random additions	1996 net undercount	Indian reserves	1996 Census RRC adjusted	1996 estimate post-censal (i)	Error of closure	Error of closure (%)
NFLD.	551,792	9,424	0	561,216	569,950	8,734	1.56
P.E.I.	134,557	1,149	175	135,881	135,960	79	0.06
N.S.	909,282	20,821	0	930,103	938,593	8,490	0.91
N.B.	738,133	14,225	518	752,876	758,259	5,383	0.71
QUE.	7,138,795	116,750	12,427	7,267,972	7,362,514	94,542	1.30
ONT.	10,753,573	301,368	20,849	11,075,790	11,183,050	107,260	0.97
MAN.	1,113,898	18,881	315	1,133,094	1,134,393	1,299	0.11
SASK.	990,237	28,051	586	1,018,874	1,014,019	-4,855	-0.48
ALTA.	2,696,826	66,327	11,287	2,774,440	2,774,832	392	0.01
B.C.	3,724,500	142,443	3,136	3,870,079	3,831,665	-38,414	-0.99
YUKON	30,766	1,022	0	31,788	31,032	-756	-2.38
N.W.T.	64,402	3,024	0	67,426	66,453	-973	-1.44
Canada	28,846,761	723,485	49,293	29,619,539	29,800,720	181,181	0.61

(i) Post-Censal Estimates for May 14th, obtained with final components for intercensal estimates.
Final Estimates (Sept. 24th, 1998) of Net Undercount, 1991 an 1996.

Table 3
Decomposition of closure error

Province/Territory	Error of closure	Difference between		Difference on enumerated	S.E. of estimates
		Dem. and RRC estimates of growth	S.E. of estimates		
NFLD.	8,734	8,634	4,889	100	5,176
P.E.I.	79	2,915	2,425	-2,836	2,462
N.S.	8,490	7,196	9,011	1,294	9,455
N.B.	5,383	1,080	7,793	4,303	7,918
QUE.	94,542	39,492	25,493	55,050	29,310
ONT.	107,260	98,125	41,212	9,135	51,300
MAN.	1,299	17,604	10,108	-16,305	10,370
SASK.	-4,855	-426	9,187	-4,429	10,200
ALTA.	392	35,042	19,067	-34,650	21,618
B.C.	-38,414	747	20,518	-39,161	22,996
YUKON	-756	N/A	N/A	-108	270
N.W.T.	-973	N/A	N/A	-284	464
Canada without Territories	182,910	210,408	43,951	-27,498	58,724
Canada	181,181	N/A	N/A	-27,890	58,762

Table 4
Estimated components (1991-1996) as compiled by demography division and RRC discrete (detailed) measurement

	NFLD	PEI	NS	NB	QUE	ONT	MAN	SASK	ALB	BC	CANADA (without terr)
Births											
Demography	31,748	8,803	55,994	44,444	453,556	730,520	81,485	70,382	199,484	229,511	1,905,927
RRC	31,779	8,782	55,984	44,444	454,332	729,744	81,485	70,382	199,484	229,511	1,905,927
Difference	-31	22	10	0	-776	776	0	0	0	0	0
Deaths											
Demography	-19,286	-5,692	-37,677	-28,567	-252,628	-376,760	-45,858	-40,652	-75,798	-126,935	-1,009,853
RRC	-18,530	-6,913	-43,820	-29,354	-273,617	-400,047	-56,108	-40,143	-74,640	-138,433	-1,081,605
Difference	-756	1,221	6,143	787	20,989	23,287	10,250	-509	-1,158	11,498	71,752
Immigration											
Demography	3,411	771	14,489	3,359	189,905	618,869	22,004	11,282	84,130	213,506	1,161,726
RRC	3,538	820	14,058	3,614	189,905	618,870	22,129	11,157	84,130	216,892	1,165,113
Difference	-127	-49	431	-255	0	-1	-125	125	0	-3,386	-3,387
Emigration											
Demography	-671	-206	-2,297	-2,429	-15,490	-48,609	-5,684	-2,493	-19,718	-17,834	-115,431
RRC	-2,227	-455	-7334	-3,889	-55,766	-168,556	-10,871	-7,133	-33,689	-31,739	-321,659
Difference	1,556	249	5,037	1,460	40,276	119,947	5,187	4,640	13,971	13,905	206,228
Interprovincial Migration											
Demography	-23,074	1,643	-5,288	-3,255	-51,176	-40,850	-25,336	-26,644	7,155	167,809	984
RRC	-32,767	-886	-1,479	-2,933	-49,395	-37,505	-29,765	-25,095	-10,321	191,222	1,076
Difference	9,693	2,529	-3,809	-322	-1,781	-3,345	4,429	-1,549	17,476	-23,413	-92
Non-permanent Residents											
Demography	-1,406	164	-950	-455	-23,353	-116,602	-1630	-777	-8,267	554	-152,722
RRC	455	236	-549	-606	-13,445	-86,934	-582	144	-5,057	4,890	-101,448
Difference	-1,861	-72	-401	151	-9,908	-29,668	-1,048	-921	-3,210	-4,336	-51,274
Total											
Demography	-9,263	5,483	24,271	13,097	300,849	766,568	24,981	11,098	186,986	466,611	1,790,681
RRC	-17,751	1,583	16,860	11,276	252,014	655,572	6,288	9,312	159,907	472,343	1,567,404
Difference	8,488	3,900	7,411	1,821	48,835	110,996	18,693	1,786	27,079	-5,731	223,277

nationally. Similarly, Newfoundland, Quebec, Alberta and Manitoba, together explain a large part of this difference.

In providing some indication as to the factors responsible for these differences, Table 4 presents comparisons using equation 11 (detailed equation). Alternative estimates are provided on (i) births, (ii) deaths, (iii) immigration, (iv) emigration, (v) interprovincial migration and (vi) net change in the number of non-permanent residents. The most important problems in the explanation of closure error are obvious in Table 4, with specific reference to emigration. As Canada does not have a complete border registration system, emigration is clearly the weakest of all the components to enter into the population estimate program. Without access to direct information on the number of persons leaving Canada, the RRC, with its exhaustive tracing, record linkage and direct interviewing procedures, is considered an improvement over any other data sources currently available. Although there are known problems in the RRC (for example, the previously mentioned frame overlap), the current evaluation points to an obvious error in the postcensal estimates, *i.e.*, an understatement of population outflow from Canada. Overall, the difference as observed nationally (206,228) explains the bulk of the closure error documented in 1996. Similarly with Ontario, difficulties in the estimation of emigration appear to be fundamental (with a difference of fully 119,947).

Without being decisive, the current decomposition also suggests other problematic components beyond emigration in the explanation of closure error for specific provinces. For example, the results suggest that estimates of interprovincial migration might be somewhat misstated for British Columbia and Newfoundland (after acknowledging the differences observed on these components and corresponding closure errors). Overall, an acceptance of the RRC on these more difficult to estimate migratory flows – would not only explain the largest part of this difference in growth – but also the largest part of 1996 closure error. With the closure error that remains, it is useful to turn to the observed difference on enumerated. In so doing, the emphasis shifts away from potential problems in the postcensal estimates.

4.3 Comparisons between estimates of enumerated

While the difference in enumerated observed nationally is much smaller than the difference documented on growth, for about half the provinces, this difference is of comparable if not larger size. In interpreting this fact, it is recognized that the RRC was never designed to target the "enumerated" population. With the priority of documenting the number "missed" in the census, the sampling design of the RRC over represents "difficult to enumerate groups"

(for example, single young adults), while under representing persons easily "enumerated". Overall, the comparison on enumerated bears well for the accuracy of the RRC – with non-significant differences across all provinces/territories. Nevertheless, the differences observed in a few provinces are reason for concern, being very close to statistical significance at the 95% level in Quebec (positive difference), and approaching statistical significance in British Columbia, Alberta and Manitoba (negative differences).

In the evaluation of the 1991 coverage study results, two alternative hypotheses have been raised in explanation of differences observed for the enumerated (Royce 1993). At one extreme, it could be argued that all of the difference (for a specific province) be explained in terms of the representativeness of the RRC sample, which implies sampling error or frame deficiencies of one sort or another. At the other extreme, it could be argued that all of the difference be explained due to a failure in documenting the true ratio of enumerated to other classification outcomes, which seems to imply some sort of misclassification error or no trace adjustment bias. A correction for the former of the two hypotheses has a relatively minor impact on the estimate of missed (*i.e.*, all classification outcomes are accordingly inflated or deflated by the proportional difference on enumerated). A correction for the latter could have potentially quite a pronounced impact, as a failure to estimate the true ratio implies that all the difference be assigned to other categories.

If the latter hypothesis applies, a correction potentially reduces the error of closure in nine out of twelve provinces/territories (*i.e.*, in all provinces under which the error of closure is in the same direction as the difference on enumerated). On the other hand, if the differences are due to problems in sample representativeness, a subsequent correction is expected to have negligible impact, if not slightly inflating closure error across most provinces. In addition, the evaluation is complicated by the difficulty in establishing the comparable census figures. Error can be potentially introduced through various sources, including: the census-based estimate of returning emigrants ($^{91FR}RE_{96PP}$), too much or too little correction for frame overlap, sampling and non-sampling error in the estimation of undercoverage in 1991 and 1996, sampling and non-sampling error in the estimation of overcoverage, and potential error in the classification by province of the enumerated. In this context, further research appears justified as to the true character of errors in the RRC estimate of enumerated.

5. Conclusion

In this paper, we have shown how there is additional information available through Canada's census coverage measurement program that is of considerable value in population estimation. Beyond the ability to estimate census undercount, it is possible to extend the classification results from these studies in order to obtain an alternative estimate of demographic growth – potentially decomposed by component. Using the most important of the coverage studies (*i.e.*, the 1996 Reverse Record Check), a new method was presented which allows for an independent estimate of demographic growth for the intercensal period. The Reverse Record Check not only provides what are considered highly accurate estimates of census coverage error, avoiding some of the correlation biases that have hindered post-enumeration studies in other countries, but also provides very valuable insight as to the magnitude of selected migratory flows of importance to population estimation.

The key to the Reverse Record Check is that it begins with a representative sample of all persons who could have theoretically been in Canada on census day, with only minor deficiencies due to the high quality of vital statistics and immigration data in Canada. Through exhaustive tracing and interviewing procedures, valuable information is then obtained as to the number and characteristics of persons successfully enumerated, missed, counted more than once, as well as useful information on the numbers leaving the country (whether temporarily or permanently), the numbers dying, living in another province, and so on. With a relatively large sample and considerable expertise and effort directed toward minimizing all forms of error, the resultant classification results can potentially inform the population estimates program. This is particularly true with some of the more difficult to estimate migratory flows.

In planning for the 2001 Census, the goal of minimizing all error in the census coverage measurement program remains a priority. As these studies have been designed with a primary target of estimating the population "missed" rather than other classification outcomes (emigrated, deceased, *etc*.), the new demographic approach presented in the current paper leads to the logical question, as to whether its current design need be reworked somewhat if its current usage is broadened. Of interest in this context is the fact that these coverage studies appear to provide an alternative estimate of growth which rivals that as currently available through the population estimates program, and is likely superior with respect to selected components. Further research about how we might more fully exploit this fact appears justified, in improving the quality of the population estimates program.

Acknowledgements

We would like to thank R.G. Carter and P. Dick (both of Statistics Canada) and G Robinson (U.S. Bureau of the Census) for their comments on an earlier draft of this paper. We also acknowledge the helpful comments and suggestions from the Associate Editor and two referees.

References

Brackstone, G.J., and Gosselin, J.-F. (1973). *Census Evaluation Program, 1971 RRC: Methodology Report*. Statistics Canada. Ottawa, Ontario.

Burgess, R.D. (1988). Evaluation of reverse record check estimates of undercoverage in the canadian census of population. *Survey Methodology*, 14, 137-156.

Carter, R.G. (1990). The Measurement of net coverage error in Canadian censuses. *Proceedings*: *Symposium 90*, *Measurement and Improvement of Data Quality*, Statistics Canada.

Fellegi, I.P. (1969). A theory of record linkage. *Journal of the American Statistical Association*, 64(328), 1183-1210.

Gosselin, J.-F. (1976). The methodology of the 1971 reverse record check. *Survey Methodology*, 2, 180-194.

Romaniuc, A. (1988). A demographic approach to the evaluation of undercoverage in the canadian census of population. *Survey Methodology*, 14,157-172.

Royce, D. (1993). Evaluation of the May 1993 Revised Results of the 1991 Census Coverage Studies. Social Survey Methods Division, Working paper, Statistics Canada, Ottawa, Ontario.

Royce, D., Germain, M.-F., Julien, C., Dick, P., Switzer, K. and Allard, B. (1994). *Coverage*: *1991 Census Technical Report*. Catalogue no. 92-341E. Ottawa: Statistics Canada.

Statistics Canada (1999). *Coverage*: *1996 Census Technical Reports*. Catalogue no. 92-370-XPB.

Statistics Canada (2000). *Annual Demographic Statistics*. Catalogue no. 21-213-XPB.

Tourigny, J., Clark C. and Provost, M. (1998). Evaluation of the March 1998 Preliminary Results of the 1996 Census Coverage Studies. Social Survey Methods Division, Working Paper, Statistics Canada, Ottawa, Ontario.

Tourigny, J., Bureau, M. and Clark, C. (1998). Revised Direct Estimates of 1991 Census Coverage Studies. Sept. 24th release. Social Survey Methods Division, Working Paper, Statistics Canada, Ottawa, Ontario.