Survey Methodology, December 1988
Vol. 14, No. 2, pp. 157–171
Statistics Canada

A Demographic Approach to the Evaluation of the 1986 Census and the Estimates of Canada's Population

ANATOLE ROMANIUC[1]

ABSTRACT

A significant increase in coverage error in the 1986 Census is revealed by both the Reverse Record Check and the demographic method presented in this paper. Considerable attention is paid to an evaluation of the various components of population growth, especially interprovincial migration. The paper concludes with an overview of two alternative methods for generating postcensal estimates: the currently-in-use, census-based model, and a flexible model using all relevant data in combination with the census.

KEY WORDS: Census undercoverage; Population estimates; Demographic component method.

1. INTRODUCTION

The accuracy of the census, and of the postcensal population estimates based thereon, is an important issue in its own right. The use of population numbers in the formulae for calculating revenue transfers between various levels of government, makes the question of accuracy all the more critical and politically sensitive (Fellegi 1980; Romaniuc and Raby 1980). The intense debates on whether or not to adjust population counts for census undercoverage in Canada and the USA, and several judicial litigations fought in the latter country in recent years, are indications of both the political importance and the technical complexity of the issue.

Yet, in spite of all that has been written on the subject, the elaborate arguments marshalled by both those for and those against adjustment, the debates remain inconclusive (Keyfitz 1979 and 1981; Kish 1980; Spencer 1980; Freedman and Navidi 1986; Stoto 1987). Eventually Statistics Canada decided (as did the US Department of Commerce) against adjustment for census undercoverage, while at the same time reaffirming its long-standing commitment to the policy of data quality evaluation (Wilk 1981). By making public both the evaluation results and the underlying methodology, the users can make adjustments to suit their particular needs, in full knowledge of the strengths and limitations of the census counts and estimates. It is in the spirit of this policy on quality evaluation that this paper has been written.

There are basically two approaches to the evaluation of the accuracy of census counts. One is the "micro" approach, involving individual verification, case-by-case record matching, in order to identify persons who have been missed, enumerated more than once, or enumerated even though, by definition, they are not part of the census universe. To this type of evaluation belong the US Bureau of the Census Post-Enumeration Program and Statistics Canada's Reverse Record Check (RRC).

[1] Anatole Romaniuc, Director, Demography Division, Statistics Canada, Ottawa, Ontario.

The second is the "macro" evaluation approach involving an analysis at aggregate levels, such as comparison of the census counts with figures derived from independent sources or with estimates arrived at by means of statistical and demographic methods. Following the pioneering work by Ansley Coale (1955), the demographic techniques of analysis have been used by the US Bureau of the Census to evaluate census coverage concurrently with the Post-Enumeration Program (see most recent report by Fay, *et al.* 1988). Some earlier attempts of this kind in Canada were also made (Lapierre 1970). The essence of the demographic method, as we shall see later, is that it brings to bear the formal relationship between population and its growth components - namely births, deaths and migration.

The evaluation of the 1986 Census coverage through the Reverse Record Check (RRC) has been carried out and reported upon elsewhere (Carter 1988; Statistics Canada 1988). It suffices to say that the RRC-based estimates of undercoverage are subject to sampling error - which can be quite significant for provinces with a small population - and to biases of unknown magnitudes (difficulties in tracing persons or matching individual records). Furthermore, the RRC has been designed primarily to measure undercoverage. The measurement of overcoverage has been attempted on an experimental basis, but at the time of writing, the results were unavailable. For these and similar reasons, an alternative assessment of the accuracy of the census counts becomes all the more important.

This paper evaluates, by means of demographic analysis, the accuracy of the three most recent censuses, with emphasis on the 1986 Census. A three-step operation is followed. First, census counts and population estimates are compared with each other. Second, demographic techniques are used to generate alternative estimates of census undercoverage which are, in turn, compared with those based on the Reverse Record Check. As a third and final step, the focus of evaluation is shifted from census counts to intercensal change in population. Two sets of independent estimates of intercensal population change are produced. One is based on the two consecutive censuses, while the other is obtained directly from data on births, deaths and migration.

Before proceeding with the actual evaluation, a word of caution is in order. Though of acceptable quality for most of the uses they serve, neither census counts nor population estimates are perfect. Indeed, there is no one set of data deemed to be perfect enough to serve as a benchmark for the validation of other data. The statistical reality is that data are imperfect in varying degrees. The fine tuning and high precision that would be required for particular uses - such as government allocations and revenue transfers referred to earlier - might not be attainable under the present state of the art. However, we hope that this evaluation, using a combination of statistical tools, imperfect as they may be, will enable us to get some sense of the direction and magnitude of errors and biases affecting census population counts and various components of population estimates. Such an undertaking will hopefully set the stage for improvements as we work toward the 1991 Census and the post-1991 population estimation methodology.

2. CENSUS COUNTS VERSUS POPULATION ESTIMATES: ERROR OF CLOSURE

The postcensal estimates of population are obtained, as per equation 1, by the so-called component method, whereby births and immigrants are added to, and deaths and emigrants are subtracted from, the base census population. The net interprovincial migration is then added to estimate population by province. The procedure is repeated annually over the five-year period

to the next census. The current estimation methodology calls for postcensal estimates to be revised retrospectively so as to bring them in line with the latest census counts (Statistics Canada 1987). The difference, as per equation 2, between estimates thus arrived at and census counts is termed "the error of closure" (EC).

$$\hat{P}_t = R_{t-5} + \left[B_{t-5,t} - D_{t-5,t} + I_{t-5,t} - \hat{E}_{t-5,t} + \hat{N}_{t-5,t}\right] \quad (1)$$

$$EC\ (\%) = \frac{\hat{P}_t - R_t}{R_t} \times 100, \quad (2)$$

where:

$\hat{P}_t$ = estimated population at time t;

R = census counts at time t or $t-5$ as the case may be;

B = number of births;

D = number of deaths;

I = number of immigrants;

$\hat{E}$ = number of emigrants as estimated;

$\hat{N}$ = net interprovincial migration as estimated;

$t-5,t$ indicates the five-year period during which the events occurred.

Table 1 presents the error of closure for the last four censuses for Canada, provinces and territories. On the whole, agreement between the census counts and the population estimates is fairly good even for provinces. This is all the more remarkable considering the fact that, in the absence of direct records, both emigration from Canada and interprovincial migration have to be estimated from administrative data (family allowance and income tax files).

Despite the high level of agreement, there are two salient features in the error of closure. One such feature is the jump to nearly one percent error of closure in 1986, a relatively large error when compared to that in the previous censuses. For the 1971 and 1976 censuses the error stood at slightly over one-half of one percent and only at one-quarter of one percent in 1981. The other feature is the negative error of closure in 1981. Whereas in the other three censuses, the estimates exceeded the census counts, in 1981 the former fell short of the latter. Almost all of this shortfall originated in the province of Alberta.

Turning to the provinces, one notes a consistently positive error of closure in 1986, whereas the sign of the error varied in the previous three censuses. Furthermore, for most of the provinces, the magnitude of the error has increased in 1986 as compared to the previous three censuses. The larger errors of closure were found in the Maritime Provinces and Quebec, and the smaller in Ontario and in the Western Provinces, with the exception of Saskatchewan.

The 1981 case of Alberta, referred to above, calls for some further remarks. In 1981, this province had to contend with an unusually large negative error of closure: the estimates fell short of the census count by 53,886 individuals or 2.41%. There are two possible explanations for this outcome. One is that the 1981 Census in this province may have suffered from a

Table 1

Error of Closure: Canada, Provinces and Territories, June 1971, 1976, 1981 and 1986

Geographic Area	Percent Error[1]			
	1971	1976	1981	1986
Canada	0.51	0.58	-0.25	0.95
Newfoundland	0.32	-0.19	1.25	2.02
Prince Edward Island	-0.76	1.58	-0.31	1.06
Nova Scotia	-2.45	0.93	-0.03	1.28
New Brunswick	-0.44	1.51	-0.28	1.57
Quebec	0.08	0.10	-0.58	1.34
Ontario	1.41	1.07	0.37	0.73
Manitoba	-0.01	1.21	0.83	0.57
Saskatchewan	0.21	0.91	-0.52	1.06
Alberta	0.31	-0.09	-2.41	0.81
British Columbia	0.47	0.07	-0.22	0.58
Yukon	-6.63	-2.34	-2.11	-4.66
Northwest Territories	3.14	-0.92	-5.60	-1.32

[1] $\frac{\text{Population Estimate - Census Count}}{\text{Census Count}} \times 100$

Source: Demography Division, Statistics Canada.

relatively large "overcount". Prompted by the booming oil-based economy, a great number of transient job-seekers from other provinces made their way to Alberta, some of whom may have been incorrectly enumerated as this province's usual residents. Yet, the fact that for 1981 Alberta showed an above-average undercount (2.54%) only adds to the puzzle. The other possible explanation is that the flow of in-migrants to Alberta, in those days of its economic prosperity and demographic boom, was not fully captured by the family allowance and taxation files - the basis of interprovincial migration estimates. In other words the large shortfall in the 1981 estimates of population might have resulted from an understatement of the net migration to Alberta.

Having demonstrated that the gap between estimates and counts widened significantly in 1986, the question to be addressed in the subsequent sections is whether this is due to the deterioration of: (a) the census coverage or (b) the data on the components of population growth over the last intercensal period.

3. DEMOGRAPHICALLY-DERIVED UNDERCOVERAGE RATE

By adjusting the census base population for undercoverage as estimated from the RRC, and by adding the net population increase (births, deaths and migrants) over the subsequent postcensal period, one obtains, as per equation 3, the population at the time of the next census. We shall call this the *expected* population, to differentiate it from the *estimated* and *enumerated* populations dealt with in the previous section.

$$P'_t = \left[R_{t-5} + \hat{U}_{t-5}\right] + \hat{G}_{t-5,t}, \tag{3}$$

where:

P'_t = expected population at time t;

R_{t-5} = enumerated population at time $t-5$;

$\hat{U}_{t-5}$ = the number of individuals missed in the census $t-5$, as estimated through the Reverse Record Check (RRC);

$\hat{G}_{t-5}$ = estimates of net population change over the intercensal period $t-5,t$ (births, deaths and migrants in equation (1)).

The difference, U'_t, between the *expected population*, P'_t, and the *enumerated population*, R_t, as per equation 4, can be taken here as a coverage error. We shall call this the *demographic* estimate of coverage error.

$$U'_t = P'_t - R_t. \tag{4}$$

And the rate of coverage error, u'_t, is simply the ratio of the demographically estimated error of coverage, U'_t, to the expected population, P'_t:

$$u'_t = \frac{P'_t - R_t}{P'_t} = \frac{U'_t}{P'_t}. \tag{5}$$

For comparison, the undercoverage rate as estimated through the RRC stands as follows:

$$\hat{u}_t = \frac{\hat{U}_t}{R_t + \hat{U}_t}. \tag{6}$$

How do the demographically estimated error of coverage and the RRC-estimated undercoverage compare? First, it should be stressed that both are subject to error and bias. The former is affected by: (a) the lack of an estimate of overcoverage; (b) the biases in the RRC-based undercoverage $\hat{U}$ at t and $t-5$ censuses, and; (c) the biases involved in the estimates of intercensal net population change $\hat{G}_{t-5,t}$, particularly its migration component. The RRC estimate of undercoverage is affected by: (a) sampling error, and; (b) various biases due to tracing of individuals, record matching, *etc.* Furthermore the undercoverage rate, $\hat{u}$, as per formula (6), is slightly downwardly biased because R_t in the denominator includes an overcount of unknown quantity. Hence, alone on these grounds, comparison between the two coverage measurements is far from being straightforward.

But there are conceptual differences as well. The RRC estimate is a pure undercoverage measurement. Demographically estimated coverage error is a more complex, difficult to define unequivocally, entity. It is neither an undercoverage nor a net undercoverage. In order, to better grasp the relationship between the two, the equation (3) of the expected population, P'_t, may be rewritten as per (7). Note that the enumerated population, R, is now expressed in terms of its two components: those who were correctly enumerated, R', and those who were overcounted, O.

$$P'_t = \left[(R'_{t-5} + O_{t-5}) + \hat{U}_{t-5} \right] + \hat{G}_{t-5,t}. \tag{7}$$

The undercoverage rate estimated by the demographic method as expressed in equation (5) now becomes:

$$u'_t = \frac{[(R'_{t-5} + O_{t-5}) + \hat{U}_{t-5} + \hat{G}_{t-5,t}] - (R'_t + O_t)}{(R'_{t-5} + O_{t-5}) + \hat{U}_{t-5} + \hat{G}_{t-5,t}}. \quad (8)$$

It follows from (8) that the overcoverage affects both the expected and the enumerated populations. Consequently, the demographic rate of undercoverage reflects the combined effect of the undercoverage per se and the difference in the overcoverage, 0, of the base census, $t-5$, and terminal census, t. Assuming that both (a) the RRC-based undercoverage, $\hat{U}$ at t and $t-5$, and (b) the population change (the net sum of the components) for intercensal period, $\hat{G}_{t-5,t}$, are correctly estimated, then the demographic coverage rate, u'_t, and the RRC rate, $\hat{u}_t$, will vary numerically depending on the level of the overcoverage of censuses at time, $t-5$ and t, so that if $O_t \gtrless O_{t-5}$ then $\hat{u}_t \lessgtr u'_t$.

Having clarified the conceptual particularities of the two measures of coverage error, we now turn to Table 2 which presents for Canada the coverage estimates for the 1981 and 1986 censuses. Both estimates reveal a significant increase in the coverage error in the 1986 Census. However, the demographically-derived rate of coverage error is consistently lower than the RRC rate of undercoverage: 2.82% and to 3.21% for 1986, and 1.70% and 2.01%, for 1981, respectively. This could mean that the overcoverage was higher in 1981 than in 1976, and higher in 1986 than in 1981, on the condition that the assumptions underlying the identities are correct. But there are no data to either confirm or deny the validity of these assumptions.

The estimates of coverage error by the two methods - demographic and RRC - by province in Table 2 are portrayed by Figure 1(a) and 1(b). The explanation of the differences at the provincial level is liable to present even greater uncertainties because the error and biases,

Table 2

Demographic and Reverse Record Check Estimates of Undercoverage Rates: By Provinces, 1981 and 1986

Geographic Area	Demographic Method		Reverse Record Check[1]			
			1981		1986	
	1981 (%)	1986 (%)	(%)	(%)	(%)	(%)
Canada (Territories not included)	1.70	2.82	2.01	(0.09)	3.21	(0.12)
Newfoundland	2.29	3.60	1.74	(0.95)	2.01	(0.32)
Prince Edward Island	0.05	2.10	1.17	(0.54)	2.16	(0.80)
Nova Scotia	0.82	2.22	1.05	(0.34)	2.63	(0.38)
New Brunswick	1.83	3.28	1.81	(0.30)	2.83	(0.36)
Quebec	2.31	3.13	1.91	(0.21)	3.06	(0.29)
Ontario	1.81	2.53	1.94	(0.14)	3.40	(0.19)
Manitoba	1.88	1.44	0.98	(0.35)	2.22	(0.40)
Saskatchewan	0.76	2.00	0.99	(0.37)	2.51	(0.36)
Alberta	-1.18	3.09	2.54	(0.36)	2.75	(0.33)
British Columbia	2.62	3.55	3.16	(0.33)	4.49	(0.39)

[1] Figures in brackets are Standard Deviations.
Source: Demography Division, Statistics Canada.

referred to above, at these levels are expected to be larger than they are at the national level. This is true in particular for sampling error in the case of the RRC undercoverage estimates, and for the biases in the interprovincial migration affecting net intercensal population change in the case of the demographic estimates of coverage error.

With the above comments regarding the biases and conceptual differences in mind, let us see how consistent are the two coverage measures at the provincial level? To this end, the following criterion of consistency is posited: if the two measures of coverage were conceptually identical and empirically correct, their respective correlation points in space should line up along the 45° bisectrix.

For the 1981 Census, disregarding the special case of Alberta referred to earlier (and also P.E.I. heavily affected by the sampling error), the correlation points follow closely the theoretical 45° straight line. The discrepancies are small: in most cases they are not statistically significant given the standard deviation affecting the RRC estimates (see Table 2).

For the 1986 Census, six provinces out of ten (Saskatchewan, Nova Scotia, Prince Edward Island, Quebec, Alberta and New Brunswick) have their respective points falling within close range of the 45° bisectrix and thus meet the consistency test. One, Newfoundland, falls far afield on the left side, suggesting a possible understatement of the RRC undercoverage rate for this province. Manitoba, Ontario and British Columbia fall well to the right side of the 45° bisectrix suggesting a possible overstatement of the RRC undercoverage or understatement of demographic coverage rate.

It should be stressed once again that the analysis of the accuracy of census coverage has been hampered by the lack of information on overcoverage. Yet, it is fair to say that notwithstanding its limitations, the analysis strongly points to a deterioration of the 1986 census coverage.

4. CENSUS AND COMPONENT-BASED INTERCENSAL POPULATION CHANGE: A CHECK FOR CONSISTENCY

The task now at hand is to compare two sets of *independent* estimates of the intercensal net population change: one set based on demographic components (births, deaths and migration), the other set derived from two consecutive censuses, unadjusted and adjusted for undercoverage. Refer to the former as *component-based estimates* and to the latter as *census-based estimates* of intercensal net population change.

$$\hat{G}_{t-5,t} = B_{t-5,t} - D_{t-5,t} + I_{t-5,t} - \hat{E}_{t-5,t} + \hat{N}_{t-5,t} \tag{9}$$

$$\bar{G}_{t-5,t} = R_t - R_{t-5} \tag{10}$$

$$\bar{G}'_{t-5,t} = (R_t + \hat{U}_t) - (R_{t-5} + \hat{U}_{t-5}). \tag{11}$$

All the above notations have been made explicit in the previous formulae.

Two independently-produced estimates might be construed as reasonably trustworthy if they are similar for a given point in time. As seen in Table 3, the difference between census-based and component-based estimates is only about 5% for the 1976-81 period. For the 1981-86 period, the two estimates differ by a substantial margin of 19% if unadjusted, and by 8% if adjusted for undercoverage.

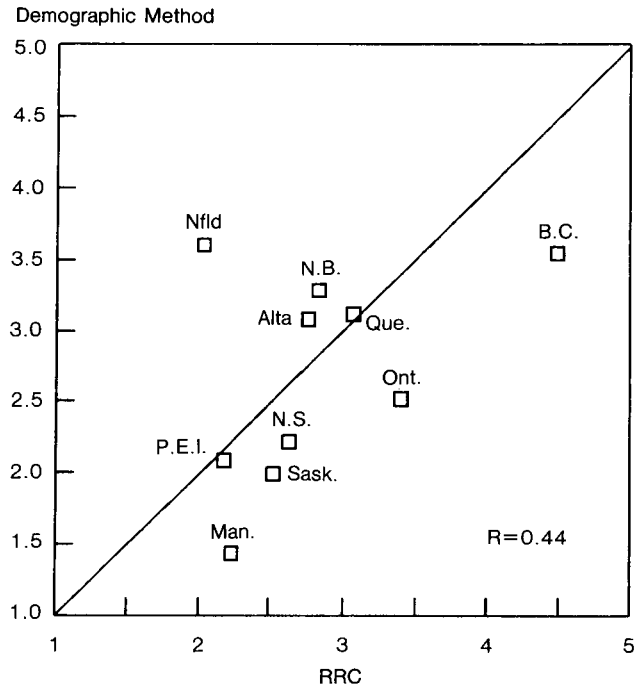

Figure 1. Relationship between Undercoverage Rates as Estimated by Reverse Record Check and Demographic Method, 1986 Census

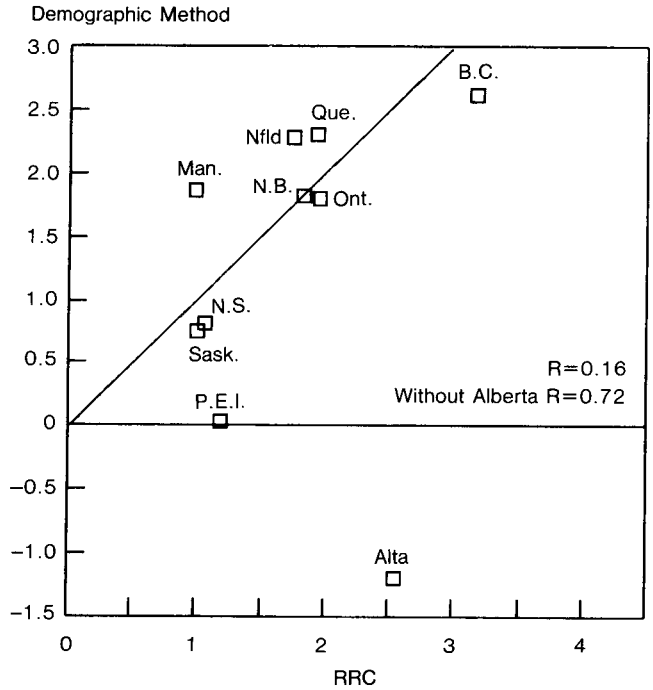

Figure 2. Relationship between Undercoverage Rates as Estimated by Reverse Record Check and Demographic Method, 1981 Census

Table 3

Ratio Between Census and Component-Based Intercensal Change in Population: By Province, 1976-81 and 1981-86

Geographic Area	Ratio between Census and Component-based Intercensal Population Change Multiplied by 100			
	1981-86		1976-81	
	Not adjusted for Census Undercoverage	Adjusted for Census Undercoverage	Not adjusted for Census Undercoverage	Adjusted for Census Undercoverage
Canada	80.9	108.4	104.5	106.1
(Territories not included)				
Newfoundland	5.7	19.6	58.3	80.9
Prince Edward Island	76.7	101.6	109.8	135.7
Nova Scotia	70.4	110.3	101.3	111.1
New Brunswick	55.6	86.7	111.3	99.2
Quebec	54.2	97.1	122.7	83.8
Ontario	88.1	115.2	92.0	103.2
Manitoba	89.2	117.3	35.6	29.0
Saskatchewan	79.4	110.3	112.0	105.5
Alberta	88.8	94.5	115.6	124.4
British Columbia	89.5	118.3	102.2	105.8

Note: The procedure cannot be applied for the period 1971-76 because, for this and earlier periods, emigration has been estimated residually from the two consecutive censuses and the remaining growth components (births, deaths and immigrants).
Source: Demography Division, Statistics Canada.

The comparison by province is a more delicate matter. On the components side, one has to contend with the reliability of the interprovincial migration estimates. On the census side, one must reckon with the variability of biases in undercoverage and overcoverage, and sampling errors in the RRC undercoverage estimates. Sampling errors alone could account for up to 15% of variations in the ratio between the two estimates of the intercensal population change for some provinces. Any variations beyond this level are more likely to have been induced by errors and biases from other than the sampling.

Hence, in the absence of a more trustworthy criterion, we have set $\pm 15\%$ as a tolerance limit for the discrepancies between the two estimates. The tolerance limit thus set, has at least the merit of screening out highly questionable cases.

With these qualifications in mind, let's turn to Table 3, which compares by province, census and component-based population changes for the last two intercensal periods. Six provinces out of ten for the 1976-81 period, and four out of ten for the 1981-86 period meet the somewhat arbitrarily set tolerance test. In general, the discrepancies are wider for the 1981-86 period than for 1976-81. Particularly conspicuous in this regard are the provinces of Newfoundland, Quebec, and New Brunswick.

Newfoundland's census-based 1981-86 population change represents only 5% of that derived from the components. It is still only 19%, even after adjustment for undercoverage. Such a low population growth would call for a net migration loss of about 26,000 over the 5-year period. Yet, all the three sources of interprovincial migration (Family Allowance, Taxation and the census mobility question) place these losses in the range of 14,800 to 16,500 (see Table 5).

Similar inconsistencies are found in the case of Quebec. The census-based population growth for the period 1981-86, which represents only 64% of the component-based growth, would imply Quebec's loss through out-migration to be twice the amount estimated by Statistics Canada, that is, 160,000 instead of 80,000. Yet again, all the three sources of information put the net-migration losses in the range of 63,000 to 81,000 over the 5-year period. The gap between the two estimates of intercensal change is almost wiped out when the 1981 and 1986 census counts are adjusted for undercoverage.

The case of New Brunswick is similar to that in Quebec and Newfoundland. The census-based estimate of population growth for the 1981-86 period suggests a net loss through out-migration of 11,200, whereas the family allowance-based figure is 2,200. The census mobility question and taxation figures are even lower, 1,376 and 65, respectively. Adjustment for undercoverage would bring New Brunswick's two estimates of the intercensal population change well within the tolerance limit.

What, then, can be concluded from the above analysis regarding the intercensal population change? It appears that both the components and the census generate reasonably consistent estimates of population change for the 1976-81 period. The discrepancies are small, within a tolerable limit for Canada and for most of the provinces. This, however, is not the case for the most recent intercensal period, 1981-86. Something seems to have deteriorated and the question remains as to whether it is the census or the components of population growth. As was seen in the preceding section, the 1986 Census experienced a significant increase in undercoverage estimated by two different methods. Adjustment for undercoverage, however, did not always produce better estimates of intercensal population growth, in fact the opposite happened in some cases. In the next section, we take a closer look at the components of population growth.

5. HOW GOOD ARE THE COMPONENTS OF POPULATION GROWTH?

What follows is a brief assessment of the quality of the data on births, deaths, immigration, emigration, and interprovincial migration. For a more complete account of the data on those components, and methodologies for estimating migration, the reader is referred to the 1987 Statistics Canada publication "Population Estimation Methods, Canada".

The registration of births and deaths is deemed to be complete in this country. Deaths or births that somehow escape registration must be by necessity very small in number in view of the prevailing regulations (need for a burial certificate) and the material (family allowance) incentives and legal requirements for registering births. Some late registration may occur, but the numbers are small. For the 1981-85 period, 3,831 or 0.02% of all births and 2,528, or 0.03%, of all deaths were registered beyond the cut-off date. This makes a net of only 1,303 persons unaccounted for in the population estimates.

Immigration statistics are regarded as reasonably accurate to the extent one speaks here of landed immigrants. The distribution of immigrants by province is based on their intended destination rather than on where they actually settle. It is, however, noteworthy, as per Table 4, that this distribution closely agrees with the 1986 Census distribution of immigrants.

Compared to the three other components reviewed above - births, deaths and immigration - interprovincial migration and emigration are weaker links in equation (1) which is used for estimating population for postcensal years. There are indeed no direct records of internal migration or emigration. Such figures must be estimated indirectly from administrative files

Table 4

Percentage Distribution of Immigrants by Province Based on the 1981 Census and Immigration Records of Intended Destination in 1980

Geographic Area	Immigration Records	Census
Newfoundland	0.4	0.3
Prince Edward Island	0.1	0.1
Nova Scotia	1.1	1.0
New Brunswick	0.8	0.8
Quebec	15.7	15.0
Ontario	43.5	42.7
Manitoba	5.4	5.4
Saskatchewan	2.5	2.6
Alberta	13.2	14.5
British Columbia + Yukon + Northwest Territories	17.2	17.6
Canada	100.0	100.0

Source: Demography Division, Statistics Canada.

- family allowance and income tax - which contain information on changes of residence. They deserve, therefore, more than a cursory consideration. In what follows, we shall focus on the significant methodological and data improvements achieved in recent years, as well as address certain persistent shortcomings inherent to these estimates. For a more complete account see Chapters IV and V of the Population Estimation Methods, Canada, 1987.

While family allowance data have been used since 1956, the most significant innovation to the system for estimating interprovincial migration was the addition of personal income tax data in 1976. As of 1981, a "two-track" estimation system was implemented: the *preliminary* quarterly and annual estimates based on family allowance data, and the *final* annual estimates based on taxation data. Both these data sources have strengths and weaknesses.

The main advantage of the family allowance file lies in its timeliness and fairly high accuracy. The information on change of address is available two months after the fact. The accuracy of the file is contingent upon two factors. The first is the comprehensiveness of coverage of child population, as every child under 18 years of age, supported by a parent, is entitled to a monthly payment. The second is the financial incentive for the beneficiaries of family allowances to report any change of address as soon as it occurs. The family allowance file does not, however, provide information on adult migration. This has to be estimated indirectly, by applying a conversion factor, *"f"*, which is obtained by calculating the ratio of the adult migration rate to the child migration rate from the taxation data available for the most recent year.

Given the key importance of the f factor in the estimation formulae, a few comments are called for. Prior to 1971, the value of f was based on 5-year migration data from the most recent census. As the annual age-specific data on migrants became available from income tax records, the decision was made to use such data since they have an advantage over census data in that they reflect a more recent age pattern of migration.

Another innovation is worth mentioning. Prior to 1981, the f factor was calculated only by province of origin. However, with the availability of relevant data from taxation, it became evident that this factor also varies significantly by province of destination. Consequently, the decision was made to calculate the f factor by both province of origin and province of destination.

Turning now to the personal income tax file as the data source for estimating interprovincial migration, the following assessment is in order. As compared to the family allowance file, the taxation file has the advantage of having a much broader demographic base: tax filers and their dependents represent roughly 90% of the population. However, there are various sources of potential errors and biases. Information on tax filers' dependents must be imputed from the dollar value of total exemptions. Various assumptions have to be made in imputing the migratory status of the tax filers' dependents, as well as that of persons who are neither filing income tax returns, nor are dependents upon those who do so, and therefore are not covered at all by the taxation system. This is particularly the case for young adults and the elderly, who may be more prone to neglect to file their tax-return or who may not earn the minimum income required for filing. Such differential age-related biases, if indeed present, affect the estimates of the age structure, and this in turn affects the value of the *f* factor, used in the family allowance-based preliminary estimates of interprovincial migration.

Table 5 presents figures on net interprovincial migration for the intercensal 1981-86 period based on family allowance, taxation, and the census question on residence five years ago. Notwithstanding some significant variations in numbers, the three sources of data provide a consistent picture of level of interprovincial net migration over the 5-year period, by province.

What has been said about interprovincial migration also holds for emigration - Canadians taking residence in another country. Prior to 1981, the aggregate emigration to countries other than the United States and the U.K. (for which data were available through the immigration services of the two countries) had to be estimated residually from consecutive censuses and the components of intercensal population growth. As of 1981, the estimation of the number of emigrants has been based on family allowance and income tax data. The procedure is similar to that described above for estimating interprovincial migration. Child-migration is estimated from family allowance data. To estimate adult emigration, and hence total emigration, a conversion factor, *f*, based on income tax data, is applied to child-emigration. This same procedure applies to both the preliminary and final estimates of emigration, except that in the latter case more complete data are used.

Table 5

Net Interprovincial Migration for the Period 1981-1986, Based on Specified Sources

Geographic Area	1986 Census[1]	Family Allowance	Income Tax
Canada	0	0	0
Newfoundland	-16,550	-14,837	-15,051
Prince Edward Island	1,540	293	751
Nova Scotia	6,275	5,204	6,895
New Brunswick	-1,370	-2,239	-65
Quebec	-63,295	-76,040	-81,254
Ontario	99,355	115,497	121,767
Manitoba	-1,555	-3,700	-2,634
Saskatchewan	-2,820	-668	-2,974
Alberta	-27,665	-34,073	-31,676
British Columbia	9,500	13,289	7,382
Yukon	-2,665	-2,381	-2,775
Northwest Territories	-755	-345	-366

[1] Population of 5 years and over.
Source: Demography Division, Statistics Canada.

Table 6

Estimates of Emigrants by Different Methods, Canada, 1981-86

Method	1981-86
Residual Method from Censuses	
(a) Unadjusted for Undercoverage	476,406
(b) Adjusted for undercoverage	134,807
Revenue Canada Tax File	165,272
Family Allowance Method (current) (using the *f* factor from the tax file)	235,481
Family Allowance Method (proposed) (using the *f* factor from the immigration file)	275,762
Reverse Record Check[1]	288,376

[1] Preliminary.
Source: Demography Division, Statistics Canada.

Table 6 compares, for the 1981-86 intercensal period, the estimates of emigration based on the family allowance files with the estimates produced by the various alternative methods. Note that the residually-derived emigration estimates, whether from adjusted or unadjusted census counts, are out of line with the more plausible estimates derived from the administrative files and the Reverse Record Check (RRC).

In brief, significant enhancements have been made to the system used to estimate interprovincial migration and emigration, particularly since 1981. While it can be surmised that the overall quality of the estimates has improved as a result, no demonstrable proof can be adduced. The family allowance and income tax data are fraught with various shortcomings inherent in any data system that has been designed for administrative rather than for statistical purposes.

6. CONCLUSIONS AND EMERGING ISSUES

Statistics Canada's population estimation system rests on two building blocks: (1) Census population counts, and; (2) components of population change, namely births, deaths and migrants. Postcensal estimates are carried forward by adding the components of population change over the subsequent years, to the base population, provided by the census. They are revised retrospectively when the next census counts become available. Thus, the census counts are both the base for the postcensal estimates, and the standard for their post-facto validation. The system has produced timely, reliable and internally consistent population estimates, and over the years has enjoyed a remarkable stability.

Much of its stability can be attributed to the high quality of the Canadian censuses. For Canada as a whole, undercoverage as measured by the Reverse Record Check (RRC) remained almost unchanged, at close to 2%, for three consecutive censuses - 1971, 1976 and 1981. Hence, even if the census fell somewhat short of the "true" population of Canada, it provided a highly reliable basis for gauging population growth.

The 1986 Census marks, however, a departure from the trend, as the rate of undercoverage, estimated by the Reverse Record Check, rose to 3.2%. The 1986 Census understates the population increase over the 1981-86 period by about 20%, if one accepts the component method as the standard of validation. Both the Reverse Record Check and the demographic analysis corroborate the deterioration of census coverage in 1986.

On the population components side of the equation - the other building blocks of the estimation system - records on births, deaths, and landed immigrants are fairly reliable. The interprovincial migration and emigration estimates have benefited from various data and

methodological enhancements, particularly since 1981, as was explained in the preceding section. But, as was also pointed out, they may suffer from various shortcomings inherent in any data sources - such as the family allowance and taxation files - that have been designed for administrative rather than statistical purposes. The estimates of interprovincial migration and emigration remain, along with census undercoverage and overcoverage, the prime sources of possible errors and biases in the postcensal estimates of population by province.

What does the future hold for the estimation system as described above? Can it continue working as it stands, or does it need some major reconceptualization? The apparently higher undercoverage rates of the 1986 Census, and its potential consequences for population estimates, has prompted the discussion of an alternative to the present census-based method of producing estimates. This alternative would no longer necessarily rely on the most recent census as a bench-mark, but instead would use relevant available information, including census counts, undercoverage and overcoverage, as well as administrative records, to generate the "best" possible estimates. In other words, the census counts remain an important ingredient of the estimation process, but not the overriding one; nor would the most recent census necessarily be used, if, say, the counts from the previous census were deemed to be more reliable.

After careful consideration, Statistics Canada has decided that the 1986 Census (unadjusted for undercoverage) would be used for the 1986 postcensal estimates and revision of the estimates for the 1981-86 intercensal period. In other words, the existing estimation procedures were reconfirmed. But at the same time, it was recognized that the evaluation of the census and estimates needed to be stepped up, and that an estimation strategy for the post-1991 Census period needed to be devised. Such an estimation strategy would have to take into account plans and realistic prospects for improvements and enhancements in the following four areas:

(1) 1991 Census coverage;

(2) Measurement of both undercoverage and overcoverage;

(3) Administrative records used for the purpose of population statistics: enhancement of the currently used sources - Family Allowance and Taxation - and the harnessing of new ones, such as Old Age Security and Provincial Health Care Files;

(4) Estimates of migration, particularly those concerning interprovincial migration, returning Canadian residents after a protracted stay abroad, and emigration from Canada.

These raise some fundamental issues concerning the philosophy and policy that ought to govern the working of a statistical system, thus transcending the rather narrow question of adjustment for undercoverage referred to at the outset of this paper. In the census-based conception, the emphasis is on the stability and internal coherence of the estimation system. In the conception of a census-divorced estimation model, a premium is placed on flexibility so as to increase the accuracy of the estimates through the utilization of the relevant available information, but possibly at the price of methodological consistency over time. The resolution of the dilemma between these two conceptions will be greatly influenced by the progress that is achieved in the four areas of statistical endeavour identified above.

ACKNOWLEDGEMENTS

The author has benefited from discussions with his colleagues in the Demography Division: Gwenaël Cartier, Gilbert Lagrange, Ronald Raby, Robert Riordan, Edward Shin and Ravi Verma. Valuable comments received from an anonymous referee and K.G. Basavarajappa,

David Binder, Malcolm Britton, Dick Carter, Ivan Fellegi and M.P. Singh, are also gratefully acknowledged.

REFERENCES

CARTER, R.G. (1988). Measuring coverage errors in the census population. Presented at the annual meeting of the Canadian Population Society, the University of Windsor, Windsor, Ontario.

COALE, A.J. (1955). The population of the United States in 1950 classified by age, sex and color. *Journal of the American Statistical Association*, 50, 16-54.

FAY, R.E., PASSEL, J.S., ROBINSON, G.J., and CONRAN, C.D. (1988). The coverage of population in the 1980 Census. U.S. Department of Commerce, Bureau of the Census.

FELLIGI, I.P. (1980). Should the census count be adjusted for allocation purposes - equity considerations. *Proceedings of the* 1980 *Conference on Census Undercount*, U.S. *Bureau of the Census*, 193-203.

FREEDMAN, D.A., and NAVIDI, W.C. (1986). Regression models for adjusting the 1980 Census, *Statistical Science*, 1, 3-39.

KEYFITZ, N. (1979). Information and allocation: Two uses of the 1980 Census, *The American Statistician*, 33, 45-50.

KEYFITZ, N. (1980). Issues in adjusting for the 1980 Census undercount. Paper presented at the Annual Meeting of the American Statistical Association, Detroit.

KISH, L. (1980). Diverse adjustments for missing data. *Proceedings of the* 1980 *Conference on Census Undercount, U.S. Bureau of the Census*, 193-203.

LAPIERRE-ADAMCYK, E. (1970). Estimation of net census underenumeration by age and sex using demographic analysis techniques. Working paper, Demographic Analysis and Research Section, Statistics Canada.

ROMANIUC, A., and RABY, R. (1980). The impact of census enumeration on selected Federal/Provincial transfer payments. Demography Division, Statistics Canada.

SPENCER, B. (1980). Issues of accuracy and equity in adjusting for census undercoverage. Paper presented at the Annual Meeting of the American Statistical Association, Detroit.

STATISTICS CANADA (1987). *Population estimation methods, Canada.* Catalogue No. 91-528E, Statistics Canada.

STATISTICS CANADA (1988). Undercoverage rates from the 1986 Reverse Record Check. User Information Bulletin, No. 2, Otttawa.

STOTO, M.A. (1987). Statement to the Subcommittee on Census and Population, Committee on Post Office and Civil Service, U.S. House of Representatives, San Francisco.

WILK, M.B. (1981). Letter to Mr. H. Breau, Chairman, Parliamentary Task Force on Federal-Provincial Fiscal Arrangements, July 3, 1981.