Canadian Studies in Population, Vol. 24(1), pp. 27-43

Census Coverage Evaluation and Demographic Analysis in Canada

Don Kerr
Denis Morissette
Demography Division
Statistics Canada

Abstract

Statistics Canada has produced population and family estimates, fully adjusted for coverage error, from 1971 to the present. Several procedures, developed by both demographers and statisticians have historically played an important role in evaluating census coverage in Canada. Most methods can be grouped as involving either (i) demographic analysis (macro-level approaches), or (ii) case by case matching procedures and record linkage techniques (micro-level approaches). The present paper reviews selected procedures, and suggests possible avenues for future research. It is argued that there is substantial potential in increasing the emphasis placed on demographic analysis in the estimation of census coverage, particularly in improving estimates for specific age and sex groups.

Résumé

Statistique Canada a produit des estimations de la population et des familles qui tiennent compte de l'erreur de couverture, de 1971 à aujourd'hui. Historiquement, plusieurs procédures élaborées conjointement par des démographes et des statisticians, ont joué un rôle important dans l'évaluation de la couverture des recensements au Canada. La plupart des méthodes peuvent être catégorisées selon qu'elles impliquent (i) une analyse démographique (approche macro) ou (ii) des procédures d'appariement cas par cas et des techniques de couplage d'enregistrements (approches micro). Le présent article examine certaines procédures et suggère des voies possibles pour la recherche à laquelle la valeur d'horizon est atteinte. Chaque composante produit venir. Il propose qu'il serait probablement opportun de faire une plus large place à l'analyse démographique dans l'estimation de la couverture des recensement -- pour améliorer les estimations relatives aux groupes d'âges ou de même sexe, en particulier.

Key Words: Census, Coverage Error, Undercount, Canada

Introduction:

Statistics Canada is expected to maintain highly accurate population figures. There are several reasons for this, not the least being the use of these figures in the formulae for calculating billions of dollars of revenue transfer payments between the federal and other levels of government. Yet at the same time, the enumeration of a population through a census is never complete, and is always hindered to a greater or lesser degree by coverage error. There are several reasons why an individual might be missed in the Canadian Census, and it is the responsibility of Statistics Canada to obtain a profile of such persons, contain this problem to a minimum, and if possible, adjust population figures for census undercount.

For the first time, following the 1991 Census, Statistics Canada decided to directly incorporate into its population estimation program, census counts adjusted for coverage error. Since this decision was made, Population Estimates Section of Demography Division continues to produce quarterly estimates of population, fully adjusted for census undercount (Statistics Canada, 1997). Similarly, Population Estimates Section has produced revised intercensal population and family estimates, fully adjusted for coverage error back to 1971 (Statistics Canada, 1994) and plans on directly adjusting the 1996 Census. Several procedures, developed by both demographers and statisticians have historically played an important role in evaluating the coverage of census data in Canada. The purpose of the present paper is to review procedures used for this purpose, and to suggest possible avenues for future research.

In estimation of the completeness of census coverage, most methods can be broadly grouped as involving either (i) demographic analysis (macro-level approaches), or (ii) case by case matching procedures and record linkage techniques (micro-level approaches). Demographic analysis begins with efforts to achieve highly precise population estimates independent of the census being evaluated (beginning for example, with data on births, deaths and migrants, derived independent of census). Population estimates can subsequently be compared with census counts, in order to obtain evidence as to how many persons are missed on census day. The crux of the matter with demographic analysis is whether or not estimates with sufficient precision can be obtained to provide valid insights as to the extent and nature of census undercount. This is a difficult problem, since the quality of administrative data and vital statistics as typically relied on by demographers in deriving population estimates are uncertain (particularly in moving back in time). Demographic analysis can be compared with micro-level approaches, as developed by statisticians, whereby case by case matching procedures and record linkage techniques are used. Such procedures systematically compare a sample of the population that should have been enumerated on census day to those as counted. In the 1991 Canadian Census, micro-level procedures provided Statistics Canada with its official adjustment for coverage errors, as based primarily on the Reverse Record Check Study (Statistics Canada, 1994). With the 1996 Census, the Reverse Record

Check will continue to be the primary vehicle for estimating undercount, nationally, and for each province and territory.

Studies on the completeness of census coverage essentially serve one or both of the following purposes: (i) to estimate and characterize persons missed by the census according to selected census characteristics (age, sex, marital status, home language, work status, urban/rural residence, etc.), and/or (ii) to form the adjustment necessary for the base population in official population estimates (requiring detailed information on persons missed by age, sex and place of residence). With respect to the first of the two purposes, micro-level /matching /record linkage techniques are typically the only source that potentially provides this sort of information (and this is true in Canada with the Reverse Record Check). With respect to the latter purpose, demographic analysis can potentially complement micro-level techniques.

The current paper focuses solely on the utility of these procedures in achieving the latter of the two purposes, i.e., in reviewing techniques relevant to maintaining high precision in official population estimates by age and sex. Overall, there appears to be substantial potential in increasing the emphasis placed on demographic analysis in the estimation of census coverage, particularly for specific age and sex groups.

Canadian Collection Procedures

With a goal of a *de jure* count of all dwellings, households and individuals in Canada, census representatives initially compile and verify exhaustive lists of dwellings within each of Canada's 46,000 enumeration areas. Within each enumeration area (approximately 220 households, on average) census representatives identify residential dwellings, classify dwellings as either private or collective, occupied or unoccupied. Within each occupied dwelling, all "usual" occupants are enumerated (primarily through self-reporting). In gathering this information, Canada's census involves two enumeration methods: mail back and canvasser.

Mail back enumeration is used in collecting information on most of Canada's population, by which census representatives drop off a questionnaire at each occupied or unoccupied dwelling with instructions to complete it (and mail it back) on census day. The canvasser method (which involves a personal interview) was reserved for a very small proportion of Canada's population (about 1%) who live in regions long considered difficult to be enumerated (i.e., in remote locations and a few difficult neighbourhoods in large cities). With both data collection methods, follow up procedures (by phone and in person) were obviously required to complete the enumeration (although initial non-response rates - at about 15% - are quite low by international standards - for example - at less than half the level observed in the U.S.). Several data quality checks are obviously introduced through all stages of data collection.

Micro-level Approaches to the Evaluation of Canadian Census Coverage

As reported in the 1991 Census Technical Report on Coverage Errors (Statistics Canada, 1994), there are four basic studies in the Coverage Error Measurement Program, which rely on tracing selected persons/record linkage/matching procedures: (1) the Vacancy Check, (2) the Temporary Residents Check, (3) the Overcoverage Study, and most important, (4) the Reverse Record Check (RRC). Briefly, the Vacancy Check involves a stratified multistage selection of enumeration areas, whereby a sample of dwellings are double checked to detect whether census representatives initially correctly classified dwellings as occupied or unoccupied. This study leads to adjustments of the census for households and individuals missed due to errors of census field representatives at the data collection stage. The Temporary Residents study involves a stratified sample of all persons who reported themselves as temporarily absent from the usual place of residence on Census day (using a special questionnaire designed for this purpose at hotels, airports, etc.), to determine if they were actually enumerated at their usual place of residence. This study leads to adjustment of the census for persons who for one reason or another were in transit during the actual enumeration process, and might have been missed at their usual place of residence.

The Reverse Record Check (RRC), the most important of the coverage studies, is a comprehensive record linkage method whereby a sample is selected assumed to represent the same target population as the census (i.e., from a supposed complete sampling frame), obtained in a manner fully independent of the census being evaluated. This includes in its sampling frame such sources as the enumerated population in the previous Census, birth registration data for all births over the intercensal period, administrative lists on recent immigrants and nonpermanent residents, and persons missed in the previous Reverse Record Check, among others. After selection of a sample in 1991 of about 56,000 persons (who should have been enumerated), a whole series of tracing, interviewing and searching procedures were introduced to determine the relative number and characteristics of persons "not enumerated." With respect to the Overcoverage Study, erroneous enumerations (double counting, false reporting, fabricated households, persons who died before census day) are detected, and involves both: (i) the sampling of persons reported residing in private and collective dwellings, in order to detect where persons listed on the census forms should have been enumerated (as well as obtaining information on possible alternate addresses to be systematically checked), and (ii) automated matching procedures on the census database, to identify whether any duplicate inclusions of households or persons occurred. The combination of the results from all of the above-mentioned studies provide official estimates of coverage error in the Canadian census.

The Reverse Record Check continues to be considered the most reliable source of information on undercoverage in Canada, both nationally and for the provinces. Table 1 provides estimates by age and sex of net undercount (estimated undercount - estimated overcount), as estimated nationally 1971-1991.[1] Net undercount has risen slightly over this period in Canada, from

1.88% in 1971 to 3.37% in 1991 among males, and 1.3% in 1971 to 2.28% in 1991 among females. With this increased coverage error, it is noteworthy that the age/sex pattern has been relatively stable, with low levels of net undercoverage throughout childhood, high levels for young adults (ages 20-24 and 25-29), and very low levels among Canadians in their latter forties and fifties. Similarly, undercount has consistently been higher among males than females, particularly among young adults.

Table 1.
Estimated Net Undercoverage Rate in Percentage of Population, by Age and Sex Group, Canada: 1971 - 1991 Censuses

Males	1971	1976	1981	1986	1991
all ages	1.88	2.06	1.95	3.18	3.37
0-4	1.24	1.61	1.16	1.80	2.26
5-14	0.82	1.21	0.92	1.57	1.86
15-19	1.92	1.34	2.27	3.12	2.70
20-24	4.24	5.17	5.17	9.00	7.74
25-34	3.13	3.39	2.49	5.01	6.76
35-44	2.32	2.04	2.98	2.87	3.19
45-54	1.43	1.23	0.83	1.43	1.53
55-64	1.26	1.19	0.59	1.55	1.37
65+	1.00	1.21	0.41	1.02	1.03
Females					
all ages	1.30	1.33	1.31	2.18	2.28
0-4	1.51	1.74	1.18	1.78	2.15
5-14	0.90	1.08	0.84	1.80	2.20
15-19	1.98	1.63	2.23	2.45	3.23
20-24	3.41	3.96	4.23	5.84	6.26
25-34	1.36	1.77	1.63	3.08	3.48
35-44	0.97	0.55	0.67	0.99	1.54
45-54	0.61	0.55	0.17	0.77	0.74
55-64	0.74	0.43	0.56	1.76	0.96
65+	0.79	0.43	0.48	1.42	1.05

Sources: Michalowski, 1993: Statistics Canada, 1994.

While the Reverse Record Check (along with related coverage studies) are widely accepted as the most reliable and valid source of information on undercount in Canada, various sources of error slightly lower their utility. It is not surprising that a primary source of error in the Reverse Record Check relates to its sampling error, which obviously increases proportionate to level of disaggregation. Furthermore, while Statistics Canada consistently publishes standard errors, the issue of non-sampling error has not been quantified. As an example, a significant percentage (4.9%) of the unweighted 1991 Reverse Record Check sample is actually "not traced" in the record matching procedure (i.e., the selected person has moved from the last known address and the person or a member of his household couldn't be traced using available information); therefore no interview was conducted to obtain a census day address and classify this person as "successfully enumerated" or "missed." These "not traced" records are then imputed (as either "enumerated, missed, deceased, or emigrated/abroad/out of scope") with roughly the same distribution as those who moved from their last known address but were successfully traced. An unknown level of bias is associated with such imputation. For a review of selected difficulties as associated with earlier versions of the Reverse Record Check methodology, see Burgess (1988).

Statistics Canada's coverage studies are understood as producing high quality estimates of net undercount, at both the national level (an estimate of 807,254 persons missed in 1991) and at the provincial level. Yet it is not surprising that the published sampling errors as associated with these coverage studies increase proportionate to the level of disaggregation, and are not always capable of producing highly precise estimates for specific age/sex groups. Furthermore, in the evaluation of census data by age and sex, it should also be appreciated that various other types of error beyond coverage error potentially impact on quality. For example, in the 1991 census database, more than 727,000 persons have imputed dates of birth and/or sex due to diverse collection, processing and data quality procedures. Not quantified here are respondent errors (proxy respondents), data capture errors and systematic processing errors that cannot be detected by the systems or certification analysis. Establishing the source of error in census data is therefore often difficult, if inconsistencies are detected in comparing census figures from one enumeration to the next.

Demographic Methods with Canadian Data

It is not surprising that selected inconsistencies surface in systematically inspecting census figures adjusted by the coverage studies. More specifically, while population figures are reasonable for Canada and the provinces, they do, however, become troublesome with respect to selected age groups. For example, in systematically examining census numbers (after adjustment by the coverage studies) for specific cohorts over time, such figures are not necessarily consistent with what demographers know of the mortality and migration experience of specific cohorts between censuses. It is in this context that demographic methods suggest that further improvements are possible, particularly with the goal of establishing intercensal consistency in adjusted

Canadian census counts, in achieving temporal coherence of population figures from one census to the next.

In the Canadian context, the utilities of demographic methods are twofold: (i) they are considered useful in the evaluation of the performance of the coverage studies, and (ii) in potentially providing more refined estimates of the Canadian population by specific age and sex categories. It is worth noting that past evaluation of Reverse Record Check results via demographic analysis has lead to revisions of Reverse Record Check based estimates of net undercoverage, for specific age/sex categories. For example, in examining the 1991 Reverse Record Check results among children aged 0-4, the estimated net undercount was significantly higher among female children (3.75%) than male (2.26%) - with no logical reason why this should have occurred beyond sampling error. Through demographic analysis, using highly precise information on the sex ratio at birth and mortality data from vital statistics, it was clear that the Reverse Record Check results when added to the census results were not possible for this age group (and consequently, the number of females was revised to correspond with the sex ratio aged 0-4 as derived via demographic analysis).[2] Similarly, demographic analysis has suggested specific difficulties in census adjusted counts for older age groups. As a result, estimates by five year age groups of net undercoverage among Canadians over 55 were generated by Demography Division (i.e., allocating the total Reverse Record Check undercount for these ages proportional to cohort size). Since the Reverse Record Check was specifically designed to estimate persons missed in the census, the sample size is not large enough to give accurate estimates for the older age groups (i.e., the sample design included a disproportional share of young adults). The estimated undercount as presented in the aforementioned Table 1 have been adjusted on the basis of demographic evaluation, and consequently differs from those previously published with the results of the 1991 coverage error measurement program (Statistics Canada, 1994).

Estimates of Net Undercount Obtained Entirely Independent of Census Operations

Using the basic demographic accounting equation, estimating Canada's population by cohort is possible, by beginning with each cohort at birth (size as determined via birth registration data), adding and subtracting all relevant data on the components of demographic change (deaths, immigration and emigration) through to the census year being evaluated. With Canadian data, a long time series is available, as historic components of demographic change have been collected and estimated back to 1921. With this time series, Bender (1992) has estimated by sex and five year age groups, the population of Canada corresponding to the last several censuses. These population estimates have been compared with unadjusted census data, to produce estimates of net census undercoverage (independent of previous censuses), nationally, by age and sex.

More specifically, the basic demographic accounting procedure relies upon the fundamental balancing equation: $C_t = B_y - D_{y,t} + I_{y,t} - E_{y,t}$, by which C_t denotes

estimated cohort size in year t, B_y represents the size of this cohort at its origin (the number of births in year y) and $D_{y,t}$, $I_{y,t}$, and $E_{y,t}$ designate the cumulative number of events (deaths, immigrants, and emigrants) affecting the size of this cohort since its origin, between the years y and t. As data on the components of demographic change are available since 1921, this demographic method can generate estimates of the age cohort 0-4 in 1926, cohorts 0-4 and 5-9 in 1931, cohorts 0-4, 5-9 and 10-14 in 1936, and so on. In applying this procedure to the estimation of the 1991 populations, the available time series allows for estimates of the population, aged 0-69.

It is not surprising that this technique suffers to the extent that there is error in the components. More specifically, this technique is hindered by errors in Statistics Canada's (i) birth registration data, (ii) registration data on deaths, by age and sex, (iii) estimates of immigration, emigration, non-permanent residents and returning Canadians (all by age and sex). Furthermore, this procedure faces obstacles to the extent that there is cumulative error in adding the demographic components up over time (as a direct function of the length of the time series involved in estimation). Fortunately, there is a consensus among Statistics Canada demographers that the two most fundamental components, i.e., births and deaths, are also those measured with the highest degree of precision (although there has never been a study of the completeness of birth and death registration in Canada).[3] On the other hand, it has also been recognized that the quality of estimates on international migration decline as we move back in time (particularly with respect to emigration and returning Canadians, by age and sex).

Due to the impact of various data quality problems, estimates of net undercount obtained with this technique for older ages are clearly unrealistic (for example, very high levels of overcoverage were estimated for Canadians born before 1951). This is not at all surprising since Demography Division has clearly greater confidence in its time series after 1951 (particularly with respect to migration - given the well-documented difficulties in the estimation of emigration from Canada). Therefore, it is possible that reasonable estimates may be possible for the youngest of ages - independent of census data - whereas with older ages, the cumulative effect of error in the components can only lead to unreliable estimates. Estimates of children and youth using this procedure are considered the most reliable, although an unknown level of bias remains.

Canadian Post Censal Estimates

Statistics Canada has for many decades used post-censal estimates in the evaluation of census results. For example, in the evaluation of the 1986 Census, Romaniuc (1988) used the component method in obtaining post-censal estimates, errors of closure, and corresponding estimates of undercoverage for the total national/provincial population counts. With post-censal estimates, births, immigrants, nonpermanent residents and returning Canadians are added to, and deaths and emigrants are subtracted from, the base census population (five years earlier in Canada). In the evaluation of the 1991 Census, Bender (1992) has

similarly relied upon this method, in estimating levels of net undercoverage by age and sex in 1981, 1986 and 1991. In so doing, both demonstrated the strengths and weaknesses of such a procedure, in deriving estimates of net undercount. Compared with the basic demographic accounting formula there is clearly less room for potential error in the components of demographic growth with this technique, relying exclusively upon the previous intercensal period. On the other hand, the base population plays a major role in the subsequent estimates, as the census base population is assumed to represent for each cohort the net effect of all components of demographic change prior to this date.

A comparison of 1991 direct census counts with post-censal estimates can allow for estimates of net undercount by age and sex, only if it is accepted that: (i) the 1986 base year was measured with complete accuracy, and (ii) that intercensal events by age and sex were measured and estimated without error. In the Canadian context, both assumptions are unrealistic to a greater or lesser extent. The first of the two assumptions is particularly problematic, as there is no evidence to suggest that error in 1986 Reverse Record Check adjusted counts be less than that in the 1991 Reverse Record Check (as the underlying Reverse Record Check methodologies were very similar in both census years). Overall, while this method provides a reasonable indication as to the level of intercensal incoherence in adjusted census counts over time, it provides little guidance as to how observed discrepancies are corrected.

Other Demographic Methods

Overall, there has not been a great deal of research using demographic techniques in the evaluation of census data, beyond that as mentioned above. Post-censal estimates (as an immediate extension of Statistics Canada's population estimation program), has long served in the evaluation of preliminary census results, and will continue to be used in this manner into the future. Some interesting exploratory research has also been enacted with previous censuses, in considering the relevance of administrative data in the evaluation process: for example, Medicare data, tax files, family allowance (for ages 0-14), old age security (for ages sixty-five and older) among others (Fortier and Raby, 1989; Michalowski, 1992). While this at first glance appears promising, particularly for older ages, much more research is necessary in examining where potential problems in the administrative files. The collection and editing of administrative data sets are typically the responsibility of each province, with varying resources and priorities in the maintenance of quality. As a result, the quality of selected data sets vary considerably by province.[4]

It is interesting to note that over 25 years ago, an interesting line of research was initiated by Lapierre-Adamcyk (1970), in an attempt to replicate with Canadian data selected demographic techniques initiated in the United States by Coale (1955). Ansley Coale has generally been considered as laying the cornerstone of demographic analysis and the evaluation of census coverage in the United States (Himes and Clogg, 1992). In applying an iterative procedure as developed by Coale, which was based on a "hypothesis of similar errors" in

coverage by age and sex across past censuses, Lapierre-Adamcyk provided alternate estimates by broad age groups of net undercount (for the "native born"). These estimates were then compared with some of the earliest estimates as available from the Reverse Record Check. A major obstacle that she faced in her research (and we continue to face in developing demographic estimates) is the uncertainty of data quality, as to the accuracy of vital statistics on births and deaths that entered her estimates. It is interesting to acknowledge the primary recommendation that came out of her research, most notably, the "necessity of studying the relative completeness of birth registration."

In recognition of the imperfect nature of data that enters demographic estimates, Dionne (1995) has suggested a series of further techniques that build upon the basic cohort component approach. In so doing, he has attempted to introduce a "multidimensional" character to subsequent estimates. By "multi-dimensional," Dionne refers to techniques under which various relationships between cohorts (of varying levels of undercount), by sex, are simultaneously taken into account when estimating the relative size of age cohorts in a specific census year. The hope in developing such models, is to establish population estimates that are relatively robust to potential data quality problems, which could then be systematically compared with unadjusted census figures to obtain estimates of undercount, by age and sex. The most comprehensive of these multidimensional methods is his intergenerational model - outlined in some detail in a recent report - with several empirical applications (Dionne and Kerr, forthcoming). Fundamental to the intergenerational model is the establishment of "descendance" and "ascendance" relations between generations, relying upon long time series on births (by age of mother/father), deaths, emigration and immigration (1921-1991). Further research continues as to the utility of this model in the evaluation (and potential correction) of the age/sex distribution of census data.

Future Research

While the Reverse Record Check is accepted as the most reliable and valid source of information on undercount in Canada, various types of error continue to weaken its utility. As previously indicated, a primary source of error relates to the sampling error of the Reverse Record Check methodology itself, which obviously increases proportionate to level of disaggregation. With this in mind, one primary advantage of demographic analysis is its potential in providing more refined estimates of population by specific age and sex categories.

In this context, it is noteworthy that the U.S. Census Bureau has long enacted demographic analysis in the estimation of census coverage, on a much larger scale than in Canada (Coale, 1955; Coale and Zelnik, 1963; Siegel and Zelnik, 1966; Fay, Passel and Robinson, 1988; Robinson, Ahmed, Das Gupta and Woodrow, 1993). Demographic methods have been relied upon for several decades in the evaluation of census coverage, and in the evaluation of the relevant components that enter demographic analysis (i.e., birth registration data, mortality data, international migration figures, etc.). In summarizing this

research, the Census Bureau has been involved in evaluating, adjusting (and justifying the adjustment) of relevant time series, all based on evidence as to the internal consistency of demographic data between and across relevant censuses. This ongoing research involves several demographers working in Population Division of the U.S. Bureau, and includes a strong tradition of collaboration between the Bureau researchers and academic demographers.

In maintaining confidence as to the quality of demographic estimates of net undercount, demographers need to be confident as to the quality of vital statistics and migration data that enter such estimates (and subsequent adjustments judged necessary). For example, research on the completeness of birth registration data has lead to systematic adjustments of the time series on births in the U.S. It is noteworthy that due to difficulties in using demographic techniques in the estimation of older age groups (i.e., cumulation of error with longer time series), the U.S. Bureau has for several censuses relied upon administrative data for ages sixty five and older (i.e., Medicare data with adjustments for under enrollment). Further research as to similar applications of administrative data sets in Canada appears advisable, particularly for older age groups.

A promising extension of this American research has recently been introduced through the development of what has come to be called an "uncertainty model" of demographic analysis (Das Gupta, 1991). Since demographic estimates are hindered by the uncertainty of underlying components, analysts at the U.S. Bureau decided to address this issue explicitly. The goal in developing this "uncertainty model" was to develop interval estimates of population (and net undercount), such that the probability was very high (95% or 99%) that the true U.S. population count fell within estimated intervals. As based on pseudo statistical theory, a model was developed which involved interval estimates somewhat analogous (yet qualitatively different from) conventional frequentist confidence intervals. In setting out to establish this "uncertainty model," the U.S. Bureau has compiled a series of reports that evaluate the uncertainty associated with the measurement of each individual component in their demographic estimates (Robinson, 1991a, 1991b, 1991c, 1991d; Woodrow, 1991a, 1991b; Robinson and Lapham, 1991; Robinson, Woodrow and Ahmed, 1991; Robinson, Ward and Spencer, 1991). In the evaluation of uncertainty, the goal was to develop interval estimates for each component, working with a "judgmental consensus" of Census Bureau experts knowledgeable about measurement error and estimation methodology.

Selected recommendations by Lapierre-Adamyck (1970) more than twenty five years ago seem relevant today, as further research is needed into the relative quality of the time series that enter demographic estimates - most notably, the accuracy of vital statistics on births and deaths and demographic estimates of international migration. This has long been a priority in the United States, and it is precisely in obtaining a quantitative fix on the validity of their demographic time series that the American research has been so innovative. Future research appears appropriate, as to the quality and uncertainty of relevant time series on demographic change, including the need for improved estimates on international

migration (current estimates on emigration and returning Canadians appear to be particularly weak).

Conclusion

Several procedures, developed by both demographers and statisticians, have historically played an important role in evaluating the coverage of census data. The purpose of the present report was to review selected procedures used for this purpose in Canada, and to make recommendations for future research. In so doing, both micro-level (record linkage/search) procedures and macro-level (demographic procedures) were briefly outlined. In concluding, it is suggested that (i) the available historical time series on births, immigration, emigration, returning Canadians, non-permanent residents, and deaths, be more thoroughly evaluated, (ii) the results obtained across different methodologies (the Reverse Record Check and demographic techniques) be more systematically appraised and compared, (iii) the utility of additional sources of administrative data be more systematically examined, particularly in evaluating the coverage of older age groups (including such sources as Medicare files, tax files, and social security), and (iv) further research into the potential impact of both sampling and non-sampling errors in the Reverse Record Check and Census edit and imputation procedures be enacted.

The importance of collaborative research on census coverage by both statisticians and demographers has been recognized in Canada, as elsewhere. At the very least, demographic methods can potentially serve useful in the evaluation of the Reverse Record Check and in providing alternate independent estimates of the population, by age and sex. Similar results across different methodologies add to a credibility and confidence of Statistics Canada's estimates of population. Given the relative costs of demographic analysis (as a regular extension of Statistics Canada's population estimation program), further resources appear justified in developing demographic estimates of census coverage error.

Disclaimer

The views expressed within are those of the authors and do not necessarily reflect those of Statistics Canada.

Endnotes:

1. Overcoverage was measured for the first time in 1991 through the aforementioned Overcoverage Study. With estimates of net undercount prior to 1971, estimates of overcoverage were provided separately by Demography Division (for details, see Michalowski, 1993).

2. The utility of sex ratios in the evaluation of RRC results should not be understated. Sex ratios are particularly robust tools in the evaluation of census data, and Statistics Canada will continue to apply them in the evaluation of census and coverage error study results, particularly with respect to younger age cohorts.

3. It should be acknowledged that a very limited study of birth registration completeness was enacted by Enid Charles (1940) in the 1930s, with exploratory research involving a few selected census tracts (with early evidence to suggest incompleteness in the order of 3%). The working assumption of analysis responsible for birth data in Canada has been that with the introduction of universal family allowance in the latter 1940s, a significant incentive to register births added to an already highly reliable registration system. Most recently, the universality of the family allowance program has been replaced by the child tax credit. It is uncertain as to whether this may have had an impact on the incentive to register births. Similarly, R. Maheu (1973) has also examined birth registration data for one province of Canada, in an attempt to estimate the extent of lateness in birth registration in Quebec.

4. Beyond using administrative records in the evaluation of Census data, Statistics Canada has actually explored the issue: as to whether administrative data sets might potentially be used as a "substitute" for census data. Over recent years, the Administrative Record Comparison Project (ARC) was specifically given the mandate to explore this possibility. In attempting to fulfill this mandate, ARC judged that they would most likely succeed in constructing an enumeration of the 1991 Canadian population, if they were to begin with tax files as presently available from Revenue Canada (with non-filers imputed). With this independently derived estimate of Canada's population, systematic comparisons were made with population counts for the 1991 Census. After disclosing significant discrepancies at all levels of geography, the expected recommendation surfaced, that being administrative records represent a data source complementary to the Census of Population and not a replacement for it (Standish, 1993). Further research is necessary, as present results (in comparing the census with tax files) at the national level are completely implausible. The obvious difficulty lies with the non-universality of most administrative data sets in Canada, and the subsequent imputation procedures that follow.

References:

Bender, R. 1992. *Demographic Estimates of 1991 Net Census Undercoverage.* Working Document. Ottawa: Demography Division, Statistics Canada.

Burgess, R. D. 1988. Evaluation of Reverse Record Check Estimates of Undercoverage in the Canadian Census of Population. *Survey Methodology* 14(2): 137-156.

Charles, E. 1940. *The Changing Size of Family in Canada.* Census Monograph No. 1. 1941. Ottawa: Dominion Bureau of the Statistics.

Coale, A. J. 1955. The Population of the United States in 1950 Classified by Age, Sex and Colour. *Journal of the American Statistical Association* 50(1): 16-54.

Coale, A. J. and M. Zelnik. 1963. *New Estimates of Fertility and Population in the United States: A Study of Annual White Births from 1855 to 1960 and of Completeness of Enumeration in the Censuses from 1880 to 1960.* Princeton, N. J.: Princeton University Press.

Das Gupta, P. 1991. *Demographic Analysis Evaluation Project D11: Models for Assessing Errors in Undercount Rates Based on Demographic Analysis. Preliminary Research and Evaluation,* Memorandum No. 84 (PREM). Washington: U.S. Bureau of the Census.

Dionne, C. 1995. Les relation intergénérationnelles et quelques modèles qui en découlent. *European Journal of Population* 11: 85-101.

Dionne, C. and D. Kerr. 1995. *Alternate Techniques for Evaluating Census Counts: Methodological Proposals and Applications.* Working paper. Ottawa: Demography Division, Statistics Canada.

Dionne, C. and D. Kerr., forthcoming. Intergenerational Relationships and Population Estimation. Demographic Document: *Current Demographic Analysis Series, No. 3.* Catalogue No. 91F0015MPE. Ottawa: Statistics Canada,.

Fay, R., J.S. Passel and J. G. Robinson. 1988. *The Coverage of Population in the 1980 Census.* Evaluation and Research Report, PHC 80-E4. Washington, D. C.: United States Bureau of the Census,

Fortier, C. and R. Raby. 1989. Évaluation de la qualité des données par age et sexe du recensement de 1986 au niveau national. *Cahier québécois de démographie* 18(2): 285-320.

Himes, C. L. and C. Clogg. 1992. An Overview of Demographic Analysis as a Method for Evaluating Census Coverage in the United States. *Population Index* 58: 587-607.

Lapierre-Adamcyk, E. 1970. *Estimation of Net Census Underenumeration by Age and Sex Using Demographic Analysis Techniques.* Working paper. Ottawa: Demographic Analysis and Research Section, Statistics Canada.

Maheu, R. 1973. Statistique des affaires sociales. *Démographie*, 1(3): 72-78.

Michalowski, M. 1992. *Preliminary Estimates of the 1991 Census Net Undercoverage Based on Administrative Sources of Data.* Working document. Ottawa: Demography Division, Statistics Canada.

Michalowski, M. 1993. *Revised Postcensal and Intercensal Estimates: Canada, Provinces and Territories, 1971-1993 - Methodology for Derivation of Annual Estimates of Population by Age and Sex.* Ottawa: Population Estimates Section, Demography Division, Statistics Canada.

Robinson, J. G. 1991a. *Demographic Analysis Evaluation Project D1: Error in the Birth Registration Completeness Estimates.* Preliminary Research and Evaluation Memorandum No. 74 (PREM). Washington, D. C.: U.S. Bureau of the Census.

Robinson, J. G. 1991b. *Demographic Analysis Evaluation Project D3: Uncertainty Intervals for Estimated White Births, 1915-1934.* Preliminary Research and Evaluation Memorandum No. 76 (PREM). Washington, D. C.: U.S. Bureau of the Census.

Robinson, J. G. 1991c. *Demographic Analysis Evaluation Project D4: Uncertainty Intervals for Estimated Black Births, 1915-1934.* Preliminary Research and Evaluation Memorandum No. 77 (PREM). Washington, D. C.: U.S. Bureau of the Census.

Robinson, J. G. 1991d. *Demographic Analysis Evaluation Project D6: Robustness of Estimates of the Population Aged 65 and Over.* Preliminary Research and Evaluation Memorandum No. 79 (PREM). Washington, D. C.: U.S. Bureau of the Census.

Robinson, J. G. , B. Ahmed, P. Das Gupta and K. A. Woodrow. 1993. Estimation of Population Coverage in the 1990 United States Census Based on Demographic Analysis. *Journal of the American Statistical Association* 88(423): 1061-1079.

Robinson, J. G. and S. Lapham. 1991. *Demographic Analysis Evaluation Project D9: Inconsistencies in Racial Classifications of the Demographic Estimates and the Census.* Preliminary Research and Evaluation Memorandum No. 82 (PREM). Washington, D. C.: U.S. Bureau of the Census.

Robinson, J. G., K. A. Woodrow and B. Ahmed. 1991. *Demographic Analysis Evaluation Project D7: Uncertainty Measure for Other Components.* Preliminary Research and Evaluation Memorandum No. 80 (PREM). Washington, D. C.: U.S. Bureau of the Census.

Robinson, J. G., D. L Ward and G. S. Spencer. 1991. *Demographic Analysis Evaluation Project D8: Uncertainty for Models to Translate 1990 Census Concepts into Historical Racial Classification.* Preliminary Research and Evaluation Memorandum No. 81 (PREM). Washington, D.C.: U.S. Bureau of the Census.

Romaniuc, A. 1988. A Demographic Approach to the Evaluation of the 1986 Census and the Estimates of Canada's Population. *Survey Methodology* 14(2): 157-172.

Siegel, J.S. and M. Zelnik. 1966. An Evaluation of Coverage in the 1960 Census of Population by Techniques of Demographic Analysis and by Composite Methods. *Proceedings of the Social Statistics Section, American Statistical Association* 71-85.

Standish, Linda. 1993 *Administrative Record Comparison (ARC): Report on Demographic Comparisons with the 1991 Canadian Census.* Working document. Ottawa: Statistics Canada.

Statistics Canada. 1994. *Coverage: 1991.* Technical Reports. Catalogue No. 92-341E. Ottawa: Statistics Canada.

Statistics Canada. 1994. *Revised Intercensal Population and Family Estimates, July 1, 1971-1991.* Catalogue No. 91-537, Occasional. Ottawa: Statistics Canada.

Statistics Canada. 1997. *Annual Demographic Statistics, 1996.* Catalogue No. 91-213, Annual. Ottawa: Statistics Canada.

Woodrow, K. A. 1991a. *Demographic Analysis Evaluation Project D2: Preliminary Estimates of Undocumented Residents in 1990,* Preliminary Research and Evaluation Memorandum No. 75 (PREM). Washington, D. C.: U.S. Bureau of the Census.

Woodrow, K. A., 1991b, *Demographic Analysis Evaluation Project D5: Preliminary Estimates of Emigration Component*, Preliminary Research and Evaluation Memorandum No. 78 (PREM). Washington, D. C.: U.S. Bureau of the Census.

Received December 1996; Revised October 1997