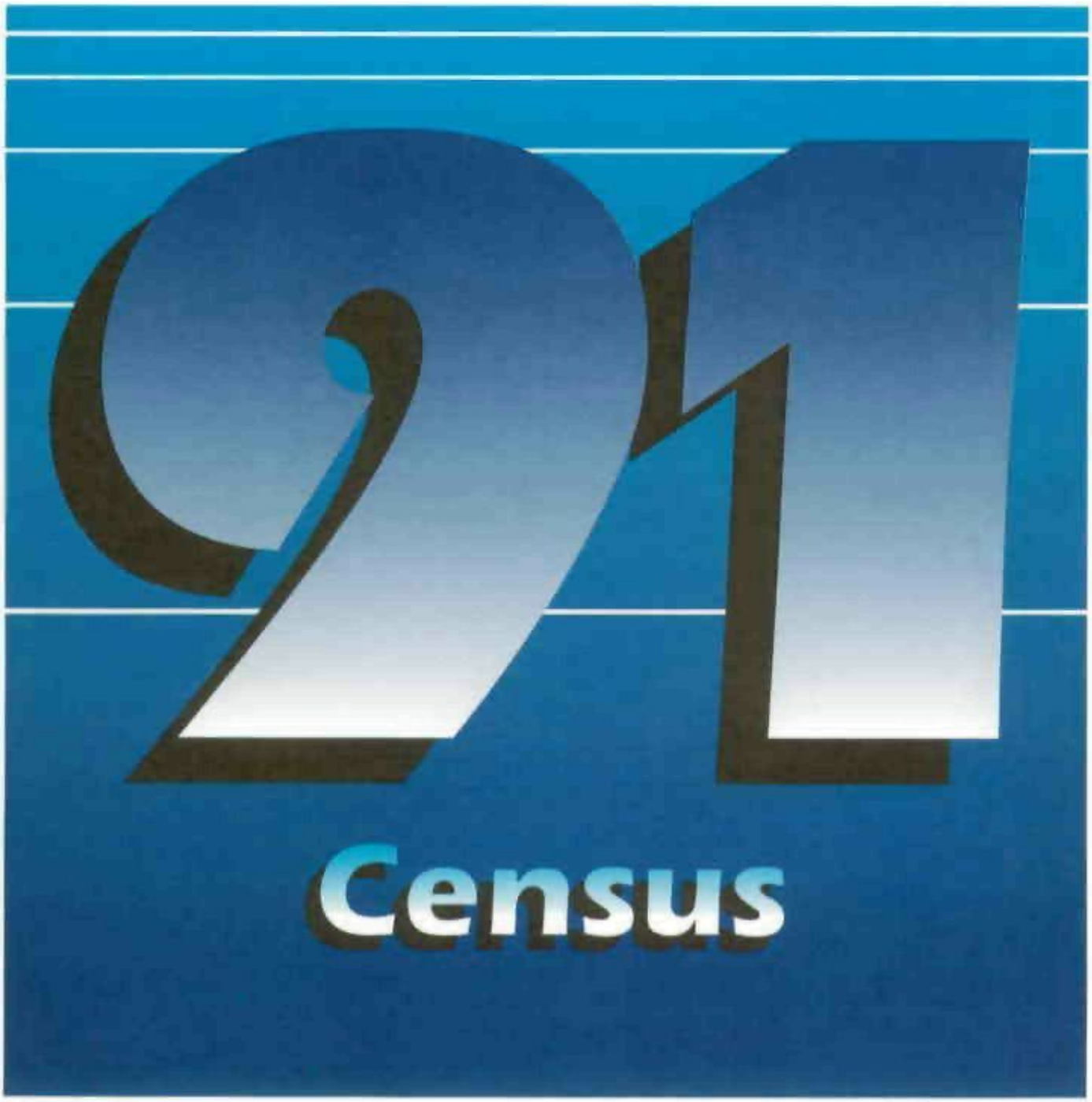
C-4

Catalogue 92-342E

Sampling and Weighting

1991 Census Technical Reports



Statistics Canada Statistique Canada

Canada

How to Reach Us

Data in Many Forms...

Statistics Canada disseminates data in a variety of forms. In addition to publications, both standard and special tabulations are offered. Data are available on CD, diskette, computer print-out, microfiche and microfilm, and magnetic tape. Maps and other geographic reference materials are available for some types of data. Direct on line access to aggregated information is possible through CANSIM, Statistics Canada's machine-readable data base and retrieval system.

How to Obtain More Information

Inquiries about products and services should be directed to your local Statistics Canada Reference Centre:

Halifax	(902 426-5331)	Regina	(306 780-5405)
Montréal	(514 283-5725)	Edmonton	(403 495-3027)
Ottawa	(613 951-8116)	Calgary	(403 292-6717)
Toronto	(416 973-6586)	Vancouver	(604 666-3691)
Winnipeg	(204 983-4020)		

Toll-free access is provided in all provinces and territories, **for users who reside outside the local dialing area** of any of the regional reference centres.

Newfoundland, Labrador, Nova Scotia, New Brunswick and Prince Edward Island	1-800-565-7192
Quebec	1-800-361-2831
Ontario	1-800-263-1136
Saskatchewan	1-800-667-7164
Manitoba	1-800-661-7828
Southern Alberta	1-800-882-5616
Alberta and the Northwest Territories	1-800-563-7828
British Columbia and the Yukon Territory	1-800-663-1551
Telecommunications Device for the hearing impaired	**1-800-363-7629**
Toll Free Order Only Line (Canada and United States)	**1-800-267-6677**

How to Order Publications

Statistics Canada publications may be purchased from local authorized agents and other community bookstores, through the nearest Statistics Canada Regional Reference Centre, or by mail order to Marketing Division, Sales and Service, Statistics Canada, Ottawa K1A 0T6. Contact the nearest Statistics Canada Regional Reference Centre for more information on electronic products and services. See the order form at the end of this product for further details.

(1-613-951-7277)
Facsimile Number (1-613-951-1548)
Toronto Credit Card Only (1-416-973-8018)

Standards of Service to the Public

To maintain quality service to the public, Statistics Canada follows established standards covering statistical products and services, delivery of statistical information, cost-recovered services and service to respondents. To obtain a copy of these service standards, please contact your nearest Statistics Canada Regional Reference Centre.

Statistics Canada

1991 Census Technical Reports

Sampling and Weighting

Reference Products series

Published by authority of the Minister responsible for Statistics Canada

September 1994

Price: Canada: $20.00
United States: US$24.00
Other Countries: US$28.00

Catalogue 92-342E

ISBN 0-660-14300-3

Ottawa

La version française de cette publication est disponible sur demande (92-342F)

Note of Appreciation

Canada owes the success of its statistical system to a long-standing cooperation involving Statistics Canada, the population of Canada, its businesses and governments. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Canadian Cataloguing in Publication Data

Main entry under title:

Sampling and Weighting

(1991 census technical reports)
(Reference product series)
Issued also in French under title: Échantillonnage et pondération
ISBN 0-660-14300-3
CS92-342E

1. Canada -- Census, 1991 -- Methodology.
2. Canada – Census, 1991 -- Evaluation.
I. Statistics Canada. II. Series.

HA741.5.1991 S25 1994 304.6'0723
C94-988074-4

How to Cite This Document

Statistics Canada. Sampling and Weighting.
1991 Census Technical Reports; Reference Products Series.
Ottawa: Minister of Industry, Science and Technology, 1994.
Catalogue number 92-342E.

Preface

Through time, the Census of Canada has become the primary source of information about Canadians and how they live. Decisions based on this information affect the social and economic affairs of all Canadians.

Statistics Canada, as the professional agency in charge of producing this information, has the responsibility for informing users of data quality. The agency must describe the concepts and methodology used in collecting and processing the data, as well as any other features that may affect their use or interpretation.

In order to describe the quality of the 1991 Census data, Statistics Canada has prepared the following publications: a census **Dictionary**, which provides concise and easy to understand textual and graphical information pertaining to census concepts; a **Handbook**, which provides an overview of how the census is conducted; and a series of **Technical Reports**, which present in greater detail, information on the quality of data for specific characteristics, such as occupation, as covered in this report.

Information on data quality is important for users. It allows them to assess the usefulness of census data for their purposes as well as the risks involved in basing conclusions or decisions on these data. The 1991 Census was a large and complex undertaking and, while considerable effort was taken to ensure high standards throughout all collection and processing operations, the resulting data are inevitably subject to a certain degree of error.

Information on data quality is also important to Statistics Canada. It is an integral part in the development and maintenance of pertinent and reliable statistical programs.

This publication is a major contribution to achieving these goals. It has been prepared by Mark Majkowski of the Census Operations Section of the Social Survey Methods Division. Support was also provided from staff of two Divisions in Statistics Canada: Social Survey Methods and Census Operations.

Finally, I would like to express my appreciation to the millions of Canadians who completed their questionnaires on June 4, 1991, as well as to those who assisted Statistics Canada in planning and conducting the census.

Ivan P. Fellegi

Chief Statistician of Canada

Table of Contents

List of Tables

I. Introduction

Sampling is an accepted practice in many aspects of life today. The quality of produce in a market may be judged visually by a sample before a purchase is made; we form opinions about people based on samples of their behaviour; we form impressions about countries or cities based on brief visits to them. These are all examples of sampling in the sense of drawing inferences about the "whole" from information for a "part".

In a more scientific sense, sampling is used, for example, by accountants in auditing financial statements, in industry for controlling the quality of items coming off a production line, and by the takers of opinion polls and surveys in producing information about a population's views or characteristics. In general, the motivation to use sampling stems from a desire either to reduce costs or to obtain results faster, or both. In some cases, measurement may destroy the product (e.g., testing the life of light bulbs) and sampling is therefore essential. The disadvantage of sampling is that the results based on a sample may not be as precise as those based on the whole population. However, when the loss in precision (which may be quite small when the sample is large) is tolerable in terms of the uses to which the results are to be put, the use of sampling may be cost-effective. Furthermore, the reduction in the scale of a study achieved through using sampling may in fact lead to a reduction in errors from non-sampling sources, thus compensating to some extent for the loss of precision resulting from sampling.

The 1991 Census of population made use of sampling in a variety of ways. It was used in the testing of question wordings during development of the questionnaire; it was used in ensuring that the quality of the Census Representative's work in collecting questionnaires met certain standards; it was used in the control of the quality of coding responses during office processing; it was used in estimating both the amount of undercoverage and the amount of overcoverage which occurred for different reasons; it was used in evaluating the quality of census data. However, the primary use of sampling in the census was during the field enumeration, when all but the basic census data were collected only from a sample of households. This guide describes this last use of sampling and evaluates the effect of sampling on the quality of census data.

Chapter II reviews the history of the use of sampling in Canadian censuses and describes the sampling procedures used in the 1991 Census. Chapter III explains the procedures used for weighting up the sample data to the population level and provides operational and theoretical justifications for these procedures. In Chapter IV, the program of studies designed to evaluate the 1991 Census sampling and weighting procedures is presented, while Chapters V through VIII present the results of these studies.

II. Sampling in Canadian Censuses

In the context of a census of population, sampling refers to the process whereby certain characteristics are collected and processed only for a random sample of the dwellings and persons identified in the complete census enumeration. Tabulations that depend on characteristics collected only on a sample basis are then obtained for the whole population by scaling up the results for the sample to the full population level. Characteristics collected on all dwellings or persons in the census will be referred to as "basic characteristics" or "2A characteristics" while those collected only on a sample basis will be known as "sample characteristics" or "2B characteristics". The 2A and 2B refer to the Forms 2A and 2B which are discussed in Section B below.

A. The History of Sampling in the Canadian Census

Sampling was first used in the Canadian census[1] in 1941. A Housing Schedule was completed for every tenth dwelling in each census subdistrict. The information from 27 questions on the separate Housing Schedule was integrated with the data in the personal and household section of the Population Schedule for the same dwelling, thus allowing cross-tabulation of sample and basic characteristics. Also in the 1941 Census, sampling was used at the processing stage to obtain early estimates of earnings of wage-earners, of the distribution of the population of working age, and of the composition of families in Canada. In this case, a sample of every tenth enumeration area across Canada was selected and all Population Schedules in these areas were processed in advance.

Again in 1951, the Census of Housing was conducted on a sample basis. This time respondents in every fifth dwelling (i.e. those whose identification numbers ended in a 2 or 7) were selected to complete a housing document containing 24 questions. In the 1961 Census, persons 15 years of age and over in a 20% sample of private households were required to complete a Population Sample Questionnaire containing questions on internal migration, fertility and income. Sampling was not used in the smaller censuses of 1956 and 1966.

The 1971 Census saw several major innovations in the method of census-taking. The primary change was from the traditional canvasser method of enumeration to the use of self-enumeration for the majority of the population. This change was prompted by the results of several studies in Canada and elsewhere (Fellegi (1964); Hansen et al. (1959)) that indicated that the effect of the enumerator was a major contribution to the variance[2] of census figures in a canvasser census. Thus the use of self-enumeration was expected to reduce the variance of census figures through reducing the effect of the enumerator, while at the same time giving the respondent more time and privacy in which to answer the census questions – factors which might also be expected to yield more accurate responses.

The second aspect of the 1971 Census that differentiated it from any earlier census was its content. The number of topics covered and the number of questions asked were greater than in any previous Canadian census. Considerations of cost, respondent burden, and timeliness versus the level of data quality to be expected using self-enumeration and sampling led to a decision to collect all but certain basic characteristics on a one-third sample basis in the 1971 Census. In all but the most remote areas of Canada, every third private household received the "long form" which contained all the census questions, while the remaining private households received the "short form" containing only the basic questions covering name, relationship to head of household, sex, date of birth, marital status, mother tongue, type of dwelling, tenure, number of rooms, water supply, toilet facilities, and certain coverage items. All households in pre-identified "remote enumeration areas" and all collective dwellings[3] received the long form. A more detailed description of the consideration of the use of sampling in the 1971 Census is given in Sampling in the Census (Dominion Bureau of Statistics (1968)).

---

1 More detailed information for specific censuses can be found in the **Administrative Report, General Review, Summary Guide** or **Census Handbook** of the appropriate census. References to these reports can be found at the end of this guide.

2 The "variance" of an estimate is a measure of its precision. Variance is discussed more fully in Chapter VIII.

3 A collective dwelling is a dwelling of a commercial, institutional or communal nature. Examples include hotels, hospitals, staff residences and work camps.

The content of the 1976 Census was considerably less than that of the 1971 Census. Furthermore, the 1976 Census did not include the questions that cause the most difficulty in collection (e.g., income) or that are costly to code (e.g., occupation, industry, and place of work). Therefore, the benefits of sampling in terms of cost savings and reduced respondent burden were less clear than for the 1971 Census. Nevertheless, after estimating the potential cost savings to be expected with various sampling fractions, and considering the public relations issues related to a reversion to 100% enumeration after a successful application of sampling in 1971, it was decided to use the same sampling procedure in 1976 as in 1971.

Most of the methodology used in the 1971 and 1976 Censuses was kept for the 1981 Census, except that the sampling rate was reduced from every third occupied private household to every fifth. Studies done at the time showed that the resulting reduction in data quality (measured in terms of variance) would be tolerable, would not be significant enough to offset the benefits of reduced cost and response burden, and would improve timeliness (see Royce (1983)). Twelve questions were asked on a 100% basis and an additional 34 questions were asked of the sample population.

The 1986 Census was the first full mid-decade census. It was decided that only a full census could meet the growing need for local labour market data, a need made more pressing by the occurrence of a major recession (1981-82) since the previous census. However, in order to keep development costs as low as possible, a policy of minimum change was adopted. Unless there were compelling reasons not to do so, 1981 Census questions and data collection and processing procedures were retained. Questions on eight subjects from the 1981 Census were not asked in 1986, while three new questions were added. After the collection of 1986 Census questionnaires, a sample of respondents was selected to participate in the post-censal Health and Activity Limitation Survey (HALS). HALS, which was conducted for the first time in 1986, was designed to provide a comprehensive picture of the lives of persons with disabilities.

In 1991, the Census of population included both permanent and non-permanent residents[4] of Canada. With the exception of the 1941 Census, only permanent residents of Canada were included in censuses prior to 1991. In order to identify the non-permanent residents, a new question for the 1991 Census had to be designed and added. In total, twelve new questions were added for the 1991 Census, while questions on four subjects from the 1986 Census were not asked in 1991. Of the twelve new questions, seven appeared for the very first time and five questions were reinstated from previous censuses. Two post-censal surveys were conducted in 1991 following the completion of the collection of 1991 Census questionnaires. The two surveys were the HALS (also conducted in 1986) and the Aboriginal Peoples Survey (APS). The APS, which was conducted for the first time in 1991, collected information from the aboriginal population living both on and off reserves. Also in the 1991 Census, there was a significant increase in the automation of data processing as well as in the way in which products and services are produced and delivered to the client.

B. The Sampling Scheme Used in the 1991 Census

A wealth of information was collected from everyone in Canada on Census Day, 1991. The bulk of the information was acquired on a sample basis. In all self-enumeration areas, four out of every five private occupied households received a short form (Form 2A) containing nine basic questions on age, sex, marital status, common-law status, mother tongue, relationship to the household reference person (Person 1), dwelling type and tenure. Every fifth household received the long form (Form 2B) containing the nine basic census questions plus 44 more questions which were asked on socio-economic and dwelling-related topics.

All dwellings in those areas enumerated by the canvasser method (generally remote areas or Indian reserves) received the Form 2D. The content of the Form 2D was identical to that of the Form 2B (except for the tenure question), but was designed to be administered in a face-to-face interview situation.

[4] Non-permanent residents are persons who hold student or employee authorizations, Minister's permits or who are refugee claimants.

A Form 2B was also created for all collective dwellings. However, the residents[5] of institutional collective dwellings were not asked the sample questions. Only the basic information was collected for residents of these dwellings. Staff members who live in these institutional collective dwellings, residents of non-institutional collective dwellings (including live-in staff) as well as Canadians stationed abroad (generally embassy or Armed Forces personnel) were asked to give long form information to questions that did not include the housing questions. However, questions about the person's usual place of residence in Canada were asked of the Canadians stationed abroad. Information on unoccupied private dwellings was recorded on a Form 2A.

The basic drop-off or delivery procedure required the Census Representative (CR) to pre-plan a route covering all dwellings in his/her enumeration area (EA) and then to visit each dwelling and leave a census questionnaire. The selection of the sample, i.e. the decision as to which type of questionnaire to leave at each occupied dwelling, was facilitated by the Visitation Record (VR), the document in which the CR listed each dwelling in his/her area. This document was printed so that every fifth line was shaded to signify that a Form 2B should be delivered. A random start was implemented by deleting either zero, one, two, three or four lines at the start of the VR according to whether the fifth, fourth, third, second or first dwelling in the EA was to be the first to receive the long form. Thereafter, the dwelling listed on each shaded line automatically received the long form. These procedures were spelled out in the CR's Manual (Form 41) and emphasized in his/her training in order to minimize the risk of any deviation from the specified procedure for selecting the sample. Quality control checks of the duties performed by the CR were done by the Census Commissioner.

In sampling terminology, the sample can be described as a stratified systematic sample of private occupied dwellings using a constant 1 in 5 sampling rate in all strata (EAs). As a sample of persons, it can be regarded as a stratified systematic cluster sample with dwellings as clusters. For a more detailed description of the concepts and terminology of sampling, see Stuart (1976), or Cochran (1977).

C. Processing the Census Sample

Once the CR had obtained the completed questionnaire (Form 2A, 2B or 2D) from each dwelling in his/her area, and his/her work had been approved, the questionnaires were sent to one of seven regional processing sites for manual processing. At these sites, questionnaires were logged, counted and prepared for key entry. Preparation included consistency checks between the questionnaires and the Visitation Record as well as legibility checks to ensure that documents were suitable for computer entry. Also, written responses to five questions were converted into numeric codes suitable for direct data entry. Transcriptions of Form 4A information (created for missing or refusal households) to Forms 2A or 2B, as well as long form information collected from persons stationed abroad or in collective dwellings to Forms 2B, were made at these sites. Complete data for each EA were captured and stored on magnetic tapes. The questionnaires and magnetic tapes were then sent for head office processing in Ottawa.

At the head office processing stage, automated structural edits were carried out at the enumeration area, household and person levels, and inconsistencies – such as person count conflicts and household number conflicts between the geographic levels – were resolved manually. An automated coding operation converted written responses for many of the questions to numeric codes. For the first time, this was done by automatically matching the captured written responses received from the head office processing operation against an automated reference file/classification structure. This structure contained a series of words or phrases and corresponding numeric codes for each of these variables. At the end of head office processing, Form 2B households with non-response to all the 2B characteristic questions were converted to Form 2A households. Doing this reduced the sample size and hence increased the size of the sample weights applied to the remaining Form 2B households. It was felt, however, that better-quality estimates would result from doing this than if all the 2B responses for these households had been imputed. After all resulting updates to the data for an EA were completed, the data were reformatted and transferred to the edit and imputation phase.

5 These persons would be inmates of correctional and penal institutions or jails; patients in hospitals; occupants of residences for senior citizens; patients in chronic care hospitals or psychiatric institutions; children in children's group homes, orphanages, or young offenders' facilities.

The data were loaded to ten edit and imputation databases, organized by sample size, i.e. 2A (100%) and 2B (20%), with five databases for each. The five databases corresponded to the four geographic regions of Canada (East, Quebec, Ontario and West) plus a database which corresponded to the Canadians stationed abroad (referred to as 2C). The 2A databases contained the basic demographic characteristics for 100% of the population, while the 2B databases contained the data for the 20% sample questions. The data were processed through a series of customized modules, where all problems of invalid, inconsistent, and missing data were resolved. The 2A databases were processed first, and a final 2A Canada Retrieval Data Base was created.

Once the 100% data were finalized, the data for the 20% sample questions were processed. A final 2B Canada Retrieval Data Base was created, which contained both the 100% and 20% data for sampled households and persons only. The weights created using the 100% data (as described in Chapter III) were placed on this database.

III. Estimation from the Census Sample

Any sampling procedure requires an associated estimation procedure for scaling sample data up to the full population level. The choice of an estimation procedure is generally governed by both operational and theoretical constraints. From the operational viewpoint, the procedure must be feasible within the processing system of which it is a part, while from the theoretical viewpoint the procedure should minimize the sampling error of the estimates it produces. In the following two sections, the operational and theoretical considerations relevant to the choice of estimation procedures for the census sample are described.

A. Operational Considerations

Mathematically, an estimation procedure can be described by an algebraic formula that shows how the value of the estimator for the population is calculated as a function of the observed sample values. In small surveys that collect only one or two characteristics, or in cases where the estimation formula is very simple, it might be possible to calculate the sample estimates by applying the given formula to the sample data for each estimate required. However, in a survey or census in which a wide range of characteristics is collected, or in which the estimation formula is at all complex, the procedure of applying a formula separately for each estimate required is not feasible. In the case of a census, for example, every cell of every tabulation based on sample data at every geographic level represents a sample estimate, which according to this approach would require a separate application of the estimation formula. In addition, the calculation of each estimate separately would not necessarily lead to consistency between the various estimates made from the same census sample.

The approach taken in the census therefore (and in most sample surveys) is to split the estimation procedure into two stages: (a) the calculation of weights (known as the weighting procedure); (b) the summing of weights to produce estimated population counts. Any mathematical complexity is then contained in step (a), which is performed just once, while step (b) is reduced to a simple process of summing weights which takes place at the time a tabulation is retrieved. Also, since the weight attached to each sample unit is the same for whatever tabulation is being retrieved, consistency between different estimates based on sample data is assured.

B. Theoretical Considerations

For a given sample design and a given estimation procedure, one can, from sampling theory, make a statement about the chances that a certain interval will contain the unknown population value being estimated. The primary criterion in the choice of an estimation procedure is minimization of the width of such intervals so that these statements about the unknown population values are as precise as possible. The usual measure of precision for comparing estimation procedures is known as the standard error. Provided that certain relatively mild conditions are met, intervals of plus or minus two standard errors from the estimate will contain the population value for approximately 95% of all possible samples.

As well as minimizing standard error, a second objective in the choice of estimation procedure for the census sample is to ensure, as far as possible, that sample estimates for basic (i.e., 2A) characteristics are consistent with the corresponding known population values. Fortunately, these two objectives are usually complementary in the sense that sampling error tends to be reduced by ensuring that sample estimates for certain basic characteristics are consistent with the corresponding population figures. While this is true in general, however, forcing sample estimates for basic characteristics to be consistent with corresponding population figures for very small subgroups can have a detrimental effect on the standard error of estimates for the sample characteristics themselves.

In the absence of any information about the population being sampled other than that collected for sample units, the estimation procedure would be restricted to weighting the sample units inversely to their probabilities of selection (e.g., if all units had a 1 in 5 chance of selection, then all selected units would receive a weight of 5). In practice, however, one almost always has some supplementary knowledge about the population (e.g., its total size, and possibly its breakdown by a certain variable - perhaps by province). Such information can be used to improve the estimation formula so as to produce estimates with a greater chance of lying close to the unknown population value. In the case of the census sample, a large amount of very detailed information about the population being sampled is available in the form of the basic 100% data at every geographic level. On the one hand, we can take advantage of

this population information to improve the estimates made from the census sample; on the other hand, this wealth of information can also be an embarrassment in the sense that it is impossible to make the sample estimates for basic characteristics consistent with all the population information at every geographic level. Differences between sample estimates and population values become visible when a cross-tabulation of a sample variable and a basic variable is produced. The tabulation has to be based on sample data, with the result that the marginal totals for the basic variable are sample estimates that can be compared with the corresponding population figures appearing in a different tabulation based on 100% data. They will not necessarily agree exactly.

C. Developing an Estimation Procedure for the Census Sample

Given that a weight has to be assigned to each unit (person, family or household) in the sample, the simplest procedure would be to give each unit a weight of 5 (because a 1 in 5 sample was selected). Such a procedure would be simple and unbiased[6] and, if nothing but the sample data were known, it might be the optimum procedure. However, although we know that the sample will contain almost exactly one-fifth of all households (excluding collective households and those in canvasser areas), one cannot be certain that it will contain exactly one-fifth of all persons, or one-fifth of each type of household, or one-fifth of all females aged 25-34, and so on. Therefore, this procedure would not ensure consistency even for the most important subgroups of the population. For large subgroups, these fractions should be very close to one-fifth, but for smaller subgroups they could differ markedly from one-fifth. The next most simple procedure would be to define certain important subgroups (e.g., age-sex groups within provinces) and, for each subgroup, to count the number of units in the population in the subgroup (N) and the number in the sample (n) and to assign to each sample unit in the subgroup a weight equal to N/n.

For example, if there were 5,000 males aged 20-24 enumerated in Prince Edward Island, and 1,020 of these fell in the sample households, then a weight of 5,000/1,020 = 4.90 would be assigned to each male aged 20-24 in the sample in Prince Edward Island. This would ensure that whenever sex and age in five-year groups were cross-classified against a sample characteristic for Prince Edward Island, the marginal total for the male 20-24 age-sex group would agree with the population total of 5,000. This type of estimation procedure is known as "ratio estimation". It should be noted in this particular example that a weight of 5 would result in a sample estimate of 5,100 (1,020 x 5). The estimation procedure that was used in the 1986 Census was a generalization of ratio estimation called the raking ratio estimation procedure (RREP). For more details on the RREP, see the User's Guide to the Quality of 1986 Census Data: Sampling and Weighting as well as Brackstone and Rao (1979).

For the 1991 Census, it was decided to use an alternative estimation procedure called the "two-step generalized least squares estimation procedure" (GLSEP). This was done to achieve a higher level of agreement between population counts and the corresponding estimates at the EA level than was possible with RREP. The standard errors of the estimates under GLSEP for small geographical areas were also reduced. In addition, the GLSEP allowed a single weight to be determined for each sampled household that could be used to produce estimates for both person and household characteristics. With the RREP, it was necessary to use different weights to produce estimates for household and person characteristics, and this sometimes led to inconsistencies. Inconsistencies also sometimes resulted because the RREP iterative procedure to calculate the weights did not always converge (see Daoust and Bankier (1989)).

With the GLSEP (which can be shown to be a regression estimator), the initial weights of approximately 5 were adjusted as little as possible for individual households while ensuring that there was perfect agreement between the estimates and the population counts for as many of the basic characteristics as possible. These so-called "constraints" are listed in Appendix A. It was required that this perfect agreement be achieved at the weighting area (WA) level. Each WA contained, on average, seven sampled EAs. More information on WAs is given in Chapter VI, Section A of this report.

6 "Unbiased" means that the average of the estimates obtained by this procedure, over all possible samples, would equal the true population value.

D. The Two-step Generalized Least Squares Estimation Procedure

The weighting calculations are carried out independently in each WA. Some of the constraints (both at the EA and WA levels) listed in Appendix A have to be discarded for each WA, and hence population/estimate agreement cannot be guaranteed for all constraints. Constraints are initially discarded at the WA level because:

- they apply to less than 60 households (these are called "small" constraints);
- they are redundant (these are called "linearly dependent" (LD) constraints); or
- they are nearly redundant (these are called "nearly linearly dependent" (NLD) constraints).

For example, since the total number of females plus the total number of males equals the total number of persons, the total number of females can be dropped as a redundant or "linearly dependent" constraint, since any two of the constraints will guarantee that the third will be satisfied. An example of a nearly redundant constraint can be seen by considering the constraints that represent persons whose marital status is "separated", and household maintainers whose marital status is "separated". If most, but not all, separated persons are household maintainers, then the two constraints are almost equal and one constraint can be considered NLD. The LD constraints were discarded, to increase the computational efficiency of the weighting algorithm. The small and NLD constraints were discarded, because otherwise the estimates might become unstable and have large standard errors.

After small, LD, and NLD constraints are discarded at the WA level, the calculation of the GLSEP weights takes place in two steps. In the first step, the initial weights, which equal the reciprocal of the EA household sampling fraction, are adjusted individually for each EA. Some constraints may be discarded due to smallness or linear dependence at the EA level which were not discarded at the WA level. The remaining constraints that have not been discarded in the EA are sorted by the number of households that they apply to at the EA level. The constraints are then split into two groups, with the even-numbered constraints in one and the odd-numbered constraints in the other. The GLSEP weights are calculated at the EA level for each group of constraints. Sometimes, the estimation procedure will produce very small weights (less than one) or very large weights (greater than 25) in order to obtain the necessary agreement for certain constraints. These weights, which are called "outlier" weights, are undesirable. Consequently, when this occurs, the constraints causing them are identified and discarded, and the weights are recalculated. Finally, the weights for the two groups of constraints are averaged together for each sampled household to produce the first step weights for each EA.

The weights produced in the first step are used as initial weights in the second step, where they are adjusted so that agreement is obtained between sample estimates and population counts at the WA level. All constraints not identified as small, LD, or NLD at the WA level are used. Again, if any outlier weights are produced, the constraints causing them are identified and discarded, and the final weights are recalculated. Although the second step destroys somewhat the agreement obtained for estimates at the EA level in the first step, the final EA level estimates are still closer to the population counts than they would have been had the first step not been done. Also, constraints requiring exact agreement for the total number of households and total number of persons at the EA level (see Appendix A) are applied in the second step weighting adjustment unless they are discarded for being small, LD, or NLD, or for causing outlier weights. For a more detailed explanation of the calculation of the weights, see Bankier, Rathwell and Majkowski (1992).

EAs where both Forms 2A and Forms 2B were distributed to the private occupied dwellings are known as sampled EAs. GLSEP weights were calculated only for Form 2B households in private occupied dwellings in sampled EAs. Private occupied dwelling households that received a Form 2A in sampled EAs were given a weight of 0. All private occupied dwelling households in non-sampled EAs received a weight of 1 along with all collective households regardless of what type of EA they belonged to.

IV. The Sampling and Weighting Evaluation Program

The sampling and weighting evaluation program was designed to determine the effect of sampling and weighting on the quality of census sample data. To this end, five studies were carried out to measure the quality of the census sample data and estimates and to provide information relevant to the planning of future censuses. These studies were:

(a) an examination of sampling bias;

(b) an evaluation of the formation of weighting areas;

(c) an evaluation of the weighting procedures;

(d) an evaluation of sample estimate and population count consistency;

(e) a study to evaluate the sampling variance for various 20% sample characteristics.

In the remainder of this chapter, these five studies are briefly described. Chapters V through VIII present the results of these studies.

A. Sampling Bias Study

Bias can be introduced into responses to any survey from a number of sources. The objective of this study was to determine if responses to basic questions on Forms 2B were biased in any way and to identify, if possible, the causes of any observed bias.

B. Evaluation of Weighting Area Formation

The objective of this study was to measure the degree to which WAs met the criteria laid down for their formation. All WAs in Canada were analyzed to determine how well they respected the size constraints and the boundaries of various types of geographic areas.

C. Evaluation of Weighting Procedures

The objective of this study was to evaluate the performance of the GLSEP. The level of agreement between the sample estimates and population counts for the constraints over all WAs in Canada was examined. The number and type of constraints discarded at the WA level as well as the reasons for them being discarded were studied to explain observed inconsistencies. In addition, the distribution of the GLSEP weights as well as differences between 1991 results and 1986 results were studied.

D. Sample Estimate and Population Count Consistency Study

This study examined the level of agreement (consistency) between sample estimates and population counts for a wide variety of basic characteristics, not just those used as constraints in the GLSEP. This consistency was studied for various geographic areas other than WAs. Comparisons were also made between the consistency achieved in 1991 and 1986 for these characteristics.

E. Sampling Variance

The "variance" of an estimate is a measure of its precision. Estimates of variance for estimators using simple weights of 5 and assuming simple random sampling are relatively inexpensive to calculate. However, estimates of variance for census estimators, taking into account the sample design and estimation techniques used, are very expensive to calculate. It is discussed how "adjustment factors" were calculated for the 1986 Census, which are the ratios of the estimates of the standard errors (the square roots of the variances) for census estimates to the simple estimates of the standard errors. An estimate of the standard error of a census estimate for any characteristic in any geographic area can then be obtained by multiplying the simple estimate of the standard error by the appropriate adjustment factor. It is then discussed how these estimates of the standard error may not be accurate because of the bias introduced into the process by the sample, the data processing and the estimation procedure.

V. Sampling Bias

A. Introduction

Estimates based on a sample survey are subject to sampling errors. One type of sampling error arises from the variability in the population. This variability means that different samples will produce different estimates, none of which will necessarily equal the true population value. The estimates will equal the true population value on average, however, provided that there is no bias in the sample creating a tendency to overestimate or underestimate. Unfortunately, bias is often difficult to eliminate completely. In the census of population, bias can be introduced into the responses from a variety of sources. These include coverage errors, non-response bias, response bias (e.g., respondents answering differently on the Form 2B than on the 2A), CR errors (e.g., not selecting the sample according to specifications), processing errors, and so on.

The purpose of the Sampling Bias Study was to search for bias in the responses to the basic questions on Forms 2B. Sample estimates for 53 basic characteristics (Appendix B describes how these characteristics relate to the Appendix A constraints) based on imputed data were compared to the population counts for all 284 sampled census divisions (CDs) in Canada. The sample estimates were produced by multiplying the sample counts at the EA level by simple weights equal to the inverse of the EA household sampling fraction (approximately 5) and then summing them to the CD level.[7] It was found that the average difference between the sample estimates and the population counts, over all CDs, was statistically significant (at the 5% level)[8] for most of the characteristics (i.e. the differences cannot be explained by sampling variability). This was determined using the statistic

$$Z^{(0)} = \frac{\hat{X}^{(0)} - X}{\sqrt{V(\hat{X}^{(0)})}} \qquad (1)$$

where $\hat{X}^{(0)}$ is an estimate based on simple weights of the known 2A population count X and $V(\hat{X}^{(0)})$ is the sampling variance of the estimator $\hat{X}^{(0)}$. The $Z^{(0)}$ values, for the 284 CDs, should approximately follow a normal distribution with mean 0 and variance 1 if a simple random sample of households was selected unbiasedly from each EA and was not affected by processing (see Appendix C for more details).

B. Main Findings

Table 1 shows the differences (in absolute and percentage terms) between the sample estimates and the population counts at the Canada level for the set of 53 2A characteristics. In most cases the bias was less than 1%. There are 43 characteristics flagged with asterisks in Table 1, however, for which the differences where found to be statistically significant at the 5% level based on the statistic $Z^{(0)}$ in Table 2. (It should be noted that the counts and percentages in Table 1 of the User's Guide to the Quality of 1986 Census Data: Sampling and Weighting were in error and should have been multiplied by a factor of 2.6.)

There was a definite tendency for the following groups of people to be over-represented in the sample: females, age groups 0-4, 5-9, 10-14 and 45-49, and census family persons, in particular married persons and census family children. The following groups of people were under-represented in the sample: age groups 20-24, 25-29 and greater than 74; divorced and separated persons; and non-census family persons.

7 These simple estimates were used instead of the GLSEP estimates because the GLSEP tends to mask the sampling bias by forcing estimates of basic characteristics to equal population counts.

8 This means that there was at most a 5% chance of obtaining such large differences in the absence of bias.

Table 1. Sample Estimate (Simple Weights) Minus Population Count at Canada Level (Sampled EAs Only) and Percentage of CDs in Which Characteristic was Over-represented

Characteristics Studied	Sample Estimate Minus Population Count			Percentage of Over-represented CDs
Person Characteristics	**Value**		**Percent**	
Males	3,275	*	(+0.03%)	63
Females	64,216	*	(+0.48%)	78
Total Persons	67,491	*	(+0.26%)	76
Age 0-4	16,950	*	(+0.92%)	63
Age 5-9	21,031	*	(+1.14%)	68
Age 10-14	21,376	*	(+1.17%)	64
Age 15-19	8,115	*	(+0.45%)	59
Age 20-24	-16,841	*	(-0.89%)	43
Age 25-29	-16,727	*	(-0.73%)	46
Age 30-34	170	*	(+0.01%)	60
Age 35-39	3,000	*	(+0.13%)	62
Age 40-44	7,938	*	(+0.39%)	61
Age 45-49	10,017	*	(+0.63%)	62
Age 50-54	5,339	*	(+0.41%)	58
Age 55-59	3,034		(+0.26%)	51
Age 60-64	4,191		(+0.37%)	51
Age 65-74	7,063		(+0.39%)	49
Age > 74	-7,165	*	(-0.68%)	35
Single Persons	12	*	(+0.00%)	56
Married Persons	95,348	*	(+0.83%)	85
Widowed Persons	-5,073	*	(-0.42%)	37
Divorced Persons	-15,198	*	(-1.22%)	38
Separated Persons	-7,598	*	(-1.31%)	39
Family Characteristics				
Total Census Families	52,069	*	(+0.72%)	88
Husband-Wife Families	54,989	*	(+0.87%)	89
Lone-parent Census Families	-2,921	*	(-0.31%)	45
Census Family Children	72,463	*	(+0.84%)	75
People in Census Families	179,522	*	(+0.81%)	85
People Not in Census Families	-112,031	*	(-2.75%)	9
Household and Dwelling Characteristics				
Owned Dwellings	46,713	*	(+0.75%)	81
Rented Dwellings	-46,713	*	(-1.28%)	19
Single-detached Dwellings	27,243	*	(+0.49%)	77
Apts with 5 or More Storeys	-931		(-0.10%)	40
Movable Dwellings	-796		(-0.46%)	60
All Other Dwelling Types	-25,516	*	(-0.80%)	26
Total Households	0		(+0.00%)	0
One-person Households	-37,392	*	(-1.66%)	15
Two-person Households	12,541	*	(+0.40%)	58
Three-person Households	8,606	*	(+0.50%)	59
Four-person Households	15,320	*	(+0.88%)	69
Five-person Households	4,857	*	(+0.68%)	58
Six-or-more-person Households	-3,932		(-1.25%)	43
Non-census-family Households	-56,518	*	(-2.07%)	11
One-census-family Households	60,762	*	(+0.87%)	92
Hhld Maintainers Aged < 25	-9,011	*	(-1.98%)	36
Hhld Maintainers Aged 25-34	-2,409		(-0.11%)	52
Hhld Maintainers Aged 35-44	7,652	*	(+0.33%)	62
Hhld Maintainers Aged 45-54	7,495	*	(+0.46%)	58
Hhld Maintainers Aged 55-64	685		(+0.05%)	49
Hhld Maintainers Aged 65-74	1,265		(+0.11%)	46
Hhld Maintainers Aged > 74	-5,676	*	(-0.76%)	36
Male Household Maintainers	-28,260	*	(-0.41%)	38
Female Household Maintainers	28,260	*	(+0.95%)	62

* These differences were found to be statistically significant at the 5% level.

In terms of household characteristics, there was a tendency for owned dwellings and female household maintainers to be over-represented in the sample, while rented dwellings and dwellings whose dwelling type (e.g., "single-detached") was classified as "other" tended to be under-represented. There was a tendency for one-census-family and husband-wife family households to be over-represented, while non-census family households were under-represented. As well, there was a tendency for three-, four- and five-person households to be over-represented while one-person households were under-represented. Household maintainers aged 45-54 were over-represented, while those aged less than 25 and greater than 74 were under-represented.

Table 2 shows that the means of the $Z^{(0)}$ values (under the "All Records" column) for many characteristics were farther from 0 than could be explained by sampling variability. The 43 mean values marked with (*) indicates that the hypothesis that the mean of the $Z^{(0)}$ values equals zero was rejected at the 5% level. The column next to the mean values, T: Mean=0, gives the t-statistic for testing the hypothesis that the mean was equal to zero.[9] The other columns of Table 2 are discussed in the following paragraphs. Plots of histograms of the $Z^{(0)}$ values overlaid with the normal distribution were produced for two characteristics to give a visual picture of the results in Table 2. The plots which appear in Appendix D are for "Total Persons" and "Male Household Maintainers". The plot for "Total Persons" shows the histogram is shifted to the right (mean=0.71) in comparison to the normal distribution. The plot for "Male Household Maintainers" shows the histogram is shifted to the left (mean=-0.32).

C. Reasons for Bias

As mentioned earlier, there are many possible explanations for the observed differences between the sample estimates based on simple weights and the population counts. One possibility arises from the fact that there were 253,156 (2.6% of the total) missed/refusal households in the 1991 Census. These were either households which completely refused to answer the questions or for which the CR was unable to get any information (usually because the members of the household were absent during the census-taking period or had moved on or after Census Day without responding). The CR was sometimes able to determine the number of persons and the tenure of the dwelling and almost always recorded the dwelling type, but usually all other responses had to be imputed for these households. Of the missed/refusal households, 43,155 were sampled households. In addition, 6,753 of the sampled households, while not of the "missed/refusal" type (i.e. they provided some responses to the basic questions), provided no answers to the questions asked on a sample basis. During data processing, these 43,155 + 6,753 = 49,908 sampled households with complete non-response to the sampled questions were removed from the sample (i.e. they were converted from Form 2B to Form 2A households so that they became non-sampled households), and the responses to the basic questions only, were imputed. This procedure of converting sampled households to non-sampled households is known as 2A/2B document conversion. It is possible that the missed/refusal households and the households without responses to the sample questions had different characteristics (e.g., they could have been smaller) than other households. Thus converting 2Bs to 2As could bias the sample. Also, if the imputation system had a tendency to impute certain characteristics for missed/refusal households more often than for other types of households, this would have caused sample estimate and population count discrepancies as well, since only non-sampled households would be affected.

To examine the impact of missed/refusal households and 2A/2B document conversion on the sampling bias, three different situations were studied. First of all, missed/refusal households were excluded (253,156 households), the simple weights were adjusted to reflect this, and the $Z^{(0)}$ statistics were recalculated (see the "Missed/Refusal Excluded" column in Table 2). Secondly, instead of missed/refusal households being dropped, the 2A documents were converted back to their original 2B document type so that they would be included in the sample (see the "Conversions Reversed" column in Table 2). This situation involved 49,908 households being converted back to 2Bs. Finally, the third situation had both the conversions being done and the missed/refusal households being dropped (see the "A & B" columns in Table 2). This situation involved 6,753 households being converted back to 2Bs, and excluding 253,156 households. The bias remained statistically significant at the 5% level for 42 of the 53 characteristics after

[9] This test should be valid given the large number of observations (284 CDs) and the high degree of normality of the $Z^{(0)}$ values for most characteristics.

Table 2. 1991 Summary Statistics for Means of $Z^{(0)}$ Values at CD Level (Sampled EAs)

Characteristics Studied	All Records		Missed / Refusal Excluded (A)		Conversions Reversed (B)		A & B	
	Mean	T:Mean=0	Mean	T:Mean=0	Mean	T:Mean=0	Mean	T:Mean=0
Person Characteristics								
Males	0.32*	5.08	0.19*	3.04	0.14*	2.14	0.16*	2.41
Females	0.76*	11.97	0.65*	10.27	0.53*	8.51	0.60*	9.55
Total Persons	0.71*	11.20	0.55*	8.55	0.44*	6.66	0.49*	7.60
Age 0-4	0.32*	5.21	0.27*	4.43	0.19*	3.14	0.25*	4.01
Age 5-9	0.50*	7.61	0.44*	6.80	0.39*	6.04	0.42*	6.43
Age 10-14	0.50*	8.24	0.44*	7.24	0.40*	6.58	0.42*	6.87
Age 15-19	0.21*	3.41	0.14*	2.35	0.13*	2.12	0.13*	2.13
Age 20-24	-0.23*	-3.60	-0.29*	-4.53	-0.29*	-4.45	-0.28*	-4.46
Age 25-29	-0.15*	-2.52	-0.21*	-3.41	-0.23*	-3.70	-0.22*	-3.57
Age 30-34	0.21*	3.18	0.16*	2.45	0.14	2.08	0.15*	2.24
Age 35-39	0.19*	3.00	0.14*	2.19	0.11*	1.77	0.12	1.90
Age 40-44	0.25*	4.30	0.20*	3.52	0.19*	3.32	0.19*	3.30
Age 45-49	0.26*	4.31	0.22*	3.72	0.21*	3.60	0.21*	3.48
Age 50-54	0.14*	2.27	0.12*	2.06	0.12*	2.03	0.12*	2.03
Age 55-59	0.00	-0.06	0.01	0.19	0.02	0.28	0.02	0.28
Age 60-64	0.05	0.99	0.08	1.46	0.07	1.25	0.08	1.47
Age 65-74	-0.01	-0.18	0.04	0.71	0.04	0.63	0.04	0.60
Age > 74	-0.38*	6.55	-0.34*	-5.73	-0.33*	-5.51	-0.33*	-5.61
Single Persons	0.20*	3.21	0.04	0.62	-0.01	-0.18	0.01	0.18
Married Persons	1.08*	17.42	1.05*	16.92	0.93*	14.80	0.99*	16.07
Widowed Persons	-0.36*	-6.11	-0.30*	-5.04	-0.24*	-4.02	-0.27*	-4.48
Divorced Persons	-0.31*	-5.19	-0.36*	-5.98	-0.34*	-5.57	-0.35*	-5.86
Separated Persons	-0.32*	-5.17	-0.35*	-5.70	-0.30*	-4.99	-0.34*	-5.56
Family Characteristics								
Total Census Families	1.20*	18.90	1.01*	16.48	0.76*	12.27	0.91*	14.99
Husband-Wife Families	1.18*	18.10	1.02*	16.14	0.77*	12.24	0.93*	14.92
Lone-parent Census Families	-0.12*	-2.05	-0.14*	-2.35	-0.10	-1.70	-0.14*	-2.41
Census Family Children	0.72*	11.69	0.61*	10.07	0.52*	8.48	0.57*	9.35
People in Census Families	1.15*	17.91	0.98*	15.80	0.77*	12.43	0.89*	14.56
People Not in Census Families	-1.47*	-18.79	-1.37*	-16.85	-1.07*	-13.82	-1.28*	-16.15
Household and Dwelling Characteristics								
Owned Dwellings	0.87*	12.85	0.69*	11.36	0.56*	9.89	0.65*	10.81
Rented Dwellings	-0.87*	-12.85	-0.69*	-11.36	-0.56*	-9.89	-0.65*	-10.81
Single-detached Dwellings	0.57*	10.63	0.36*	7.55	0.23*	5.07	0.33*	6.95
Apts with 5 or More Storeys	0.00	0.12	0.01	0.32	0.04	0.96	0.03	0.78
Movable Dwellings	-0.04	-0.90	-0.03	-0.53	-0.02	-0.37	-0.02	-0.35
All Other Dwelling Types	-0.56*	-10.50	-0.36*	-7.32	-0.23*	-4.92	-0.33*	-6.71
Total Households	0.00	-	0.00	-	0.00	-	0.00	-
One-person Households	-1.04*	-17.07	-0.84*	-14.26	-0.65*	-10.69	-0.75*	-12.67
Two-person Households	0.18*	3.07	0.15*	2.54	0.08	1.34	0.12*	2.08
Three-person Households	0.20*	3.56	0.15*	2.69	0.12*	2.15	0.13*	2.41
Four-person Households	0.48*	7.62	0.41*	6.59	0.35*	5.58	0.38*	6.15
Five-person Households	0.27*	4.55	0.23*	3.83	0.21*	3.53	0.21*	3.61
Six-or-more-person Households	-0.08	-1.36	-0.12	-1.93	-0.12	-1.92	-0.13*	-2.05
Non-census-family Households	-1.31*	-20.42	-1.13*	-18.17	-0.87*	-14.16	-1.03*	-16.81
One-census-family Households	1.37*	20.99	1.19*	18.86	0.94*	15.19	1.10*	17.64
Hhld Maintainers Aged < 25	-0.32*	-5.69	-0.35*	-5.98	-0.32*	-5.44	-0.33*	-5.71
Hhld Maintainers Aged 25-34	0.08	1.33	0.04	0.72	0.03	0.43	0.04	0.58
Hhld Maintainers Aged 35-44	0.27*	4.40	0.23*	3.69	0.22*	3.53	0.22*	3.50
Hhld Maintainers Aged 45-54	0.21*	3.82	0.19*	3.39	0.20*	3.56	0.19*	3.32
Hhld Maintainers Aged 55-64	-0.04	-0.80	-0.02	-0.28	-0.01	-0.25	-0.01	-0.13
Hhld Maintainers Aged 65-74	-0.10	-1.76	-0.03	-0.55	-0.02	-0.38	-0.03	-0.51
Hhld Maintainers Aged > 74	-0.39*	-6.52	-0.33*	-5.41	-0.32*	-5.23	-0.32*	-5.24
Male Household Maintainers	-0.32*	-4.78	-0.39*	-5.80	-0.49*	-7.05	-0.43*	-6.28
Female Household Maintainers	0.32*	4.78	0.39*	5.80	0.49*	7.05	0.43*	6.28

* These differences were found to be statistically significant at the 5% level.

missed/refusal households were excluded, for 39 of the 53 characteristics after the document conversions were reversed, and for 42 of the 53 characteristics after both the document conversions were reversed and the missed/refusal households were dropped. Thus, although these factors definitely contributed to the bias, much of the bias still remains. The bulk of the bias still present is probably due to one or more factors such as non-response bias, response bias, and/or the selection of a biased sample by the CRs.

VI. Evaluation of Weighting Procedures

A. Weighting Area (WA) Formation

The first stage of the weighting procedures was the formation of WAs. A WA is the smallest geographic area for which agreement for characteristics of the population between certain sample and population counts can be ensured. A WA satisfies the following conditions:

(a) a WA should contain between 2,000 and 7,000 persons (population count);

(b) WA boundaries must respect the boundaries of census divisions (CDs), and as far as possible, of census subdivisions (CSDs), census tracts (CTs), and federal electoral districts (FEDs);

(c) WAs should be made up of contiguous EAs (i.e. be connected).

The sampled EAs were formed into 5,736 WAs with an average population (excluding persons in collective dwellings) of 4,583. Of the 5,736 WAs, 5,727 (99.8%) fell within the population range of 2,000-7,000. The nine WAs outside this range all had populations below 2,000, since each one of them consisted solely of an entire CD with a population less than 2,000. Only two of these nine were run through the GLSEP, with acceptable results being produced for both of the WAs thus analyzed. The other seven were not run through the GLSEP. One was custom-weighted while the other six contained no sample EAs and, therefore, did not require any weighting. The WA that was custom-weighted by a GLSEP prototype consisted of an entire CD which had sampled EAs that contained a population of only 38 persons.

The extent to which WAs respected the boundaries of various geographic areas was examined separately for CTs, CSDs in census-tracted areas, CSDs in non-census-tracted areas and FEDs. Since CD boundaries were always respected, no study was necessary for them. Only the sampled portions of geographic areas were considered in verifying the respect for boundaries. Geographic areas which did not contain any sampled EAs were excluded from the study.

Table 3 shows how well the boundaries of CTs, CSDs and FEDs were respected by WAs. The first column shows the percentage of geographic areas which contained only entire WAs. The second column shows the percentage of geographic areas which were too small to form entire WAs, but were completely contained within one WA. The third column shows the percentage which contained parts of different WAs.

Table 3. Extent to Which Weighting Areas Respected Various Geographic Boundaries

Geographic Areas	Contained Only Entire WAs	Contained Entirely Within One WA	Contained Parts of Different WAs
		Percentage	
Census Divisions	100	0	0
Census Tracts	58	31	11
Census Subdivisions in Census-tracted Areas	59	32	9
Census Subdivisions in Non-census-tracted Areas	8	84	8
Federal Electoral Districts	17	0	83

Table 5. Frequency of Discarding WA Level Constraints in 1991

Constraint	Small	LD	NLD	Outlier	Total	Constraint	Small	LD	NLD	Outlier	Total
FAMCHGE4	5109	609	11	0	5729	AGEHM74	495	18	720	1256	2489
AGEC1718	1496	4181	22	12	5711	FAMCHLD0	1	24	1946	363	2334
HHSIZEG6	3791	1636	257	18	5702	AGE9	68	814	16	1070	1968
AGEC1517	5257	64	302	43	5666	HHSIZE4	56	149	841	912	1958
MOVABLE	4838	432	114	138	5522	AGE49	3	117	231	1589	1940
SEP	1307	3892	86	226	5511	AGE24	1	208	261	1401	1871
AGEHM24	2939	1571	896	56	5462	CHILD	0	1237	290	269	1796
HHSIZE1	176	371	4875	13	5435	AGEHM64	53	34	313	1360	1760
APT5PL	4046	796	186	49	5077	AGE74	200	0	270	1156	1626
LONEPARF	629	49	3725	618	5021	AGEHM54	3	35	269	1120	1427
AGEHM75P	1690	618	2435	195	4938	AGE44	1	5	84	1316	1406
AGEC014	865	1	3497	574	4937	AGE29	2	44	244	1078	1368
FAMCHLD3	588	2136	1626	601	4851	SINGDET	316	548	344	147	1355
AGEC617	3315	12	625	655	4607	AGEHM34	11	8	297	859	1175
NONMEMB	1	4463	38	14	4516	AGE39	1	5	54	1036	1096
AGE75PL	1108	2369	626	323	4426	AGECLE17	7	1	746	310	1064
AGECGE18	41	41	3726	608	4416	HHSIZE2	1	2	580	461	1044
FAMCHLD1	2	20	3804	255	4081	AGE34	1	25	30	896	952
HHSIZE5	515	1013	845	1288	3661	FAMCHLD2	50	113	120	546	829
AGE64	242	446	1781	966	3435	AGEHM44	1	0	4	620	625
OTHDWLS	577	1519	1012	261	3369	MARRIED	0	1	63	266	330
AGE4	34	2546	6	687	3273	CENFAM	0	0	243	55	298
AGE54	12	478	1578	1185	3253	OWNED	22	1	60	200	283
AGE14	75	2286	4	739	3104	SINGLE	0	0	7	268	275
HHSIZE3	3	66	2189	670	2928	HUSBAND	0	28	54	116	198
AGE59	74	496	701	1526	2797	MALEGE15	0	0	104	51	155
WIDOWED	254	386	155	1857	2652	MALEHM	0	0	0	85	85
AGECLE5	71	0	609	1935	2615	MALE	0	0	0	11	11
DIVORCED	185	42	244	2124	2595	TOTPERS	0	0	0	5	5
AGE19	18	409	894	1267	2588	TPERGE15	0	0	0	5	5
AGEC614	96	0	607	1847	2550	TOTHHLD	0	0	0	0	0

One of the aims of the weighting procedure is to minimize the discrepancies between population counts and the corresponding sample estimates for the constraints. These discrepancies are the result of sampling variability and bias (see Chapter V). Even after the weighting procedure is completed, however, some discrepancies may remain. Discrepancies are measured by the difference between the sample estimate and the population count, expressed as a percentage of the population count, i.e.

$$\text{discrepancy} = \frac{\text{sample estimate} - \text{population count}}{\text{population count}} \times 100 \qquad (2)$$

The numerator of the above expression (sample estimate - population count) is often referred to as the "difference".

Table 6 shows the differences (DIFF) and discrepancies (DISC) at the Canada level in 1991 for the 62 constraints. It should be noted that DISCs have been rounded to two decimal places. All of these characteristics were WA level constraints that were used in determining GLSEP weights in 1991. The sample estimates and population counts are based on occupied private dwellings in sampled EAs. The same abbreviations for the constraints that were used in Table 5 and which were defined in Appendix A are used in this table.

Table 6 shows that 4 out of the 62 constraints had a DISC of 0.00. These four constraints were the ones that were the four least frequently discarded constraints, as was shown in Table 5. Table 6 shows that 24 constraints underestimated the population for the constraint, while 34 constraints overestimated the population. Constraints with the largest underestimates in percentage terms were HHSIZEG6 (-3.21), FAMCHGE4 (-2.71) and MOVABLE (-0.78), while constraints with the largest overestimates were AGEC617 (1.69), FAMCHLD3 (1.52) and HHSIZE5 (1.26). All of these constraints with large underestimates and large overestimates do not apply to a large proportion of the population, and they were among the constraints that were the most frequently discarded, as illustrated in Table 5.

Table 6. 1991 Estimate/Population Discrepancies at the Canada Level

Constraint	DIFF	DISC (%)	Constraint	DIFF	DISC (%)	Constraint	DIFF	DISC (%)
TOTPERS	-150	0.00	DIVORCED	-4131	-0.33	AGEHM24	581	0.13
TPERGE15	-135	0.00	WIDOWED	-6695	-0.55	AGEHM34	111	0.01
MALE	-396	0.00	SEP	3708	0.64	AGEHM44	3430	0.15
MALEGE15	-1022	-0.01	CENFAM	-438	-0.01	AGEHM54	5857	0.36
AGE4	-2151	-0.12	NONMEMB	-9916	-0.24	AGEHM64	1582	0.12
AGE9	-1789	-0.10	HUSBAND	630	0.01	AGEHM74	-6122	-0.53
AGE14	3925	0.21	CHILD	9574	0.11	AGEHM75P	-5439	-0.73
AGE19	8705	0.48	LONEPARF	1927	0.25	FAMCHLD0	-8031	-0.31
AGE24	4890	0.26	TOTHHLD	0	0.00	FAMCHLD1	-1874	-0.10
AGE29	-8762	-0.38	OWNED	1039	0.02	FAMCHLD2	4637	0.24
AGE34	580	0.02	MALEHM	-1616	-0.02	FAMCHLD3	10277	1.52
AGE39	-3777	-0.17	SINGDET	316	0.01	FAMCHGE4	-5435	-2.71
AGE44	1278	0.06	MOVABLE	-1358	-0.78	AGECHLE5	-1163	-0.12
AGE49	2665	0.17	APT5PL	-313	-0.03	AGEC614	-4299	-0.45
AGE54	3122	0.24	OTHDWLS	1354	0.04	AGEC1517	-1418	-0.65
AGE59	1639	0.14	HHSIZE1	-14571	-0.65	ACEC014	1524	0.26
AGE64	1005	0.09	HHSIZE2	3250	0.10	AGEC617	5497	1.69
AGE74	-4312	-0.24	HHSIZE3	6227	0.36	AGECLE17	892	0.03
AGE75P	-7169	-0.68	HHSIZE4	6158	0.35	AGECGE18	3976	0.35
MARRIED	4927	0.04	HHSIZE5	9029	1.26	AGEC1718	2739	0.56
SINGLE	2041	0.02	HHSIZEG6	-10092	-3.21			

A study was done comparing the absolute differences between sample estimates and population counts for 62 characteristics in 1991 and 1986 for various geographical levels. The 62 characteristics that were part of this study of absolute differences are listed in Appendix B. The results of the study are summarized in the Table 7 that follows. The table contains the percentage of characteristics that had an "R value" within a certain range for the six geographical levels that are denoted in the table. An R value is a ratio between 1991 and 1986 differences, as the following equation shows:

$$R = 100 * \frac{\sum_{1}^{N_{91}} |\hat{X}^{91} - X^{91}| \; / \; \sum_{1}^{N_{91}} X^{91}}{\sum_{1}^{N_{86}} |\hat{X}^{86} - X^{86}| \; / \; \sum_{1}^{N_{86}} X^{86}} \qquad (3)$$

where X^{91} and X^{86} are, respectively, the 1991 and 1986 population counts for a given characteristic. The sample estimate in 1991 based on GLSEP weights is $\hat{X}^{91}$, while the sample estimate in 1986 based on RREP weights is $\hat{X}^{86}$. R values were calculated for each of the six geographic levels (EA, WA, CSD, CD, PROV., and Canada). The sum of the absolute values of the population/estimate differences was calculated, where N_{91} equals the number of areas for the particular geographical level in 1991 and N_{86} equals the number of areas for the particular geographical level in 1986. An R value in the range of 95 to 105 means that the 1991 estimation system and 1986 estimation system performed almost equally. An R value less than 95 means that the 1991 system performed better than the 1986 system for the characteristic at the particular geographical level, while an R value greater than 105 means that it did worse. Table 7 also gives the percentages of the 37 person characteristics, and of the 25 household characteristics, that fall within these three ranges of R values. For more information on this study, see Majkowski (1994).

Table 7. Percentage of the Characteristics with R Values Falling in Certain Ranges

Characteristics	R Value	EA	WA	CSD	CD	PROV.	Canada
Person (37)	< 95	84	51	76	41	22	22
	95-105	16	14	10	16	19	14
	> 105	0	35	14	43	59	65
Household (25)	< 95	92	68	88	56	44	40
	95-105	4	8	4	20	8	4
	> 105	4	24	8	24	48	56
All (62)	< 95	87	58	81	47	31	29
	95-105	11	11	8	18	14	10
	> 105	2	31	11	35	55	61

Table 7 shows that for the 62 characteristics combined, 87% of them had an R value less than 95% at the EA level. Only 2% (or one characteristic) had an R value greater than 105% at the EA level. This shows that the 1991 estimation system was effective at reducing the population/estimate differences at the EA level compared to the 1986 estimation system. However, as the table shows, this effectiveness of the 1991 estimation system continually decreases as the geographical levels become larger. At the provincial and Canada levels, the percentage of characteristics having an R value greater than 105 is over 50%. A similar pattern is also present for the person and household characteristics when these are studied separately. In comparing the two sets of characteristics, the table shows that a smaller percentage of household characteristics have an R value greater than 105 compared to the person characteristics. This is true for all geographical levels except for the EA level, where there is a difference of 4%, or one household characteristic with an R value greater than 105.

The results that are displayed in Table 7 indicate that the positive and negative differences that result at the WA level do not cancel out as well in 1991 as they did in 1986 when these differences are summed to higher geographical levels (CD, PROV., and Canada levels). The 1991 differences in percentage terms at these higher geographical levels, however, are still very small.

VII. Sample Estimate and Population Count Consistency

In order for the GLSEP to work well, some of the constraints had to be discarded within each WA before the weights could be calculated. Consequently, many important characteristics were discarded in a number of WAs. As a result, the level of agreement (consistency) between sample estimates and population counts for these characteristics was reduced. Furthermore, many geographic areas of interest do not always consist of complete WAs (see Chapter VI, Section A). Consequently, in these areas the consistency for all characteristics depends on how close the areas come to consisting of complete WAs.

The consistency study examined the discrepancies between sample estimates and population counts (expressed as percentages of the population counts) for the same basic set of 53 characteristics as used in the Sampling Bias Study (see Chapter V) for the following geographic areas:

(a) census divisions;

(b) census subdivisions;

(c) census tracts and provincial census tracts;

(d) enumeration areas.

Appendix B contains the list of characteristics whose discrepancies are studied in this chapter. As in Chapter VI, Section B, the discrepancies between sample estimates and population counts were calculated as follows:

$$\text{discrepancy} = \frac{\text{sample estimate} - \text{population count}}{\text{population count}} \times 100$$

A. Census Divisions (CDs)

The percentiles in Table 8 summarize the level of consistency for all 284 sampled CDs in Canada for a wide variety of basic characteristics with a population count[10] greater than 50. Generally, the discrepancies (either positive or negative) produced for characteristics with population counts $\leq$ 50 for most geographic areas were found to be relatively large. Therefore, it was decided to not include geographic areas where the characteristic count was less than or equal to 50 because a few of these areas could significantly alter the percentiles of discrepancies in Tables 8 through 12. This would occur if many of these areas had either relatively large positive discrepancies or relatively large negative discrepancies. In Table 8, for each characteristic, N% of the CDs had discrepancies that were less than the Nth percentile while 100 - N% of the CDs had discrepancies that were greater than the Nth percentile. Thus, the discrepancy was between the 10th and 90th percentiles for 80% of the CDs, was between the 25th and 75th percentiles for 50% of the CDs, etc. For example, the discrepancy for age 0-4 was between -2.86% and 2.38% for 80% of the CDs.

10 The population count here refers to that of the characteristic. For example, the level of consistency for age 0-4 is summarized for all CDs in which there were more than 50 people in the age group 0-4. The same definition applies to Tables 9, 10, 11, and 12.

Table 8. Percentiles of Sample Estimates and Population Count Discrepancies (as a Percentage of the Population Count) for CDs - 1991 and 1986 Censuses

Characteristics Studied	1991 Percentiles of Discrepancies					1986	
	10th	25th	50th	75th	90th	10th	90th
Person Characteristics							
Males	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Females	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Total Person Population	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Age 0-4	-2.86	-1.07	0.00	1.06	2.38	0.00	0.00
Age 5-9	-2.14	-1.05	0.00	0.63	2.06	0.00	0.00
Age 10-14	-1.80	-0.56	0.00	0.99	2.36	0.00	0.00
Age 15-19	-1.87	-0.51	0.55	1.76	3.15	-0.88	1.02
Age 20-24	-3.71	-0.95	0.32	2.35	4.14	-1.35	1.15
Age 25-29	-3.07	-1.40	-0.20	0.20	1.78	-4.12	3.52
Age 30-34	-1.67	-0.39	0.00	0.62	2.11	-3.58	3.73
Age 35-39	-3.14	-0.84	0.00	0.59	2.26	-4.03	3.46
Age 40-44	-2.33	-0.69	0.00	1.02	3.04	-5.07	5.89
Age 45-49	-2.95	-1.14	0.00	1.80	4.26	-4.82	4.98
Age 50-54	-4.96	-2.14	0.13	2.02	5.03	-5.43	5.17
Age 55-59	-6.13	-2.33	0.00	1.60	4.21	-4.47	5.69
Age 60-64	-3.69	-1.75	0.07	1.93	4.88	-5.28	4.91
Age 65-74	-2.28	-1.08	0.00	0.59	2.03	-2.55	4.06
Age > 74	-7.87	-3.66	-1.07	0.67	4.65	-7.49	5.19
Single Persons	-0.08	0.00	0.00	0.00	0.12	-0.28	0.29
Married Persons	0.00	0.00	0.00	0.08	0.33	-0.33	0.20
Widowed Persons	-4.22	-2.26	-0.55	0.57	2.27	-4.49	5.55
Divorced Persons	-4.47	-1.82	-0.14	1.88	4.73	-8.43	9.46
Separated Persons	-9.33	-3.96	0.44	4.90	11.12	-10.60	10.61
Family Characteristics							
Total Census Families	-0.01	0.00	0.00	0.00	0.00	-0.13	0.08
Husband-Wife Families	-0.01	0.00	0.00	0.00	0.07	-0.13	0.10
Lone-parent Census Families	-0.80	0.00	0.00	0.00	0.16	-0.14	0.04
Census Family Children	-0.08	0.00	0.03	0.21	0.42	-0.07	0.21
People in Census Families	-0.04	0.00	0.02	0.09	0.18	0.00	0.02
People Not in Census Families	-1.42	-0.63	-0.14	0.00	0.32	-0.11	0.03
Household and Dwelling Characteristics							
Owned Dwellings	-0.05	0.00	0.00	0.00	0.09	0.00	0.00
Rented Dwellings	-0.20	0.00	0.00	0.00	0.08	0.00	0.00
Single-detached Dwellings	-0.07	0.00	0.00	0.00	0.07	-0.53	0.50
Apts with 5 or More Storeys	-6.68	-1.24	0.00	1.01	4.29	-5.36	8.49
Movable Dwellings	-10.64	-3.85	-0.95	1.22	6.02	-11.67	13.54
All Other Dwelling Types	-0.78	-0.18	0.11	0.67	1.84	-3.48	2.39
Total Households	0.00	0.00	0.00	0.00	0.00	0.00	0.00
One-person Households	-2.54	-1.44	-0.60	0.03	0.73	0.00	0.00
Two-person Households	-0.51	-0.04	0.00	0.22	0.87	-1.64	2.17
Three-person Households	-2.43	-0.92	0.21	1.23	3.17	-4.36	3.95
Four-person Households	-1.66	-0.57	0.04	0.92	2.15	-3.33	4.17
Five-person Households	-3.83	-0.59	1.56	4.19	6.79	-7.78	7.22
Six-or-more-person Households	-14.49	-8.69	-4.20	0.18	5.21	-11.41	7.37
Non-census-family Households	-1.18	-0.67	-0.30	0.03	0.45	0.00	0.00
One-census-family Households	-0.28	-0.03	0.19	0.39	0.69	-0.23	0.31
Hhld Maintainers Aged < 25	-9.54	-3.32	-0.09	5.03	12.64	-7.78	6.76
Hhld Maintainers Aged 25-34	-1.40	-0.35	0.00	0.47	1.48	-1.75	1.66
Hhld Maintainers Aged 35-44	-0.48	0.00	0.00	0.38	1.43	-1.98	1.80
Hhld Maintainers Aged 45-54	-1.81	-0.28	0.14	1.11	3.41	-2.53	3.16
Hhld Maintainers Aged 55-64	-2.99	-1.11	0.00	1.05	2.76	-2.63	2.99
Hhld Maintainers Aged 65-74	-2.91	-1.47	-0.34	0.51	2.64	-3.75	5.03
Hhld Maintainers Aged > 74	-8.09	-4.09	-1.00	1.30	4.77	-8.94	5.78
Male Household Maintainers	-0.04	0.00	0.00	0.00	0.00	-0.74	0.42
Female Household Maintainers	0.00	0.00	0.00	0.00	0.12	-1.56	2.73

All CDs consist of complete WAs. Thus the characteristics that were constraints in 1991 which were rarely or never discarded in a WA had nearly perfect consistency at the CD level.[11] These characteristics were: males, total person population, single persons, married persons, total census families, husband-wife families, owned dwellings, total households and male household maintainers. As Table 5 showed, this group of characteristics was not discarded in many of the WAs. Married persons, the constraint which was the most frequently discarded from this group, was discarded in 330 WAs (about 5.8% of all WAs) by the estimation system. The level of consistency for the remaining characteristics was not perfect but was still quite good, except for those characteristics which represent only a small percentage of the population in most CDs, such as separated persons, movable dwellings, six-or-more-person households and household maintainers aged < 25. A general relationship does exist between the discrepancies and the population counts for all characteristics, in that the consistency improves as the population count for the CD increases.

The final two columns of Table 8 give the 10th and 90th percentiles of the 1986 discrepancies for CDs. Of course, the 1986 discrepancies are based on sample estimates that are the result of the RREP that was used in the 1986 Census. Tables 9, 10, 11 and 12 which follow also contain these two columns for the other geographical levels. In comparison to the same percentiles in 1991, the 1991 discrepancies at the CD level are the same or significantly smaller than the 1986 discrepancies for two-thirds of the characteristics in Table 8. The sizes of the discrepancies at the CD level are quite small compared to the discrepancies at other geographical levels that are studied in the sections that follow.

B. Census Subdivisions (CSDs)

Table 9 summarizes the level of consistency between sample estimates and population counts for all sampled CSDs in Canada with a population count for a given characteristic which is greater than 50. It includes the same characteristics as Table 8. CSDs do not always consist uniquely of complete WAs. They are also much smaller on average than CDs. Consequently, the consistency was not as good for CSDs as for CDs. In general, as with CDs, the consistency improved as the population count for the CSD increased, for all characteristics. In comparison to the 1986 discrepancies for the 10th and 90th percentiles, the 1991 discrepancies are dramatically smaller for many of the characteristics.

C. Census Tracts (CTs) and Provincial Census Tracts (PCTs)

Table 10 summarizes the level of consistency for all sampled CTs in Canada and Table 11 summarizes the level of consistency for all sampled PCTs in Canada. Both tables only include CTs or PCTs where population counts for the characteristic were greater than 50. Both CTs and PCTs also have larger populations on average than CSDs. PCTs have slightly larger populations on average than CTs; however, CT boundaries were respected better than PCT boundaries were when WAs were formed. In both 1991 and 1986, the consistency for CTs was consequently better than for PCTs for most characteristics, while the consistency for PCTs was better than for CSDs for most characteristics. The characteristics for which this was not true were generally those with poor consistency at all geographic levels. In comparison to the 1986 discrepancies at the 10th and 90th percentiles, the 1991 discrepancies are again dramatically smaller for many of the characteristics at both the CT and PCT levels.

11 Even for characteristics with perfect consistency, published tabulations of basic characteristics based on sample data will not agree exactly with tabulations of the same characteristics based on 100% data. This is because those residents of collective dwellings who were not asked the sample questions (see Chapter II, Section B) are included in tabulations based on 100% data, but are excluded from tabulations based on sample data.

Table 9. Percentiles of Sample Estimates and Population Count Discrepancies (as a Percentage of the Population Count) for CSDs – 1991 and 1986 Censuses

Characteristics Studied	1991 Percentiles of Discrepancies					1986	
	10th	25th	50th	75th	90th	10th	90th
Person Characteristics							
Males	-5.47	-1.69	0.00	1.62	5.05	-9.15	9.76
Females	-5.49	-1.83	0.00	1.83	5.56	-9.38	8.97
Total Person Population	-3.36	-0.37	0.00	0.26	3.28	-7.56	7.95
Age 0-4	-19.44	-7.14	0.00	6.23	18.50	-19.36	17.63
Age 5-9	-16.53	-6.11	0.00	5.73	17.01	-19.17	18.71
Age 10-14	-16.99	-6.15	0.00	7.06	18.01	-19.24	18.92
Age 15-19	-16.79	-6.03	0.00	6.92	19.04	-21.32	20.48
Age 20-24	-20.98	-7.71	0.00	7.94	21.04	-20.41	20.18
Age 25-29	-20.34	-7.54	0.00	5.96	19.29	-21.36	20.59
Age 30-34	-17.90	-6.21	0.00	6.02	17.39	-21.07	21.74
Age 35-39	-19.22	-6.66	0.00	5.86	18.71	-20.97	20.66
Age 40-44	-19.19	-6.43	0.00	7.42	19.92	-22.08	21.03
Age 45-49	-19.21	-7.37	0.00	8.45	22.15	-20.62	21.18
Age 50-54	-23.11	-9.06	0.00	8.62	21.29	-21.80	22.17
Age 55-59	-22.25	-8.28	0.00	8.34	22.23	-21.48	21.71
Age 60-64	-22.33	-9.43	0.00	9.14	22.91	-21.20	22.18
Age 65-74	-21.21	-8.49	0.00	6.72	19.36	-20.62	23.17
Age > 74	-26.23	-11.99	-1.31	7.74	22.02	-24.90	23.60
Single Persons	-7.26	-2.34	0.00	2.33	7.14	-13.63	13.11
Married Persons	-6.33	-1.99	0.00	2.31	6.98	-8.50	8.72
Widowed Persons	-18.85	-8.35	-0.01	6.03	17.18	-18.21	20.06
Divorced Persons	-19.62	-7.28	0.00	7.72	19.43	-21.57	20.79
Separated Persons	-20.50	-7.85	0.00	8.55	22.51	-22.90	20.28
Family Characteristics							
Total Census Families	-4.42	-1.52	0.00	1.54	4.69	-7.15	7.48
Husband-Wife Families	-4.91	-1.69	0.00	1.77	5.48	-8.19	8.18
Lone-parent Census Families	-9.90	-1.43	0.00	1.88	10.34	-10.39	9.57
Census Family Children	-7.74	-2.33	0.00	2.60	8.25	-14.51	14.45
People in Census Families	-4.96	-1.29	0.00	1.57	5.01	-9.44	9.54
People Not in Census Families	-16.76	-6.51	0.00	4.59	14.28	-19.36	19.04
Household and Dwelling Characteristics							
Owned Dwellings	-4.97	-1.61	0.00	1.73	4.84	-6.74	6.85
Rented Dwellings	-11.45	-2.88	0.00	2.74	9.52	-13.97	13.60
Single-detached Dwellings	-4.22	-1.31	0.00	1.39	3.96	-5.93	5.99
Apts with 5 or More Storeys	-3.52	-0.80	0.00	0.55	4.24	-6.43	7.26
Movable Dwellings	-11.81	-4.43	0.00	3.15	9.29	-12.54	15.80
All Other Dwelling Types	-7.35	-1.69	0.00	2.71	9.81	-11.77	11.77
Total Households	-1.55	0.00	0.00	0.00	1.39	-4.45	4.50
One-person Households	-11.93	-5.07	-0.54	3.51	10.46	-15.74	15.47
Two-person Households	-11.13	-3.71	0.00	4.02	11.56	-15.78	16.38
Three-person Households	-15.33	-5.54	0.00	6.18	17.21	-20.29	20.59
Four-person Households	-14.78	-5.03	0.00	4.44	13.50	-19.22	18.31
Five-person Households	-14.04	-4.94	0.40	8.11	19.45	-21.86	22.06
Six-or-more-person Households	-20.84	-10.41	-3.38	4.13	12.07	-25.20	22.42
Non-census-family Households	-10.37	-3.84	-0.29	3.16	9.27	-15.97	16.09
One-census-family Households	-4.72	-1.58	0.21	2.09	5.47	-7.10	7.42
Hhld Maintainers Aged < 25	-15.86	-7.74	0.00	8.78	23.06	-19.76	16.65
Hhld Maintainers Aged 25-34	-14.42	-4.73	0.00	4.43	13.73	-16.94	15.25
Hhld Maintainers Aged 35-44	-12.84	-4.01	0.00	5.00	14.29	-15.76	15.05
Hhld Maintainers Aged 45-54	-15.87	-4.83	0.00	5.98	17.57	-16.63	15.60
Hhld Maintainers Aged 55-64	-18.33	-6.89	0.00	6.62	18.45	-17.44	18.62
Hhld Maintainers Aged 65-74	-18.01	-6.98	0.00	5.84	16.95	-17.97	19.44
Hhld Maintainers Aged > 74	-20.81	-9.52	-0.42	7.08	17.22	-21.07	20.86
Male Household Maintainers	-5.23	-1.65	0.00	1.75	5.24	-7.63	7.22
Female Household Maintainers	-10.58	-3.02	0.00	3.06	10.82	-15.84	18.12

Table 10. Percentiles of Sample Estimates and Population Count Discrepancies (as a Percentage of the Population Count) for CTs – 1991 and 1986 Censuses

Characteristics Studied	1991 Percentiles of Discrepancies					1986	
	10th	25th	50th	75th	90th	10th	90th
Person Characteristics							
Males	-0.88	0.00	0.00	0.00	0.87	-2.13	1.94
Females	-0.90	0.00	0.00	0.00	0.79	-1.95	1.94
Total Person Population	-0.24	0.00	0.00	0.00	0.24	-1.67	1.53
Age 0-4	-10.77	-1.16	0.00	1.03	9.77	-6.50	6.83
Age 5-9	-10.97	-1.55	0.00	0.94	10.16	-6.80	6.68
Age 10-14	-11.19	-1.34	0.00	2.23	13.00	-6.80	6.48
Age 15-19	-10.92	-2.72	0.00	3.93	11.62	-6.86	7.03
Age 20-24	-10.09	-1.48	0.00	1.90	10.82	-6.79	6.92
Age 25-29	-9.97	-1.57	0.00	0.28	7.31	-12.80	12.15
Age 30-34	-7.78	-0.57	0.00	0.29	7.46	-12.59	12.74
Age 35-39	-9.25	-1.32	0.00	0.00	7.88	-13.00	12.62
Age 40-44	-10.00	-1.43	0.00	1.02	9.05	-14.42	14.56
Age 45-49	-12.65	-3.06	0.00	3.06	12.00	-15.39	15.01
Age 50-54	-14.93	-5.00	0.00	5.84	16.01	-15.88	15.84
Age 55-59	-16.32	-4.96	0.00	5.84	17.34	-16.05	16.05
Age 60-64	-17.42	-6.53	0.00	6.54	18.61	-16.96	17.34
Age 65-74	-14.16	-3.00	0.00	1.61	12.41	-12.49	12.89
Age > 74	-21.54	-8.89	0.00	8.23	21.00	-21.43	20.62
Single Persons	-1.21	0.00	0.00	0.00	1.27	-2.85	2.83
Married Persons	-1.42	0.00	0.00	0.00	1.59	-2.32	2.02
Widowed Persons	-15.41	-5.27	0.00	3.87	14.87	-16.21	17.02
Divorced Persons	-15.84	-5.22	0.00	4.08	14.20	-21.36	21.80
Separated Persons	-23.67	-8.79	0.00	8.86	25.51	-24.42	24.41
Family Characteristics							
Total Census Families	-1.08	0.00	0.00	0.00	1.01	-2.03	1.71
Husband-Wife Families	-1.17	0.00	0.00	0.00	1.24	-2.22	2.06
Lone-parent Census Families	-6.03	0.00	0.00	0.00	5.40	-5.68	6.15
Census Family Children	-1.77	0.00	0.00	0.26	2.12	-3.28	3.40
People in Census Families	-0.91	0.00	0.00	0.08	1.02	-2.07	1.88
People Not in Census Families	-4.60	-0.36	0.00	0.00	3.24	-4.37	4.26
Household and Dwelling Characteristics							
Owned Dwellings	-1.59	0.00	0.00	0.00	1.53	-1.81	2.03
Rented Dwellings	-2.17	0.00	0.00	0.00	1.90	-2.87	2.73
Single-detached Dwellings	-1.22	0.00	0.00	0.11	1.32	-2.90	2.97
Apts with 5 or More Storeys	-2.42	0.00	0.00	0.16	2.22	-7.15	8.32
Movable Dwellings	-13.00	-2.87	0.00	1.96	9.31	-10.68	10.77
All Other Dwelling Types	-2.08	-0.13	0.00	0.39	2.40	-5.94	5.44
Total Households	-0.01	0.00	0.00	0.00	0.03	-1.05	1.03
One-person Households	-7.60	-3.17	-0.36	1.38	5.39	-4.58	4.36
Two-person Households	-3.90	-0.32	0.00	0.40	4.23	-6.90	7.70
Three-person Households	-8.90	-2.30	0.00	3.32	9.95	-14.29	13.32
Four-person Households	-7.58	-0.90	0.00	2.19	9.90	-14.18	14.13
Five-person Households	-15.85	-4.81	0.00	7.47	18.54	-22.73	22.91
Six-or-more-person Households	-23.65	-11.04	-0.54	7.11	16.34	-31.21	26.78
Non-census-family Households	-5.70	-2.20	-0.29	1.29	4.08	-4.34	3.95
One-census-family Households	-2.09	-0.79	0.24	1.23	2.43	-2.10	2.08
Hhld Maintainers Aged < 25	-19.94	-8.97	0.00	8.95	20.02	-22.02	20.37
Hhld Maintainers Aged 25-34	-6.78	-0.63	0.00	0.63	6.50	-7.67	7.84
Hhld Maintainers Aged 35-44	-5.31	0.00	0.00	0.00	5.33	-8.45	8.15
Hhld Maintainers Aged 45-54	-7.89	-0.57	0.00	1.79	9.10	-9.73	9.66
Hhld Maintainers Aged 55-64	-10.30	-1.92	0.00	2.44	11.77	-10.65	11.33
Hhld Maintainers Aged 65-74	-14.99	-5.06	0.00	2.94	11.90	-13.86	14.72
Hhld Maintainers Aged > 74	-19.24	-7.33	0.00	6.87	17.96	-21.14	20.26
Male Household Maintainers	-1.25	0.00	0.00	0.00	1.17	-2.89	2.53
Female Household Maintainers	-2.29	0.00	0.00	0.00	2.67	-6.77	7.09

Table 11. Percentiles of Sample Estimates and Population Count Discrepancies (as a Percentage of the Population Count) for PCTs – 1991 and 1986 Censuses

Characteristics Studied	1991 Percentiles of Discrepancies					1986	
	10th	25th	50th	75th	90th	10th	90th
Person Characteristics							
Males	-1.48	-0.58	0.00	0.54	1.33	-2.55	2.68
Females	-1.44	-0.56	0.00	0.59	1.54	-2.71	2.60
Total Person Population	-0.84	-0.26	0.00	0.29	0.87	-2.11	2.06
Age 0-4	-10.98	-4.99	0.00	4.02	10.23	-8.56	8.72
Age 5-9	-9.57	-4.01	0.00	3.90	9.46	-8.47	8.82
Age 10-14	-9.53	-3.86	0.00	4.22	9.91	-8.59	8.60
Age 15-19	-8.88	-3.78	0.00	5.00	11.54	-8.64	8.78
Age 20-24	-13.17	-5.47	0.00	5.75	14.26	-9.80	9.41
Age 25-29	-11.71	-4.77	0.00	3.41	9.43	-12.54	12.62
Age 30-34	-9.22	-3.35	0.00	3.69	9.88	-12.24	12.01
Age 35-39	-10.28	-4.22	0.00	3.75	9.92	-12.40	13.11
Age 40-44	-10.20	-4.36	0.00	4.64	11.19	-13.88	14.99
Age 45-49	-12.13	-5.43	0.00	6.27	13.76	-15.71	15.43
Age 50-54	-16.36	-7.16	0.00	7.58	16.29	-15.90	16.13
Age 55-59	-16.66	-7.56	0.00	6.82	15.40	-16.24	16.60
Age 60-64	-16.11	-7.68	0.00	7.49	16.88	-16.40	16.40
Age 65-74	-11.33	-4.33	0.00	3.82	10.00	-11.13	12.57
Age > 74	-21.49	-10.26	-1.17	6.13	16.49	-21.08	17.98
Single Persons	-1.90	-0.68	0.00	0.80	1.98	-3.86	3.91
Married Persons	-1.67	-0.68	0.00	0.83	1.95	-2.51	2.41
Widowed Persons	-14.28	-6.34	0.00	3.92	10.82	-14.31	14.77
Divorced Persons	-16.70	-7.78	0.00	6.80	16.68	-22.77	25.18
Separated Persons	-23.27	-10.35	0.00	11.45	25.79	-25.85	26.49
Family Characteristics							
Total Census Families	-1.29	-0.52	0.00	0.55	1.25	-2.06	2.19
Husband-Wife Families	-1.45	-0.61	0.00	0.63	1.55	-2.40	2.43
Lone-parent Census Families	-10.56	-3.92	0.00	3.73	10.25	-11.62	10.83
Census Family Children	-2.18	-0.79	0.00	1.08	2.64	-4.25	4.25
People in Census Families	-1.26	-0.44	0.00	0.58	1.32	-2.47	2.59
People Not in Census Families	-6.38	-2.68	0.00	1.74	4.89	-7.20	7.11
Household and Dwelling Characteristics							
Owned Dwellings	-1.53	-0.62	0.00	0.64	1.65	-2.18	2.08
Rented Dwellings	-6.03	-1.77	0.00	1.63	5.32	-6.58	6.40
Single-detached Dwellings	-1.36	-0.49	0.00	0.53	1.31	-2.35	2.45
Apts with 5 or More Storeys	-6.05	-0.60	0.00	0.76	2.22	-11.44	9.99
Movable Dwellings	-14.68	-5.67	0.00	4.85	12.50	-17.31	17.11
All Other Dwelling Types	-6.41	-1.74	0.00	2.41	6.90	-9.60	9.03
Total Households	-0.54	-0.11	0.00	0.14	0.56	-1.34	1.38
One-person Households	-7.11	-3.60	-0.54	2.17	6.28	-8.37	8.05
Two-person Households	-4.65	-1.77	0.00	1.88	4.76	-7.94	8.03
Three-person Households	-10.00	-4.39	0.00	4.65	10.27	-14.18	14.42
Four-person Households	-8.76	-3.57	0.00	3.81	8.51	-11.95	13.06
Five-person Households	-14.77	-5.82	0.88	9.24	18.62	-20.93	19.79
Six-or-more-person Households	-27.14	-15.35	-4.22	7.39	18.29	-27.92	27.64
Non-census-family Households	-5.24	-2.50	-0.30	1.73	4.34	-8.14	7.08
One-census-family Households	-1.70	-0.63	0.20	1.07	2.00	-2.20	2.30
Hhld Maintainers Aged < 25	-23.55	-11.01	-0.50	11.66	25.39	-22.44	20.89
Hhld Maintainers Aged 25-34	-7.61	-3.37	0.00	3.03	7.97	-8.14	8.53
Hhld Maintainers Aged 35-44	-6.44	-2.34	0.00	2.93	7.26	-8.76	9.27
Hhld Maintainers Aged 45-54	-8.78	-3.83	0.00	4.41	10.85	-10.72	10.96
Hhld Maintainers Aged 55-64	-11.71	-5.00	0.00	5.04	11.28	-11.00	11.08
Hhld Maintainers Aged 65-74	-12.21	-5.13	0.00	4.15	11.59	-13.29	15.06
Hhld Maintainers Aged > 74	-21.70	-10.26	-0.87	6.59	16.36	-22.91	18.59
Male Household Maintainers	-1.48	-0.65	0.00	0.59	1.49	-2.77	2.55
Female Household Maintainers	-4.99	-1.77	0.00	2.07	4.95	-8.33	9.09

Table 12. Percentiles of Sample Estimates and Population Count Discrepancies (as a Percentage of the Population Count) for EAs – 1991 and 1986 Censuses

Characteristics Studied	1991 Percentiles of Discrepancies					1986	
	10th	25th	50th	75th	90th	10th	90th
Person Characteristics							
Males	-6.25	-2.59	0.07	2.64	6.06	-12.81	13.01
Females	-6.29	-2.66	-0.08	2.55	6.24	-12.01	12.16
Total Person Population	-3.56	0.00	0.00	0.00	3.35	-9.83	9.99
Age 0-4	-26.37	-12.73	-0.22	12.50	26.49	-29.25	30.17
Age 5-9	-24.17	-11.49	0.03	11.23	24.00	-28.54	28.75
Age 10-14	-25.42	-12.24	0.07	12.23	25.28	-28.48	28.87
Age 15-19	-25.38	-12.06	0.37	13.15	27.12	-29.55	30.57
Age 20-24	-27.94	-13.58	-0.23	13.51	28.48	-29.57	30.39
Age 25-29	-26.08	-12.60	-0.51	11.45	25.23	-30.21	29.88
Age 30-34	-24.43	-11.60	-0.23	11.43	24.63	-30.13	30.59
Age 35-39	-25.09	-12.36	-0.68	11.28	24.60	-29.83	30.18
Age 40-44	-26.31	-12.80	-0.37	11.93	26.53	-31.06	31.66
Age 45-49	-27.42	-13.12	-0.02	13.65	29.16	-31.97	32.29
Age 50-54	-30.00	-15.15	-0.10	14.85	31.77	-31.25	33.71
Age 55-59	-29.56	-15.01	0.29	15.76	31.40	-31.59	34.43
Age 60-64	-30.34	-15.37	-0.50	16.01	31.98	-32.51	35.26
Age 65-74	-26.62	-12.59	-0.63	11.72	27.41	-30.08	31.33
Age > 74	-27.47	-12.72	-0.32	12.50	27.83	-30.54	32.05
Single Persons	-8.45	-3.43	0.04	3.42	8.03	-17.47	17.67
Married Persons	-8.78	-3.55	0.07	3.76	8.92	-13.10	12.95
Widowed Persons	-23.51	-11.54	-0.60	10.68	22.99	-26.94	29.00
Divorced Persons	-29.67	-15.22	-0.93	14.50	29.84	-32.41	35.25
Separated Persons	-28.90	-11.74	3.21	17.38	32.19	-33.44	36.57
Family Characteristics							
Total Census Families	-5.72	-2.53	-0.02	2.48	5.80	-10.62	10.61
Husband-Wife Families	-6.54	-2.95	0.01	2.93	6.60	-12.31	12.16
Lone-parent Census Families	-17.57	-8.93	0.61	10.14	19.57	-26.38	24.83
Census Family Children	-9.98	-3.88	0.19	4.13	9.81	-19.94	20.38
People in Census Families	-6.39	-2.31	0.08	2.37	6.31	-13.30	13.31
People Not in Census Families	-19.17	-8.92	-0.25	7.96	18.09	-24.31	25.02
Household and Dwelling Characteristics							
Owned Dwellings	-7.16	-3.01	0.09	3.02	7.05	-10.15	10.13
Rented Dwellings	-9.94	-3.82	0.11	4.00	9.77	-13.82	13.32
Single-detached Dwellings	-6.17	-2.51	0.02	2.48	6.13	-8.73	8.87
Apts with 5 or More Storeys	-6.37	-2.23	0.13	2.39	6.40	-9.60	9.98
Movable Dwellings	-12.40	-4.74	0.10	5.14	11.19	-14.69	16.40
All Other Dwelling Types	-8.44	-3.25	0.08	3.43	8.48	-12.44	11.88
Total Households	-0.18	0.00	0.00	0.00	0.04	-6.31	6.23
One-person Households	-13.75	-6.67	-0.35	5.69	12.38	-21.47	21.42
Two-person Households	-14.41	-7.08	-0.04	7.06	15.11	-22.07	23.55
Three-person Households	-22.27	-11.37	-0.12	11.25	23.48	-27.75	28.63
Four-person Households	-17.59	-8.41	-0.04	8.49	18.45	-24.95	26.08
Five-person Households	-22.17	-10.99	1.09	13.92	26.33	-30.33	31.62
Six-or-more-person Households	-22.63	-9.30	2.48	14.81	26.63	-38.54	27.49
Non-census-family Households	-11.54	-5.28	-0.04	4.91	11.03	-20.18	20.25
One-census-family Households	-6.78	-2.95	0.34	3.44	7.01	-10.82	11.06
Hhld Maintainers Aged < 25	-21.82	-10.84	0.95	12.93	24.53	-28.01	31.55
Hhld Maintainers Aged 25-34	-17.55	-8.40	0.06	8.43	18.27	-23.63	24.19
Hhld Maintainers Aged 35-44	-18.12	-8.66	-0.11	8.34	18.77	-23.81	24.39
Hhld Maintainers Aged 45-54	-21.43	-10.20	-0.07	10.61	22.95	-25.43	26.29
Hhld Maintainers Aged 55-64	-23.30	-11.39	-0.08	11.35	24.20	-26.66	27.51
Hhld Maintainers Aged 65-74	-19.82	-9.90	-0.51	9.19	19.91	-26.35	28.44
Hhld Maintainers Aged > 74	-18.88	-8.90	-0.08	9.68	19.34	-25.04	27.55
Male Household Maintainers	-6.88	-3.19	0.05	3.25	6.84	-11.67	11.46
Female Household Maintainers	-12.61	-6.19	-0.09	6.10	13.10	-21.31	21.89

D. Enumeration Areas (EAs)

EAs are the components of WAs, and WAs are the lowest level at which sample estimates are forced to agree with population counts for most characteristics. EAs are also the components of higher geographical levels (CDs, CSDs, CTs, PCTs, etc.) and a number of the WAs are, as Table 3 earlier showed, components of these higher levels. Consequently, the consistency at the EA level cannot be expected to be as good as that exhibited at higher geographical levels that have been studied. Table 12 confirms this as it shows that for most characteristics studied, in sampled EAs with a population count for the characteristic greater than 50, the discrepancies are larger than the discrepancies for the geographical levels studied earlier. This is the case in both 1991 and 1986. In comparison to the 1986 discrepancies for the 10th and 90th percentiles, the 1991 discrepancies are dramatically lower for the vast majority of the characteristics studied and similar to 1986 for the few remaining others.

A similar study was done earlier for the same geographical levels as above. A total of 68 characteristics, which included all the 53 above, were studied, with discrepancies for both 1991 and 1986 estimates for the 10th, 25th, 50th, 75th and 90th percentiles being produced. For more information on this study, see Majkowski (1992a).

VIII. Sampling Variance

Sampling error can be divided into two components: variance and bias. The variance measures the variability of an estimate about its average value in hypothetical repetitions of the survey process, while the bias is defined as the difference between the average value of the estimate in hypothetical repetitions and the true value being estimated. The mean square error (MSE) measures the variability of the estimate about the true value in hypothetical repetitions of the survey process. It can be shown that the MSE equals the variance plus the square of the bias. The MSE is the most accurate measure of how far the estimate is from the true population value on average. If the bias is small relative to the variance, the variance is a good approximation of the MSE. There is evidence, however, that the bias accumulates as census estimates for progressively larger geographical areas are produced. Thus, the bias can be insignificant for small geographical areas but can become large relative to the variance for large geographical areas. This can result in the variance being much smaller than the MSE for large geographical areas. The variance of an estimate can be estimated from the sample, but the bias of an estimate cannot. This means that it is not possible to estimate the MSE accurately from the sample unless the bias is small relative to the variance.

In previous censuses, a study to provide estimates of the sampling variance was carried out. A few results from the 1986 study are provided in Section A (for more information, see the User's Guide to the Quality of 1986 Census Weighting: Sampling and Weighting). Because it was felt, however, that the sampling variance would not provide an accurate estimate of the MSE for large geographical areas, it was decided not to repeat this study for the 1991 Census. A discussion is given in Section B, however, of what impact the estimation methodology used in the 1991 Census had on the sampling variance compared to the 1986 Census.

A. 1986 Census Sampling Variance Study

Chapter V presented results of the Sampling Bias Study, describing the nature and extent of bias in the census sample prior to weighting. Chapters VI and VII presented results on the sampling bias following the application of the weighting procedure. Even with a perfectly unbiased sampling method, the results would still be subject to variance, simply because the estimates are based only on a sample. The variance may be estimated using the data collected by the sample survey.[12] The 1986 Sampling Variance Study was carried out to estimate the effect of the sampling and estimation procedures on those census figures that are based on sample data.

On the basis of the 2B sample data, thousands of tables are produced by Statistics Canada. Conceptually, a measurement of precision, the estimated sampling variance, can be associated with every estimate calculated in these tables. This measurement takes into account both the sample design and the estimation method. In practice, however, it cannot be calculated for every census estimate because of high data processing costs. Sampling variance is thus estimated for only a subset of census estimates. From this, the combined effect of the sample design and the estimation method on the sampling variance can be estimated. Simple estimates of sampling variance, which are inexpensive to calculate, can then be adjusted for this impact in order to produce estimates of sampling variance for any census estimates.

Table 13 gives non-adjusted (simple) standard errors of census sample estimates. The figures in this table were determined by assuming that the techniques of 1-in-5 simple random sampling and simple weighting by 5 were used. The standard errors are expressed in Table 13 as a function of the size of both the census estimate and the geographic area. For example, for an estimate of 250 persons in a geographic area with a total of 1,000 persons, the non-adjusted standard error is 25.

12 Unfortunately, the sampling variance does not provide any indication of the extent of non-sampling error.

Table 13. Estimates of Standard Errors of Sample Estimates

Estimated Total	Total Number of Persons, Households, Dwellings or Families in the Area								
	500	1,000	2,500	5,000	10,000	25,000	50,000	100,000	250,000
50	15	15	15	15	15	15	15	15	15
100	18	19	20	20	20	20	20	20	20
250	22	25	30	30	30	30	30	30	30
500	0	30	40	40	45	45	45	45	45
1,000		0	50	55	60	60	65	65	65
2,500			0	70	85	95	95	100	100
5,000				0	100	130	130	140	140
10,000					0	150	180	190	200
25,000						0	220	270	300
50,000							0	320	400
100,000								0	490
250,000									0

Estimated Total	Total Number of Persons, Households, Dwellings or Families in the Area					
	500,000	1,000,000	2,500,000	5,000,000	10,000,000	25,000,000
50	15	15	15	15	15	15
100	20	20	20	20	20	20
250	30	30	30	30	30	30
500	45	45	45	45	45	45
1,000	65	65	65	65	65	65
2,500	100	100	100	100	100	100
5,000	140	140	140	140	140	140
10,000	200	200	200	200	200	200
25,000	310	310	310	320	320	320
50,000	420	440	440	440	450	450
100,000	570	600	620	630	630	630
250,000	710	870	950	970	990	990
500,000	0	1,000	1,260	1,340	1,380	1,400
1,000,000		0	1,550	1,790	1,900	1,960
2,500,000			0	2,240	2,740	3,000
5,000,000				0	3,160	4,000
10,000,000					0	4,900

Standard errors are given in Table 13 for only a limited number of values for the estimated total and the total number of persons, households, dwellings or families in the area. The following formula may be used to calculate the non-adjusted standard errors for any estimated total for an area of any size:

$$NASE = \sqrt{\frac{4E(N-E)}{N}} \qquad (4)$$

where NASE is the non-adjusted standard error, E is the estimated total and N is the total number of persons, households, dwellings or families in the area. For example, for an estimated total of 750 persons in an area with a total of 9,000 persons, the non-adjusted standard error would be:

$$\sqrt{\frac{4(750)(9,000 - 750)}{9,000}} = 52$$

It should be noted that if E is very much smaller than N, this will cause N-E to equal N approximately. Then NASE in equation (4) will approximately equal two times the square root of the estimate itself ($NASE \simeq 2\sqrt{E}$).

The 1986 Sampling Variance Study provides adjustment factors[13] by which the non-adjusted standard errors should be multiplied to adjust for the combined effect of the sample design and the estimation procedure. To calculate these adjustment factors, a sample of 401 WAs (out of a total of 5,341 WAs) was selected. The sample was allocated among the ten provinces[14] in such a way as to obtain good estimates of the sampling variance at the provincial level without greatly sacrificing the quality of the estimates at the national level. For each WA in the sample, estimates of the sampling variances for raking ratio estimates were calculated for different categories of all of the characteristics given in Table 9 of the 1986 Census User's Guide. The estimates of sampling variance at the provincial and national levels were obtained by weighting up the WA level estimates. The adjustment factors for each category of each characteristic were calculated by dividing the square roots of these estimates by the non-adjusted standard errors. Adjustment factors were calculated at the provincial and national levels for each characteristic by averaging the adjustment factors for all of its categories. For further information on how these adjustment factors were calculated, see Béland (1990).

To estimate the standard error for a given census sample estimate, the adjustment factor applying to the characteristic was determined from Table 9 of the 1986 Census User's Guide. The adjustment factor at the national or provincial level for sample characteristics was generally in the range 0.40 to 1.60. This factor was then multiplied by the non-adjusted standard error selected in Table 13.

The following example illustrates how to calculate the adjusted standard errors. Suppose the estimate of interest is the immigrant population in Ontario. The 1986 estimate for this characteristic was 2,081,200. The 1986 Census count for the population of Ontario was 9,001,170. Since neither number is very close to any of the values given in Table 13, equation (4), which calculates the non-adjusted standard error, should be used. In this case the result would be 2,530. From Table 9 of the 1986 Census User's Guide, the provincial-level adjustment factor for the characteristic "immigrant" is 1.12. Consequently, the adjusted standard error for this estimate is 2,530 x 1.12 = 2,834.

13 The squares of the adjustment factors are commonly known as "design effects".

14 The Yukon and Northwest Territories were grouped with British Columbia.

A second example, however, casts doubt on the accuracy of these adjusted standard errors as estimates of the square root of the MSE. The estimated number of persons in the 1986 Census with marital status "married" who lived in private dwellings in sampled EAs was 11,771,126. The number of persons enumerated in the 1986 Census who lived in private dwellings in sampled EAs was 24,369,559. Applying equation (4) generates a non-adjusted standard error of 4,934. From Table 9 of the 1986 Census User's Guide, the national-level adjustment factor for the characteristic "married" is 0.25. Consequently, the adjusted standard error for this estimate is 4,934 x 0.25 = 1,233. Because marital status is a basic characteristic, however, it is known that the population count of the number of persons in the 1986 Census with marital status "married" who lived in private dwellings in sampled EAs was 11,778,842. The difference between the estimate and the population count is -7,716. The ratio of this difference to the adjusted standard error is -7,716/1,233 = -6.26. A 95% confidence interval for an estimate would normally be defined as plus or minus two times the adjusted standard error. The fact that the ratio of the difference to the standard error is -6.26 suggests that the adjusted standard error of 1,233 is an underestimate of the square root of the MSE.

B. Sampling Variance and Bias in the 1991 Census

In Bankier, Rathwell and Majkowski (1992), the coefficients of variation (CVs) of the GLSEP for some sample characteristics were compared to the corresponding CVs of the RREP. In both cases, 1986 Census data were used. The CV of an estimate is the square root of the estimated variance expressed as a percentage of the estimate. For 79 WAs, the estimated CVs were calculated for estimates of 507 EA level and 642 WA level sample characteristics (all of which applied to at least an estimated 60 households in the population). The WA level and EA level estimates were each classified into small estimates (less than or equal to the median value of the estimates) and large estimates (greater than the median value of the estimates). It was found that the median value for the CVs for large WA estimates was 5% for the GLSEP, while it was 6% for the RREP. The median value for the CVs for small WA estimates was 13% for the GLSEP, while it was 15% for the RREP. The median value for the CVs for large EA estimates was 10% for the GLSEP, while it was 12.5% for the RREP. The median value for the CVs for small EA estimates was 15% for the GLSEP, while it was 17.5% for the RREP. Thus, there was some reduction in the CVs for the GLSEP, compared to the RREP at both the EA and WA levels. Because the variances at higher geographical levels are just the sum of the variances at the WA level, these reductions in the CVs should also hold at higher geographical levels.

Chapter V indicates that the census sample has small but significant biases. These biases are insignificant compared to the sampling variance at the WA level. For higher geographical levels, however, the bias for a characteristic can accumulate if the bias almost always results in overestimates or underestimates. It appears that the effect of the bias is more significant with the GLSEP than with the RREP. This can be seen from Table 7 of Chapter VI, where the GLSEP has smaller population/estimate differences than the RREP for smaller geographical areas. The opposite situation holds true, however, for larger geographical areas. Besides bias introduced by sampling and processing, Bankier, Rathwell and Majkowski (1992) show in a Monte Carlo study that the GLSEP estimator itself is biased, though the relative bias is less than 1% for 50% of the characteristics studied. More serious, however, is the fact that the estimated variances of GLSEP estimators have a median relative bias of -25% at the WA level. Thus, they tend to underestimate the true variance. The RREP estimators may suffer from similar biases but no study of them has been done.

For the 1996 Census, enhancements will be made to the estimation procedures to reduce the size of population/estimate differences at higher geographical levels. This should allow more accurate estimates of the MSEs of the sample characteristics to be produced.

IX. Conclusion

Sampling is now an accepted and integral part of census-taking. Its use can lead to substantial reductions in costs and respondent burden associated with a census, or alternatively, can allow the scope of a census to be broadened at the same cost. The price paid for these advantages is the introduction of sampling error to census figures that are based on the sample. The effect of sampling is most important for small census figures, whether they are counts for rare categories at the national or provincial level or counts for categories in small geographic areas. It should be noted that response errors and processing errors also contribute to the overall error of census figures and that it is the same small census figures that are particularly susceptible to the effects of these non-sampling errors. Therefore, even with a 100% census, many small figures would be of limited reliability. As a general rule of thumb for the 1991 Census, figures of size 50 or less that are based on sample data are of very low reliability, while figures up to size 500 tend to have standard errors in excess of 10% of their size.

For many of the characteristics, a certain amount of bias was detected in the sample. A portion of the bias was found to have been introduced during data processing and Edit and Imputation. The remaining bias must have been due to one or more factors such as non-response bias, response bias, the selection of a biased sample by the CRs, etc. The procedures for weighting the sample data up to the population level were carried out successfully, and generally achieved the levels of sample estimate and population count consistency anticipated. The consistency that was achieved at the provincial and Canada levels was somewhat lower than expected given the improved consistency for smaller geographical levels that was achieved during testing. This is probably the result of the bias in the sample plus a small amount of additional bias introduced by the estimation procedure itself.

The census estimation methodology will be reassessed for the 1996 Census to see if it is possible to improve sample estimate and population count consistency at the provincial and Canada levels while maintaining good consistency at the EA level. Doing this should also allow more reliable estimates of the mean square error of the census estimates to be produced.

Appendix A

WA and EA Level Constraints Applied to 1991 Census Weights With Short-form Names for the Constraints

Person WA Level Constraints

TOTPERS	–Total persons
TPERGE15	–Total persons aged ≥ 15
MALE	–Males
MALEGE15	–Males aged ≥ 15
AGE4	–Persons aged 0 to 4
AGE9	–Persons aged 5 to 9
AGE14	–Persons aged 10 to 14
AGE19	–Persons aged 15 to 19
AGE24	–Persons aged 20 to 24
AGE29	–Persons aged 25 to 29
AGE34	–Persons aged 30 to 34
AGE39	–Persons aged 35 to 39
AGE44	–Persons aged 40 to 44
AGE49	–Persons aged 45 to 49
AGE54	–Persons aged 50 to 54
AGE59	–Persons aged 55 to 59
AGE64	–Persons aged 60 to 64
AGE74	–Persons aged 65 to 74
AGE75P	–Persons aged ≥ 75
MARRIED	–Married persons
SINGLE	–Single persons
DIVORCED	–Divorced persons
WIDOWED	–Widowed persons
SEP	–Separated persons
CENFAM	–Census families
NONMEMB	–Non-members of census families
HUSBAND	–Husbands
CHILD	–Census family children
LONEPARF	–Lone-parent females

EA Level Constraints

HHEACT	–Total households in EA
PPEACT	–Total persons in EA

Household WA Level Constraints

TOTHHLD	–Total households
OWNED	–Owned dwellings
MALEHM	–Households with male household maintainer
SINGDET	–Single-detached dwellings
MOVABLE	–Movable dwellings
APT5PL	–Apartments in a building with more than 5 storeys
OTHDWLS	–All other types of dwellings
HHSIZE1	–Households of size 1
HHSIZE2	–Households of size 2
HHSIZE3	–Households of size 3
HHSIZE4	–Households of size 4
HHSIZE5	–Households of size 5
HHSIZEG6	–Households of size ≥ 6
AGEHM24	–Households with household maintainer aged ≤ 24
AGEHM34	–Households with household maintainer aged 25 to 34
AGEHM44	–Households with household maintainer aged 35 to 44
AGEHM54	–Households with household maintainer aged 45 to 54
AGEHM64	–Households with household maintainer aged 55 to 64
AGEHM74	–Households with household maintainer aged 65 to 74
AGEHM75P	–Households with household maintainer aged ≥ 75
FAMCHLD0	–Census families with no children at home
FAMCHLD1	–Census families with one child at home
FAMCHLD2	–Census families with two children at home
FAMCHLD3	–Census families with three children at home
FAMCHGE4	–Census families with ≥ 4 children at home
AGECLE5	–Census families with all children at home aged ≤ 5
AGEC614	–Census families with all children at home aged 6 to 14
AGEC1517	–Census families with all children at home aged 15 to 17
AGEC014	–Census families with some children at home aged ≤ 5 and the rest aged 6 to 14
AGEC617	–Census families with some children at home aged 6 to 14 and the rest aged 15 to 17
AGECLE17	–Census families with all children at home aged ≤ 17
AGECGE18	–Census families with all children at home aged ≥ 18
AGEC1718	–Census families with some children at home aged ≤ 17 and the rest aged ≥ 18

Appendix B

Additional Characteristics Studied

For the Sampling Bias and Sample Estimate Consistency studies of Chapters V and VII, a set of 53 characteristics was used. These 53 characteristics consisted of 46 of the 62 constraints from Appendix A (TPERGE15, MALEGE15, LONEPARF, plus the 13 constraints whose alphanumeric codes begin with the letters FAMCH and AGEC, were excluded) as well as the 7 characteristics listed below:

- females;
- lone parents;
- census family persons;
- rented dwellings;
- non-census-family households;
- one-census-family households;
- female household maintainers.

For the study of absolute differences (Table 7 of Chapter VI), 62 characteristics were studied. Besides the above 53 characteristics, the three constraints TPERGE15, MALEGE15, LONEPARF from Appendix A, plus the characteristics

- "married or separated persons";
- "lone-parent males";
- "females aged 15 or above";
- "non-members aged 15 or above";
- "children aged 15 or above"; and
- "households with household maintainer aged > 64";

were included.

Some of the constraints excluded from these studies were not used as constraints in the 1986 Census. Thus, it was felt that the results of different censuses would be more comparable if they were excluded. Some of the additional characteristics added were "indirectly" constraints in 1991, as they were linearly dependent on characteristics that were constraints in 1991. For example, the characteristic "females" was linearly dependent on those for "males" and "total persons".

Appendix C

Additional Information on Statistics Used in Sampling Bias Study

Let X represent the known value for a 2A characteristic at the census division (CD) level and let $\hat{X}^{(0)}$ represent the Horvitz-Thompson estimator of X. $\hat{X}^{(0)}$ was calculated by multiplying the unweighted sample total for the characteristic of each sampled EA by the inverse of the realized household sampling fraction for the EA, and then summing the results to the CD level. Non-sampled EAs were excluded from the analysis. The standard deviation of $\hat{X}^{(0)}$, $\text{std}(\hat{X}^{(0)}) = \sqrt{V(\hat{X}^{(0)})}$, was calculated under the assumption that simple random samples of households were drawn independently in each EA (in fact, independent systematic random samples were drawn). Consequently, the variances were calculated at the EA level and summed to the CD level. The population S^2 values which measure the variance in the population counts were used in the variance calculations. See Cochran (1977) pp. 23-24 for variance formulas for person and family characteristics and pp. 50-52 for variance formulas for household and dwelling characteristics.

Since the $\hat{X}^{(0)}$ values are Horvitz-Thompson estimators, they are unbiased for X. Sampling was done independently in different EAs. Therefore the $\hat{X}^{(0)}$ values are the sum of n independent random variables, where n is the number of EAs in the CD. Since 90 percent of the CDs had more than 25 EAs, with an average number of EAs of 140, n is quite large in most CDs. Thus, according to the central limit theorem, $Z^{(0)} = (\hat{X}^{(0)}-X)/\text{std}(\hat{X}^{(0)})$ should follow an approximately normal(0,1) distribution (see Kendall and Stuart (1963), p. 193). This, however, would not be the case if 2B responses were significantly biased for any reason.

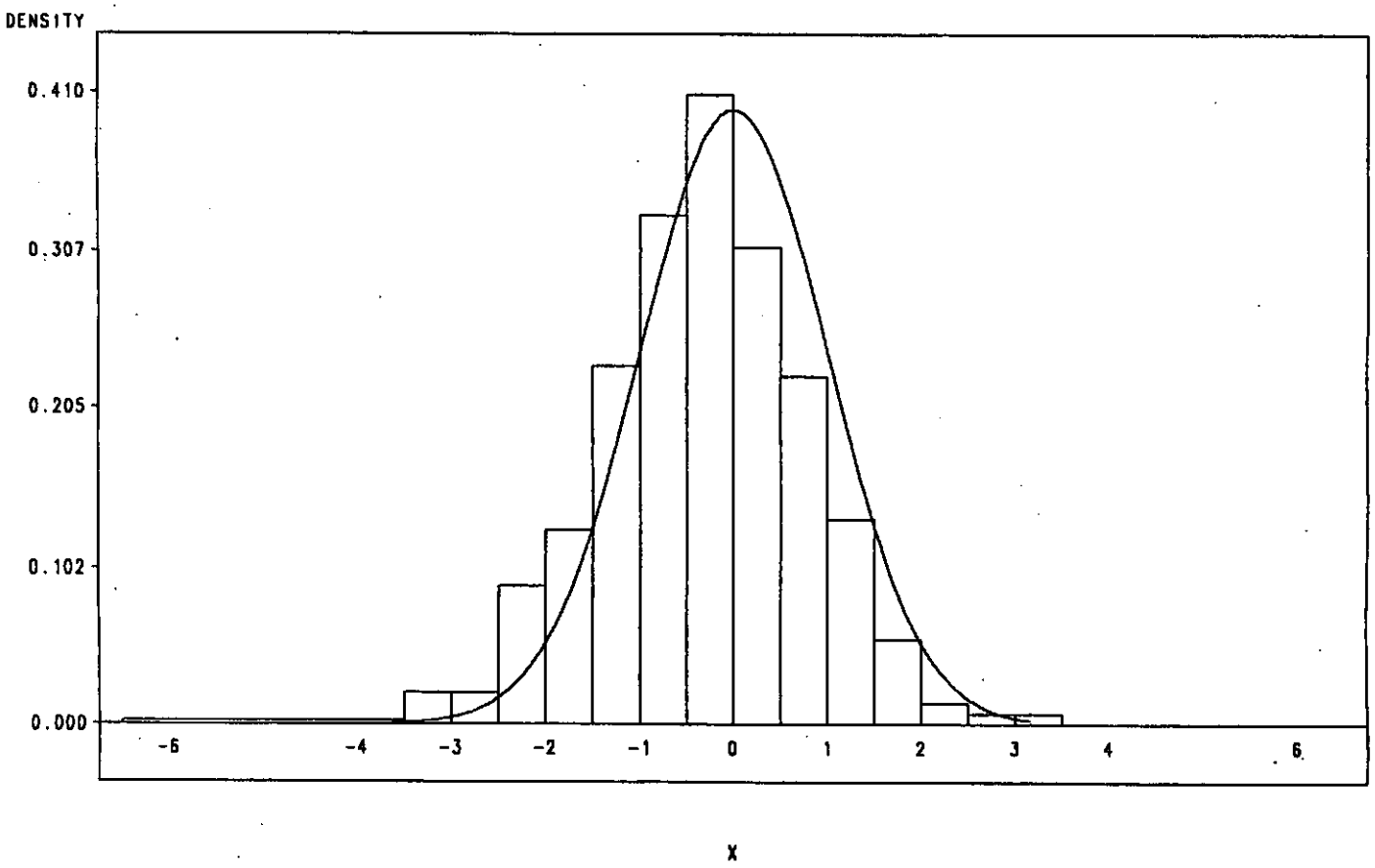
The $Z^{(0)}$ values were produced for all 284 sampled CDs in Canada, for the 2A characteristics given in Chapter V. In order to evaluate the normality of the $Z^{(0)}$ values at the CD level, histograms of the $Z^{(0)}$ values overlaid with a normal Probability Density Function (PDF) were produced. See Appendix D for examples of such plots for two 2A characteristics.

To test whether $Z^{(0)}$ was being selected from a normal distribution whose mean is zero (i.e. the sample selection procedure was unbiased), the mean $\bar{Z}^{(0)} = \sum_{i=1}^{m} Z_i^{(0)}/m$ was calculated where m = 284 (the number of CDs) and $Z_i^{(0)}$ is the value of $Z^{(0)}$ for the i^{th} CD. In addition, the standard deviation of the $Z_i^{(0)}$ was determined, where $\text{std}^2(Z^{(0)}) = \sum_{i=1}^{m}(Z_i^{(0)}-\bar{Z}^{(0)})/(m-1)$. The T statistic $T_Z = \sqrt{m}\,\bar{Z}^{(0)}/\text{std}(Z^{(0)})$ was then calculated. If the sample selection procedure was unbiased, T should follow Student's t distribution with m-1 degrees of freedom. The probability of $|T_Z| > 1.960$ if the sample selection procedure was unbiased is less than 0.05. Thus, if $|T_Z| > 1.960$, the hypothesis that the sample selection procedure was unbiased will be rejected and the difference between the sample estimate and the population count will be said to be statistically significant at the 5% level.

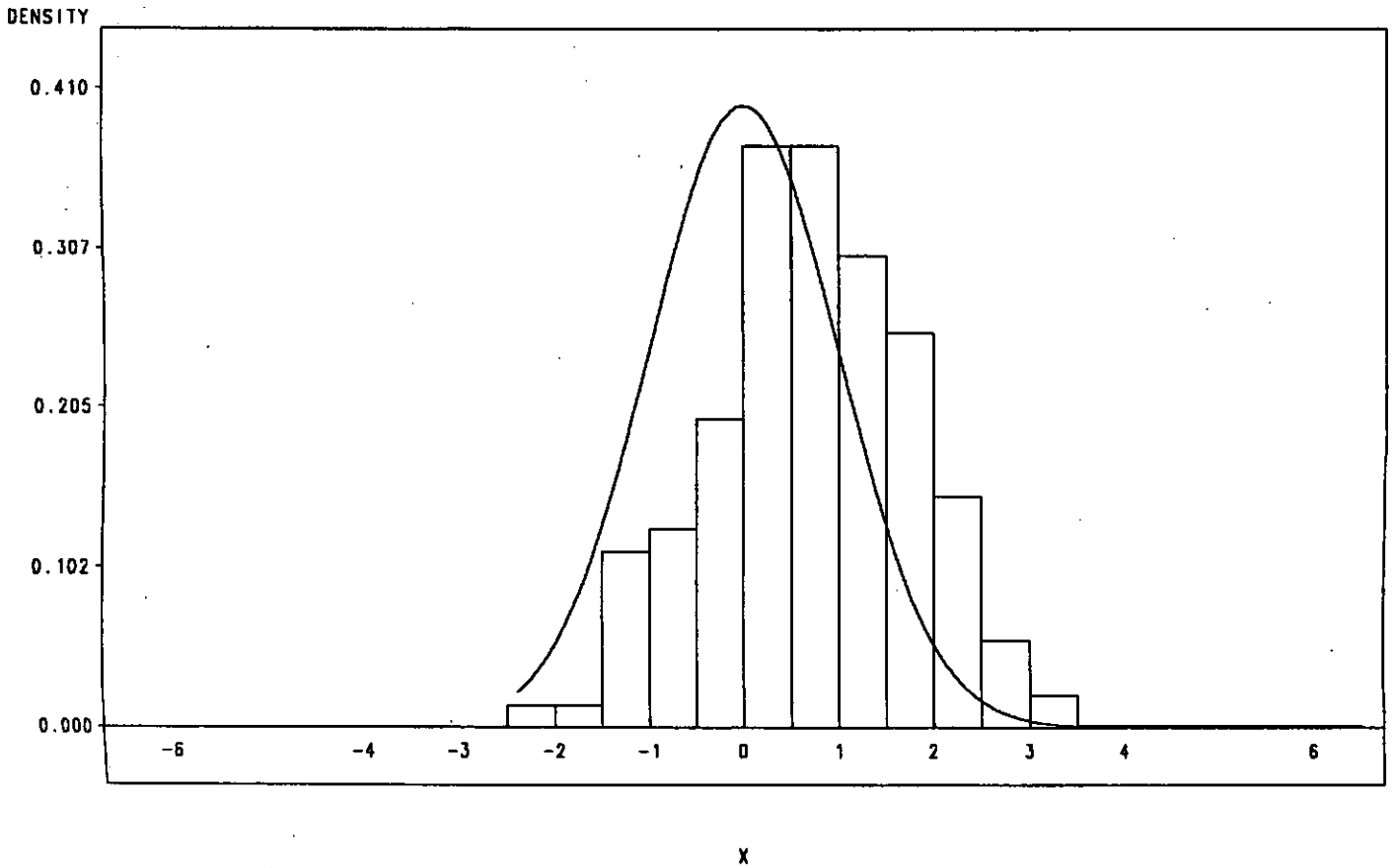
Appendix D

Histograms of the $Z^{(0)}$ Values Overlaid with a Normal PDF (Probability Density Function)

SAMPLING BIAS STUDY
MALE HOUSEHOLD MAINTAINERS
Z(0) VALUES FOR CDS – CANADA



SAMPLING BIAS STUDY
TOTAL POPULATION
Z(0) VALUES FOR CDS – CANADA



Bibliography

Béland, Y. (1990) ***"Results and Methodology of the 1986 Sampling Variance Study"***, Statistics Canada internal report.

Bankier, M., Rathwell, S. and Majkowski, M. (1992) ***"Two Step Generalized Least Squares Estimation in the 1991 Canadian Census"***, Methodology Branch Working Paper, SSMD 92-007E.

Brackstone, G.J. and Rao, J.N.K. (1979) ***"An Investigation of Raking Ratio Estimators"***, Sankhya, Volume 41, Series C, Part 2, pp. 97-114.

Cochran, W.G. (1977) ***Sampling Techniques*** (Third Edition), New York, John Wiley & Sons.

Daoust, P. and Bankier, M. (1989) ***"Evaluation of Weighting for the 1986 Census of Population"***, Statistics Canada internal report.

Dominion Bureau of Statistics (1945), ***Eighth Census of Canada, 1941, Administrative Report of the Dominion Statistician,*** Ottawa: King's Printer.

_____ (1955), ***Ninth Census of Canada, 1951***, Vol. XI, ***Administrative Report***, Ottawa: Queen's Printer.

_____ (1968), ***Sampling in the Census***, S.M.S.03.5.

_____ (1970), ***1961 Census of Canada***, **General Review**, Bulletin 7.2-12, Catalogue No. 99-537, Ottawa: Queen's Printer.

Fellegi, I.P. (1964) ***"Response Variance and its Estimation"***, Journal of the American Statistical Association, 59, pp. 1016-1041.

Hansen, M.H., Hurwitz, W.N., and Bershad, M.A. (1959) ***"Measurement Errors in Censuses and Surveys"***, Bulletin of the International Statistical Institute, 38, pp. 359-374.

Kendall, M.G. and Stuart, A. (1963) ***The Advanced Theory of Statistics***, Volume 1, Charles Griffin and Company Limited, London.

Majkowski, M. (1992a) ***"1991 Census 2A/2B Discrepancies"***, Statistics Canada internal report.

Majkowski, M. (1992b) ***"1991 Weighting Area Formation Evaluation Report"***, Statistics Canada internal report.

Majkowski, M. (1994) ***"Investigation into Large Population/Estimate Differences in the 1991 Census"***, Statistics Canada internal report.

Royce, D. (1983) ***"The Use of Sampling in the 1981 Canadian Census"***, Statistics Canada internal report.

Statistics Canada (1976), ***1971 Census of Canada***, **General Review**, Vol. VI, Part 1, Catalogue No. 99-740, Ottawa.

_____ (1980), ***1976 Census of Canada***, **Administrative Report**, Part 1, Catalogue No. 99-850, Ottawa.

_____ (1980), ***1976 Census of Canada, Quality of Data, Series 1: Sources of Error – Sampling and Weighting***, Catalogue No. 99-844, Ottawa.

_____ (1983), ***1981 Census of Canada, Summary Guide: Total Population***, Catalogue No. 99-902, Ottawa.

_____ (1988), ***1986 Census of Canada, Census Handbook***, Catalogue No. 99-104E, Ottawa.

_____ (1990), ***User's Guide to the Quality of 1986 Census Data: Sampling and Weighting***, Catalogue No. 99-136E, Ottawa.

_____ (1992), ***1991 Census of Canada, Census Handbook***, Catalogue No. 92-305E, Ottawa.

Stuart, A. (1976) ***Basic Ideas of Scientific Sampling*** (Second Edition), New York: Hafner Press.

Regional Reference Centres

Statistics Canada's regional reference centres provide a full range of census products and services. Each reference centre is equipped with a library and a sales counter where users can consult or purchase publications, microcomputer diskettes, microfiche, maps and more.

The staff of the regional reference centres provides consultative and research services in addition to providing after-sales service and support, including seminars and workshops on the use of Statistics Canada information.

Each centre has facilities to retrieve information from Statistics Canada's computerized data retrieval systems CANSIM and E-STAT. A telephone inquiry service is also available with toll-free numbers for regional users outside local calling areas. Call, write, fax or visit the nearest regional reference centre for more information.

Atlantic Region

Serving the provinces of Newfoundland and Labrador, Nova Scotia, Prince Edward Island and New Brunswick.

Advisory Services
Statistics Canada
Viking Building, 3rd Floor
Crosbie Road
St. John's, Newfoundland
A1B 3P2

Toll-free service: 1-800-565-7192
Fax number: (709) 772-6433

Advisory Services
Statistics Canada
North American Life Centre
1770 Market Street
Halifax, Nova Scotia
B3J 3M3

Toll-free service: 1-800-565-7192
Local calls: (902) 426-5331
Fax number: (902) 426-9538

Quebec Region

Advisory Services
Statistics Canada
200 René Lévesque Blvd. W.
Guy Favreau Complex
Suite 412, East Tower
Montréal, Quebec
H2Z 1X4

Toll-free service: 1-800-361-2831
Local calls: (514) 283-5725
Fax number: (514) 283-9350

National Capital Region

Statistical Reference Centre (NCR)
Statistics Canada
R.H. Coats Building Lobby
Holland Avenue
Ottawa, Ontario
K1A 0T6

If outside the local calling area, please dial the toll-free number for your region.

Local calls: (613) 951-8116
Fax number: (613) 951-0581

Ontario Region

Advisory Services
Statistics Canada
Arthur Meighen Building, 10th Floor
25 St. Clair Avenue East
Toronto, Ontario
M4T 1M4

Toll-free service: 1-800-263-1136
Local calls: (416) 973-6586
Fax number: (416) 973-7475

Pacific Region

Serving the province of British Columbia and the Yukon Territory.

Advisory Services
Statistics Canada
Sinclair Centre, Suite 300
757 West Hastings Street
Vancouver, British Columbia
V6C 3C9

Toll-free service: 1-800-663-1551
Local calls: (604) 666-3691
Fax number: (604) 666-4863

Prairie Region

Serving the provinces of Manitoba, Saskatchewan, Alberta and the Northwest Territories.

Advisory Services
Statistics Canada
MacDonald Building, Suite 300
344 Edmonton Street
Winnipeg, Manitoba
R3B 3L9

Toll-free service: 1-800-563-7828
Local calls: (204) 983-4020
Fax number: (204) 983-7543

Advisory Services
Statistics Canada
Avord Tower, 9th Floor
2002 Victoria Avenue
Regina, Saskatchewan
S4P 0R7

Toll-free service: 1-800-563-7828
Local calls: (306) 780-5405
Fax number: (306) 780-5403

Advisory Services
Statistics Canada
First Street Plaza, Room 401
138 – 4th Avenue South-East
Calgary, Alberta
T2G 4Z6

Toll-free service: 1-800-563-7828
Local calls: (403) 292-6717
Fax number: (403) 292-4958

Advisory Services
Statistics Canada
Park Square, 8th Floor
10001 Bellamy Hill
Edmonton, Alberta
T5J 3B6

Toll-free service: 1-800-563-7828
Local calls: (403) 495-3027
Fax number: (403) 495-5318

Telecommunications Device for the Hearing Impaired: 1-800-363-7629

Toll Free Order Only Line (Canada and United States): 1-800-267-6677

Depository Libraries

The Statistics Canada Library in Ottawa maintains complete current and historical records of all Statistics Canada publications, both catalogued and non-catalogued. The library staff is available to help users find the required information.

Statistics Canada Library
R.H. Coats Building, 2nd Floor
Tunney's Pasture
Ottawa, Ontario
K1A 0T6
Local calls: 613-951-8219/20
Fax: 1-613-951-0939

The following is a list of full depository libraries that receive all Statistics Canada publications and all other federal government publications.

Canada

Newfoundland

St. John's

Memorial University of Newfoundland
Queen Elizabeth II Library
St. John's, Newfoundland
A1B 3Y1

Prince Edward Island

Charlottetown

Government Services Library
Charlottetown, Prince Edward Island
C1A 3T2

Nova Scotia

Halifax

Dalhousie University
Killam Memorial Library
Halifax, Nova Scotia
B3H 4H8

Wolfville

Acadia University
Vaughan Memorial Library
Wolfville, Nova Scotia
B0P 1X0

New Brunswick

Fredericton

Legislative Library
Fredericton, New Brunswick
E3B 5H1

University of New Brunswick
Harriet Irving Library
Fredericton, New Brunswick
E3B 5H5

Moncton

Université de Moncton
Bibliothèque Champlain
Moncton, New Brunswick
E1A 3E9

Sackville

Mount Allison University
Ralph Pickard Bell Library
Sackville, New Brunswick
E0A 3C0

Quebec

Montréal

Municipal Library of Montréal
Montréal, Quebec
H2L 1L9

Services documentaires multimédia
Montréal, Quebec
H2C 1T1

Concordia University Library
Montréal, Quebec
H3G 1M8

McGill University
McLennan Library
Montréal, Quebec
H3A 1Y1

Université de Montréal
Bibliothèque des sciences humaines et sociales
Montréal, Quebec
H3C 3T2

Université du Québec à Montréal
Bibliothèque
Montréal, Quebec
H2L 4S6

Québec

National Assembly Library
Québec, Quebec
G1A 1A5

Sherbrooke

Université de Sherbrooke
Bibliothèque générale
Cité universitaire
Sherbrooke, Quebec
J1K 2R1

Sainte-Foy

Université Laval
Bibliothèque générale
Sainte-Foy, Quebec
G1K 7P4

Ontario

Downsview

York University
Scott Library
Downsview, Ontario
M3J 2R6

Guelph

University of Guelph
Library
Guelph, Ontario
N1G 2W1

Hamilton

Hamilton Public Library
Hamilton, Ontario
L8R 3K1

McMaster University
Mills Memorial Library
Hamilton, Ontario
L8S 4L6

Kingston

Queen's University at Kingston
Douglas Library
Kingston, Ontario
K7L 3N6

London

The University of Western Ontario
D.B. Weldon Library
London, Ontario
N6A 3K7

Ottawa

Library of Parliament
Canadian Government Information Section
Ottawa, Ontario
K1A 0A9

National Library of Canada
Ottawa, Ontario
K1A 0N4

University of Ottawa
Morisset Library
Ottawa, Ontario
K1N 9A5

Sudbury

Laurentian University of Sudbury
Library
Sudbury, Ontario
P3C 2C6

Thunder Bay

Lakehead University
Chancellor Paterson Library
Thunder Bay, Ontario
P7B 5E1

Thunder Bay Public Library
Thunder Bay, Ontario
P7E 1C2

Toronto

Legislative Library
Toronto, Ontario
M5S 1A5

Metropolitan Toronto Reference Library
Toronto, Ontario
M4W 2G8

University of Toronto
Robarts Library
Toronto, Ontario
M5S 1A5

Waterloo

University of Waterloo
Dana Porter Arts Library
Waterloo, Ontario
N2L 3G1

Windsor

Windsor Public Library
Windsor, Ontario
N9A 4M9

Manitoba

Winnipeg

Legislative Library
Winnipeg, Manitoba
R3C 0V8

The University of Manitoba
Elizabeth Dafoe Library
Winnipeg, Manitoba
R3T 2N2

Saskatchewan

Regina

Legislative Library
Regina, Saskatchewan
S4S 0B3

Saskatoon

University of Saskatchewan
The Main Library
Saskatoon, Saskatchewan
S7N 0W0

Alberta

Calgary

The University of Calgary
MacKimmie Library
Calgary, Alberta
T2N 1N4

Edmonton

Edmonton Public Library
Edmonton, Alberta
T5J 2V4

Legislative Library
Edmonton, Alberta
T5K 2B6

The University of Alberta
Library
Edmonton, Alberta
T6G 2J8

British Columbia

Burnaby

Simon Fraser University
Library
Burnaby, British Columbia
V5A 1S6

Vancouver

The University of British Columbia
Library
Vancouver, British Columbia
V6T 1Y3

Vancouver Public Library
Vancouver, British Columbia
V6Z 1X5

Victoria

Legislative Library
Victoria, British Columbia
V8V 1X4

University of Victoria
McPherson Library
Victoria, British Columbia
V8W 3H5

Northwest Territories

Yellowknife

Northwest Territories
Government Library
Yellowknife, Northwest Territories
X0E 1H0

Other Countries

Federal Republic of Germany

Preussischer Kulturbesitz
Staatsbibliothek
Abt. Amtsdruckchriften U. Tausch
Postfach 1407
1000 Berlin 30
Germany

United Kingdom

The British Library
London, WC1B 3DG
England, United Kingdom

Japan

National Diet Library
Tokyo, Japan

United States of America

Library of Congress
Washington, D.C. 20540
United States of America

ORDER FORM

Statistics Canada

MAIL TO:

Marketing Division
Publication Sales
Statistics Canada
Ottawa, Ontario
Canada K1A 0T6

(Please print)

PHONE: 1-800-267-6677

Charge to VISA or MasterCard. Outside Canada and the U.S. call (613) 951-7277. Please do not send confirmation.

FAX TO: (613) 951-1584

VISA, MasterCard and Purchase Orders only. Please do not send confirmation. A fax will be treated as an original order.

Company

Department

Attention Title

Address

City Province

Postal Code ( ) Phone ( ) Fax

Please ensure that **all information** is completed.

METHOD OF PAYMENT:

(Check only one)

☐ **Please charge my:** ☐ VISA ☐ MasterCard

Card Number

Signature Expiry Date

☐ **Payment enclosed** $

Please make cheque or money order payable to the *Receiver General for Canada – Publications.*

☐ **Purchase Order Number** *(Please enclose)*

Authorized Signature

Catalogue Number	Title	Date of Issue or Indicate an "S" for subscriptions	Annual Subscription or Book Price			Quantity	Total
			Canada $	United States US$	Other Countries US$		$

Note: Catalogue prices for U.S. and other countries are shown in US dollars.

GST Registration # R121491807

Cheque or money order should be made payable to the *Receiver General for Canada – Publications.*

Canadian clients pay in Canadian funds and add 7% GST. Foreign clients pay total amount in US funds drawn on a US bank. Prices for US and foreign clients are shown in US dollars.

SUBTOTAL	
DISCOUNT (if applicable)	
GST (7%) (Canadian clients only)	
GRAND TOTAL	

PF 093238

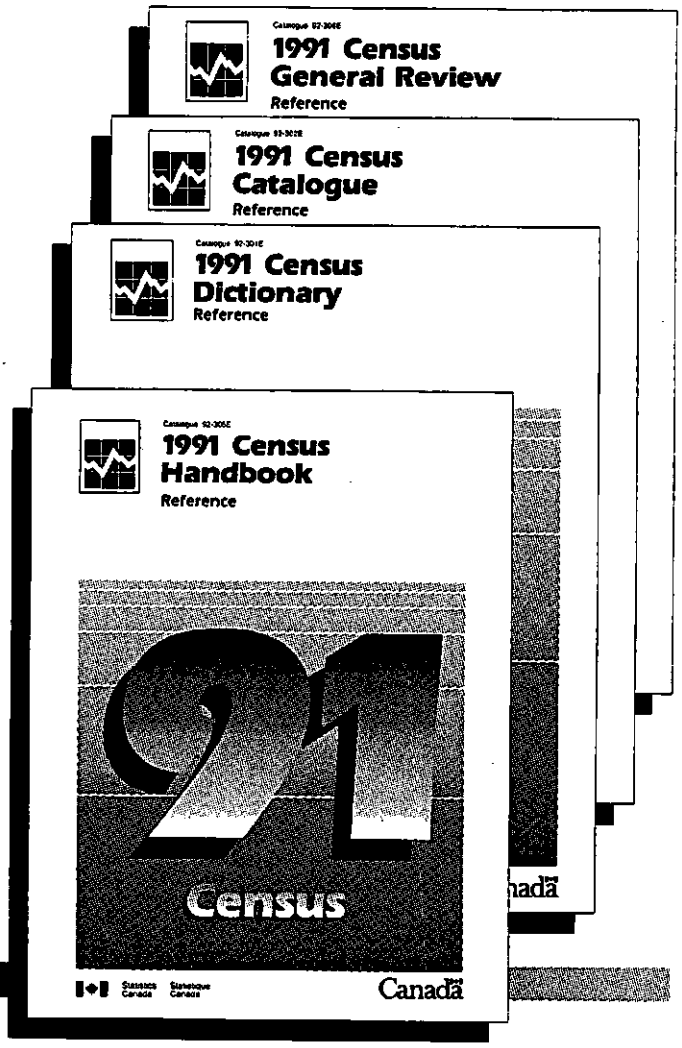
THANK YOU FOR YOUR ORDER!

Statistics Canada Statistique Canada

Canada

The Census is the most comprehensive database available in Canada today...

are you sure you're getting everything out of it?



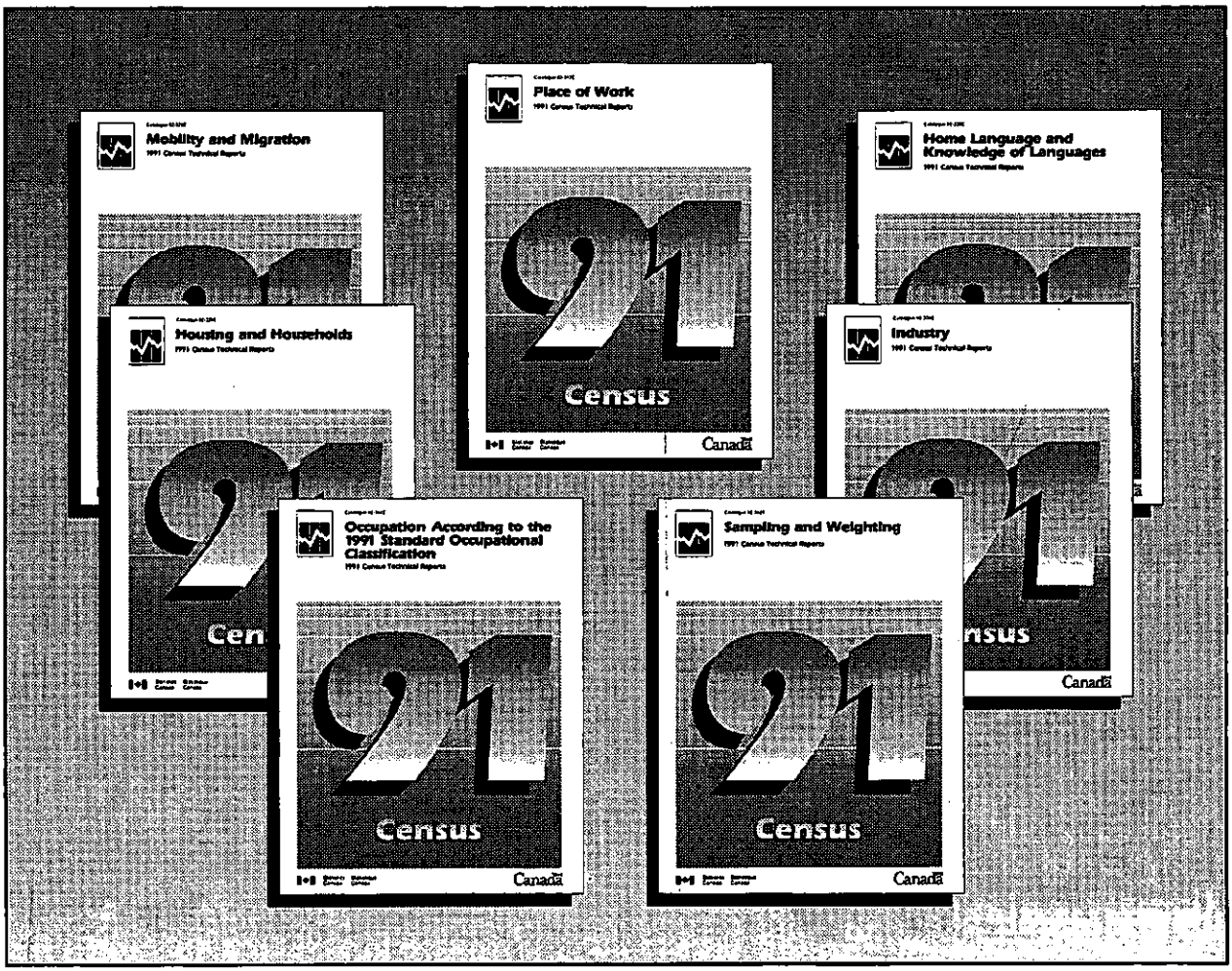
Taking full advantage of Canada's largest, most comprehensive social and economic database is often overwhelming, but the Census can be the most valuable business tool you will ever use. Statistics Canada has designed a series of 1991 General Reference Products to put the Census to work for you.

To order the *Census Dictionary, Census Handbook, Census General Review* or a *Census Catalogue* of products and services, call your nearest Statistics Canada Regional Reference Centre or our national

☎ 1 800 267-6677 ☎

Nuts and Bolts of Census Data

1991 Census Technical Reports provide users with data quality information. Census concepts, variables and their components, definitions, coverage, processing, data evaluation and limitations and much more are explained in detail in each report.



For a complete list of *Technical Reports* available, call your nearest Statistics Canada Regional Reference Centre or our national order line...

1 800 267-6677