

THE 1971 CENSUS OF CANADA

WALTER E. DUFFETT*

INTRODUCTORY REMARKS

SOMETHING like two hundred and fifty years ago Alexander Pope observed, "The proper study of mankind is man. . . . Great lord of all things, yet a prey to all." This expresses the dilemma that still confronts man in attempting to define his role and his needs, in adapting the resources of the world to his own wants, constantly circumscribed and frustrated by his fellow creatures and by the hard facts of life. And so we find ourselves and our society—the actuaries, the economists, the businessmen, the politicians, and, last of all, the statisticians who are supposed to measure it all—endlessly preoccupied with man and his works and wants.

This preoccupation with man and his world is by no means diminishing; it is growing. Two hundred years ago the basic functions of government were conceived to be the maintenance of order and the provision of a climate in which business and other economic activities could prosper with little concern for the social consequences of such activities. These objectives are still apparent, but the emphasis is largely and increasingly on control of economic affairs in the interests of mankind and in direct services to individuals.

This changing emphasis was noted recently in the Eighth Annual Review of the Economic Council of Canada, which drew attention to the preoccupation of government at all levels with what could be called welfare-type expenditures. Combining federal, provincial, and municipal expenditures, the council found that health, social assistance, and education accounted for 35½ per cent of expenditures in fiscal year 1957–58, and only eleven years later had grown to over 48 per cent of expenditures—both a very large amount and a very marked increase. As a proportion of the national product, this group of social expenditures grew from 9½ per cent to 16 per cent in the same period. Estimates for the last two fiscal years indicate that these trends are continuing.

It is therefore not surprising that social statistics—statistics about people—should be a constant preoccupation of society and of government, since programs of this kind and magnitude have to be developed in the light of statistical measures, and such measures are also necessary

* Mr. Walter E. Duffett, not a member of the Society, is Chief Statistician of Canada.

at a later stage to monitor and control the programs. At the apex of this kind of data stands the census, which in the Canadian system includes not only the characteristics of persons but also information on housing and agriculture.

COMPARISON BETWEEN UNITED STATES AND CANADIAN CENSUSES

In terms of information collected on the population and the methods used to collect this information, Canadian and United States censuses are very similar. Since I am sure that most of you are familiar in general terms with the nature of census operations, I do not propose to describe Canadian census features in detail; instead, I shall concentrate on some of the less-known aspects of census-taking, in particular the planning and design of the census, the evaluation of its accuracy, and the dissemination of results.

PLANNING AND DESIGNING THE CENSUS

Public attention to census activities normally is concentrated on the few weeks surrounding census day, and the impression can thus be created that the census is an event which occupies statisticians for only a few months every decade. Such, however, is far from the case, and I would like to say something about the nature of the planning operations necessary for a successful census.

Content of Questionnaires

The Statistics Act specifies that the subject fields covered in the census shall be prescribed by the governor-in-council; in preparation for this, however, census staffs were engaged for almost three years prior to June, 1971, in resolving conflicting and excessive demands from users, to keep the list of questions short enough to ensure that the operation was feasible in terms of public acceptability, quality, timeliness, and cost and yet long enough to meet vital data needs. A series of tests of the proposed questions and new census collection methods culminated with a "dress rehearsal" of the 1971 census in the fall of 1969, almost a year and a half before the census date. The results supported the conclusion that the proposed questions would be acceptable, with a minimum of objection from householders, and that the intended methods for collection would be effective.

In the interests of simplicity, only five basic population questions were asked of everybody, and nine housing questions were asked of all heads of household. All other questions were asked of only a sample of households on the "long" questionnaire, covering one-third of the population.

Methodology

In addition to the establishing of the design and content of questionnaires, a great deal of planning and testing is required on the actual methods to be used in taking a census. The development of the methodology for the 1971 census received its initial impetus from the evaluation of the quality of the 1961 census and from the study of international experience, particularly that of the United States. These studies were directed mainly toward reduction of the important sources of error to which census statistics are subject. For example, respondents may inadvertently or deliberately provide erroneous information, enumerators may adversely influence answers in a number of ways, and additional errors are possible at the data-processing stage. The studies indicated that by far the largest reduction in error could be expected if information was secured as directly as possible from respondents themselves. Consequently, extensive research was conducted into the use of self-enumeration techniques, and, as a result, for the first time in Canadian census history, the self-enumeration system was used on a large scale.

Another technique studied and tested extensively during the planning of the 1971 census was the use of sampling. The main advantages of sampling are that it significantly increases operational efficiency and control and that it reduces costs. In addition, by reducing the editing and processing work load, sampling can make a major contribution to the timely release of census results. Sampling is also an essential element in reducing the burden on the respondent. Without sampling, additional questions could not have been included in the 1971 census, because of the response burden which would have been imposed on the public and the processing burden imposed on the bureau.

Sampling does, of course, contribute to error, particularly for tabulation of "cells" with very small numbers of observations, but the extent of such sampling error can be measured, and information will be included in census publications indicating the reliability of data contained in the various tabulations, drawn from sample results. Looking at the whole census operation, it is believed that any increase in error due to sampling will also be more than offset by the improved accuracy arising from self-enumeration.

RESULTS FROM THE 1971 CENSUS

Unfortunately, it is still somewhat too early to be able to say very much about the results of the 1971 census, in terms of the information collected. Preliminary population estimates for cities, towns, villages,

and municipalities have been published, but the vast majority of the data are still being tabulated.

Some of the preliminary results are, however, of considerable interest and reveal significant trends in population growth. Two main trends are (1) that growth rates have, on the whole, slowed down in the last five years because of lower birth rates and lower immigration and (2) that most of the growth which has taken place has occurred in the fringe or suburban parts of the larger metropolitan centers, usually at the expense of "core" cities and the smaller towns and rural communities.

Although it is still too early to say very much about the results of the 1971 census so far as actual data are concerned, it does appear that the long years of planning and design were well worth the effort in terms of the operational efficiency with which the census was conducted. The self-enumeration technique, when used in conjunction with the questionnaire mail-back feature, produced an initial response rate of approximately 94 per cent, well above expectations and substantially above the 86 per cent experienced in the United States census a year earlier. Costs were kept under control and, in fact, were well under budget estimates. Processing operations to date have been generally satisfactory, and no major difficulties are foreseen.

This is not to say, however, that the entire census operation went off without a hitch. Distribution of a small fraction of questionnaires in the wrong language, difficulties in connection with pay scales and expense allowances for field and processing staff, and even protests by women's liberation proponents over the designation of the husband as the head of the household all caused some problems during the taking of the census. However, such problems were overcome as the census progressed, and I can say that, in general, we are well pleased with the outcome.

EVALUATION OF THE ACCURACY OF CENSUS RESULTS

With the growing use of population census data for important government and business decisions, accuracy is a matter of continuing and growing concern to statisticians. Beyond a certain point the evaluation and preservation of accuracy call for the exercise of very specialized skill and judgment, and the user community owes a considerable debt to the survey specialists and census officers whose advice plays such a large part in the final decisions.

Responsible statisticians have probably been carrying out some kind of "verification" or assessment of the reliability of survey data since the earliest days of survey-taking, and a comprehensive understanding of the different sources of error that surveys are subject to has been gradually

developed during the last thirty years. In fact, the evaluation of sources of errors in surveys has played a major role in the development of present-day survey methodology, which in turn has had important implications for census-taking.

OBJECTIVES OF EVALUATION PROGRAMS

Four major objectives of evaluation programs can be identified: (1) analysis of sources of error with a view to subsequent improvement; (2) the measurement of accuracy of survey results in order to guide users; (3) evaluation of alternative methods of survey design; and (4) evaluation for purposes of continuing control on the performance of a survey.

In the remainder of my remarks on the subject of evaluation I would like to describe, in some detail, the various areas in which such techniques are of particular importance to census operations, namely, in the evaluation of coverage errors, response errors, and processing errors.

Evaluation of Coverage Errors

In the case of a census, a coverage error occurs when a person or household that, according to the definitions, is within the scope of the inquiry is either not included or is included more than once. Thus coverage errors result in biases. For this reason, unlike sampling errors, they cannot be estimated from the census itself; their evaluation involves a comparison with an independent, superior standard. There are three well-known methods of estimating coverage errors in censuses:

1. The first, and most widely used, method of evaluation consists essentially of a postcensus re-enumeration of selected sample areas, with the results of the re-enumeration being reconciled with the census records. Experience has shown, however, that this method of estimating coverage errors may grossly understate errors, since persons who are difficult to enumerate in the census are also difficult to enumerate in the evaluation survey. In view of this weakness, this method will not be used in connection with the 1971 census.

2. A second method of estimating coverage errors is the so-called analytical method, which uses demographic techniques of age-cohort analysis to estimate the current population by age and sex. The method involves adjusting previous census data, using information on such factors as birth rates, death rates, immigration, and emigration. The method, however, is not well suited to a country such as Canada, where unrecorded external and internal migration is a major phenomenon, and hence this technique, too, will not be used in evaluating coverage errors of the 1971 census.

3. A third and relatively new method is called "reverse record check." It was used with considerable success in Canada after both 1961 and 1966 censuses, and it will be the method used to evaluate coverage errors for the 1971 census.

The method consists of a tracing operation, involving a sample of persons selected from a group of lists (such as previous census records, birth and immigration records, and files of drivers' licenses) which, among them, include everyone who lives in the country at the time of the census. After a sample has been selected from each of these lists, the first task in carrying out the reverse record check is to identify the current address of every selected person. This operation is called "tracing" and is the most burdensome operation of the study. The next operation, after the determination of current address, involves searching the current census reports to ascertain whether the person was, in fact, enumerated at the address where he is found to be living or, alternatively, at another address. Once a person is found to have been missed by the census, some additional processing permits the determination of whether he was missed as part of a whole household missed or as part of a partially enumerated household. Some data may also be collected from these persons in order to analyze the characteristics of missed persons; hence the impact of coverage errors on the substantive census estimates.

The major advantage of the reverse record check over the re-enumeration method of estimating coverage errors is that it provides a completely independent approach. Thus it avoids the major problem of evaluating coverage errors, that is, that the evaluation method might itself miss the same persons who were missed in the census.

The reverse record check was carried out for the first time in Canada in connection with the evaluation of the 1961 census. This procedure was in the nature of an exploration of the method and, as such, was carried out using a relatively small sample size of some 6,000 persons. A much larger study of the same kind was carried out following the 1966 census (sample size almost 27,000 persons), permitting a more refined breakdown of the underenumeration by age and sex.

We are repeating the reverse record check study for the 1971 census and are adding several new features to it, most important of which is the fact that, having identified the sample of persons who were missed in the census, we will contact them and ask a series of census-type questions of them. A tabulation of the results will indicate not only how many people were missed in 1971 but also what the impact is of having missed them on the different census tabulations. This will be an instance of guiding users with respect to the reliability of census tabulations. The same data will

also enable us to identify the characteristics of persons most likely to be missed by the census—thereby enabling us to attempt to search for new methods of census-taking which might come closer to including such persons in future censuses.

Evaluation and Measurement of Response Errors

Response errors may occur whenever data are requested, provided, received, or recorded. Questions may be misinterpreted by the respondent; he or she may not know, may not remember, or may purposely want to distort the correct answer. Different enumerators may have different tendencies in putting and explaining questions or interpreting responses. Response errors can be classified in two categories.

1. The first type of error, response variance, is that component of the response errors which has a chance to cancel over a large number of responses. Because this type of error may be in either direction, the net error from this source may be quite small for large areas, but it can be very large for small areas or rare characteristics. It also tends to be much larger for sensitive characteristics which are normally difficult to measure.

As part of the evaluation program of the 1961 census, a project was carried out to measure the response variance, particularly the component of the response variance that can be attributed to interviewers. Although this study was restricted to a purposively selected area of the country (an area considered to be generally similar in characteristics to the remainder of the country), the project had an enormous impact on the procedures eventually adopted for the 1971 census.

The results of the study indicated that, on the average, the response variance due to enumerators was of the same magnitude as the sampling variance would have been if the census had been carried out using a 25 per cent sample instead of 100 per cent enumeration. It became quite clear that for many characteristics the contribution of enumerators to the response variance was overwhelmingly the largest source of error of small-area estimates. As a consequence, it was clear that our census methodology was not in optimal balance with respect to cost and error; if the response variance due to enumerators could be significantly reduced, one could introduce sampling into the census and still achieve a net reduction of errors, at the same time achieving some reductions in costs and a definite improvement in timeliness. As a result of these considerations, plus a good deal of testing, a new methodology for the 1971 census emerged, involving, as outlined earlier in this paper, the use of sampling and self-enumeration techniques. The point to be emphasized here, however, is

that the methodology of the 1971 census was influenced in a very significant way by the response variance study.

2. The second type of error is response bias, which is, roughly speaking, that portion of the total response error which is "left over" after all cancellation involved in the response variance has occurred. It is made up of response errors which have a tendency to occur more in one direction than in the other. It may be the result of the training and attitude of interviewers, or the reluctance of respondents to admit certain characteristics or a tendency to consistently overrate or consistently underrate them; or it may be the fault of the questionnaire, in that it may invite errors in one direction more than in the other. Such errors do not cancel out even over large areas or a large number of respondents and may be particularly damaging for statistics at the metropolitan-area, province, or national level. As in the case of response variance, response bias has a tendency to be much larger for characteristics that are difficult to measure.

Response bias may sometimes be detected (though not measured) on the basis of analyzing departures in the data from hypothetical models (internal consistency analysis). Response bias, however, is notoriously difficult to measure or even approximate.

Evaluation of Processing Errors

Errors are of course introduced into survey data not only during the reporting of data but during processing as well. In order to minimize such errors, quality-control techniques were applied to several processing operations in connection with the 1971 census. In addition to the immediate feedback into the quality of the present census, these control operations provide a wealth of evaluation-type data concerning the respective operations, which will be analyzed from the point of view of future censuses.

THE 1971 CENSUS DISSEMINATION PROGRAM

With census day now behind us, Statistics Canada is concentrating its energy on the processing and dissemination of census results. The objective sought for is that the almost unlimited potential of this unique source of information be available with the highest possible flexibility to all those who may or could apply census data to their administrative, planning, or research needs. To meet its objective, the census dissemination program must not only ensure that all possible users be made aware of what is immediately available in a preplanned tabulation program, but it must also permit users to appraise all the potentialities of the census data base in relation to their specific needs. The program put forward by

the Census Division to meet these objectives can be divided into two parts, that is, a tabulation program and a data access program.

The Tabulation Program

The object of the tabulation program is to approach the high volume of demand for census data in an orderly and systematic fashion. This calls for both a regular and a special tabulation program.

The regular tabulation program was planned in order to meet the most common needs for census data, as known from steady contacts with census data users in all fields of activity. The amount of census information is so great, however, that it cannot be covered by a single, preplanned tabulation program. Nevertheless, meeting special data requests for such users is looked upon as an integral part of the 1971 census dissemination program, and a User Inquiry Service, supported by powerful computer programs, will handle special requests of every nature as quickly and as effectively as possible.

A *Data Access Newsletter*, available free on request from the Census Division, will provide users with information on the data content, geographical possibilities, and other aspects of the regular and special tabulation programs.

The Access Program

The census data access program is aimed at getting available data into the hands of users. The program can be described as an understanding of and an approach to data access. The program is immediately concerned with defining the principles and setting up the machinery by means of which any organization or individual who needs data may directly or indirectly search this data source and find out to what extent and in what way it can meet his needs. It is based on the principle that any statistical information, however simple or complex, is necessarily and essentially composed of data elements or characteristics which can be identified and specified in a logical pattern, so that a tabulation program or a data dissemination program can be used to call and retrieve information from the data base according to what combination of factors is specified. A series of descriptive documents and standardized request forms will be used to assist data users to specify exactly the data they wish to obtain.

Geocoding

A development of major importance for the dissemination of census information is the geocoding program which has recently been completed at Statistics Canada, which will enable us to provide quickly and cheaply,

by computer methods, census information for areas of almost any size and shape desired by a user. Officially, this program is named the Geographically Referenced Data Storage and Retrieval Program, or GRDSR for short. Although the program is not limited to census use, the first major application of the system is to geocode a copy of the 1971 census data, and, once this file has been geocoded, it will be possible to answer special requests for cross-tabulations for user-specified areas, such as planning districts or radio-television contours.

For fourteen of the largest urban centers in Canada (which account for about one-third of the Canadian population), data will be coded at the street block-face level of detail, and it will thus be possible for the first time to obtain census information for very small user-specified areas within these fourteen urban centers. For other areas of the country, data will be geocoded at the level of the census enumeration areas within electoral districts.

Statistics Canada Life Tables

One aspect of the bureau's program which may be of particular interest to actuaries is the preparation of life tables. Census population data, of course, are an essential ingredient of life tables, and Statistics Canada since 1931 has been producing a series of life tables centered around June 1 decennial or quinquennial census data and three-year death statistics around the census year. This year, for the first time, life tables (1965-67) for Canada and for each of the ten provinces were published, thus providing considerably more geographic detail than had hitherto been available. When results of the current census become available, with the necessary mortality rates for 1970-72, similar Canada and provincial tables will again be prepared. The demand for the life tables produced by the bureau is substantial, and the tables are widely used in legal proceedings to assist in settling claims for damages.

In addition to ordinary life tables, the bureau also has produced a number of specialized life tables, such as abridged tables, tables for specific population groups (such as Indians and Eskimos), for cities of more than 100,000 population, and for particular ethnic groups. Tables based on the exclusion of certain causes of death (such as accidents or cancer) have also been prepared in order to illustrate what would happen to life expectancy if these conditions could be controlled. Joint life tables for various age combinations of husbands and wives have also been produced. Much of this work has been done in response to outside requests, and we expect to continue to provide this service in the future.

The 1971 census will permit an updating of the pioneering series of

working-life tables for Canadian males that were prepared by Frank Denton and Sylvia Ostry for the 1961 Census Monograph Programme. The working-life table is an extension of the ordinary life table, incorporating labor force participation rates. Two important measures obtained from the working-life table are the average number of years of working life remaining for a person aged X and the average number of years of retirement that can be expected by persons aged X . We expect that the working-life tables will be further developed in future, and we would certainly be interested in receiving comments or suggestions on their development from actuaries or other experts in the business of insurance.

CONCLUSION

In bringing my remarks to a close, I should like to observe that in my contacts with the insurance industry and with other related fields such as pensions I am aware of the breadth of interest of the actuarial community, a fact which is borne out by your program on this occasion. It may well be that Statistics Canada has resources, beyond those which we have been considering today, which are related to these broader interests. Our doors, and particularly mine, are open to all of you, and I would be happy to see any of you at any time.

