



52879



UNITED NATIONS
ECONOMIC AND SOCIAL COUNCIL

Distr.
LIMITED

E/CN.14/POP/91
26 June 1973

Original: ENGLISH

ECONOMIC COMMISSION FOR AFRICA

Seminar on Techniques of Evaluation
of Basic Demographic Data

Accra, Ghana, 16-28 July 1973

THE USE OF SAMPLING TO EXTEND AND SUPPLEMENT INFORMATION
OBTAINED FROM POPULATION CENSUSES⁺

+ This document was issued for the United Nations Seminar on Evaluation and Utilization of Population Census Data in Latin America, held in Santiago, Chile, from 30 November to 18 December 1959 and the United Nations Seminar on Evaluation and Utilization of Population Census data in Asia and the Far East, Bombay, India, from 20 June to 8 July 1960. It is intended to use this document also as a basis for discussion at the ECA Seminar on Techniques of Evaluation of Basic Demographic Data, Accra, Ghana, from 16 to 28 July 1973.

I. Introduction

The demand for extensive and reliable demographic data is growing in most countries of the world, and especially those engaged in the task of planning for economic and social development feel this need keenly. The population censuses have been one of the chief sources of information on a national basis, but census programmes even in the more advanced countries are now unable to meet all the requirements. Censuses are expensive and if all the needs were to be met through the traditional census method of complete enumeration and wholesale tabulation of the data collected, the cost, time and organization required would place a serious strain on the resources normally available for census-taking. The use of sampling methods in population censuses has of late afforded the opportunity of getting much additional information which would otherwise not be available. In addition, the application of modern sampling theory has also opened up new ways in which information on economic and social characteristics of the population can be obtained by means other than population censuses. Such sample surveys, however, depend largely on the information provided by censuses for their proper execution. In the present day, therefore, the census and sampling procedures are not competing techniques and should be looked upon as interrelated parts of a concerted programme of data collection and research. It is the aim of this paper to emphasize this relationship and point out the need for an integrated approach if the increasing demand for demographic information is to be met. ^{1/}

The use of sampling in connexion with population censuses to obtain additional information will first be dealt with and then will follow a discussion of sample surveys which may be carried out separately from the census. The use of sample surveys when a census is not feasible will be briefly touched upon at the end of the paper.

This paper deals primarily with Asia and the Far East and with the findings and recommendations of a regional Working Group of Experts on Sampling Methods held in Bangkok in 1959. ^{2/} However, the principles set forth here are almost equally applicable to the Latin American region.

II. Use of Sampling in Population Censuses

A. Sampling methods in census enumeration

Sampling has been used in a number of countries for broadening the scope of the census by asking supplementary questions of only a sample of the population. Modern experience in the use of sampling techniques has confirmed that it is not necessary to gather all demographic

- ^{1/} For further information reference may be made to the United Nations documents: Handbook of Population Census Methods, vol. I. General aspects of a population census (Statistical Papers, Series F, No.15, Rev.I) and Principles and Recommendations for National Population Censuses (Statistical Papers, Series M, No.27).
- ^{2/} United Nations Economic Commission for Asia and the Far East and United Nations Statistical Commission, Conference of Asian Statisticians: Report of the Working Group of Experts on Sampling Methods. Bangkok, 14-25 September 1959.

information on a complete basis; the use of sampling actually saves a good deal of time and money and, furthermore, under certain circumstances the sampling approach alone ensures data of acceptable accuracy.

Supplementary information may be obtained by one of two different methods. Either the same canvassers may ask the supplementary questions or specially trained canvassers may be engaged in enumerating the sample. The first method may be used where it is fairly simple to obtain the supplementary information by conventional enumeration techniques and where keeping the cost of the census low is the principal objective; the second method is used where ordinary enumeration methods are not expected to elicit accurate answers to the supplementary questions and where it is more important to secure reliable information than to save costs. Where special canvassers are used to obtain supplementary data, such information can be obtained at the time of the main enumeration or within a short period before or afterward.

In Ceylon, sampling was introduced in the 1953 census of population for the purpose of getting information on fertility, education, income and unemployment. It was felt that obtaining reliable data on these questions would be an onerous task for the ordinary canvassers. Consequently it was decided to assign to the sample enumeration a few well-selected and well-trained canvassers whose work could be generally confined to these items of inquiry and who therefore could devote particular attention to the accuracy of the answers.

In the 1953-1954 censuses of Burma only the basic items were taken up in the complete enumeration and such characteristics as religion, citizenship, internal migration, education, labour force, income and fertility were covered on a sampling basis.

The Working Group of Experts on Sampling Methods which met recently in Bangkok discussed the problems involved in supplemental enumeration by sampling. The Group noted that only a few countries in the ECAFE region proposed to incorporate this procedure in the censuses to be taken about 1960. The Group was of the opinion that for the sampling enumeration, the items of inquiry should be as few as possible, and should be confined to those which are time-consuming or require better-trained investigators and for which tabulation might not be required by detailed geographical or other break-downs. The Group did not consider it desirable to make the sample enumeration simultaneously with the main enumeration, as to do so might adversely affect the main enumeration. The Group was in favour of making the sample enumeration after the main enumeration.

Various sampling units can be used, such as individuals, households or enumeration districts. The main alternatives which may arise in practice are the choice of a random sample of persons or households from each enumeration district as against the inclusion of every person or household from a sample of enumeration districts. The first procedure

would generally be considered better on theoretical grounds, but there may be practical reasons to choose the latter. This would necessitate the training of fewer enumerators in the collection of the data on the supplementary questions. Data processing would also be simpler, as enumeration districts are often the smallest units used in the preparation of tabulations. Further, if enumeration districts are used as sampling units, the choice of the units can be done in a central office well in advance of the actual enumeration.

B. Sampling methods in census tabulation

The use of sampling at the tabulation stage received warmer support from the Working Group that its use for collection of data, as being more easily accomplished in many countries in the ECAFE region. Sample tabulations may be undertaken from two points of view: (i) to provide provisional results and (ii) to provide additional data, as explained below.

(i) Tabulation of provisional results. A complete national census is a huge undertaking and unless sampling methods are introduced, several years may elapse after the census enumeration before some of the tabulations of the greatest economic and social interest are published. The probability of such a lag is especially great in the economically less developed countries, where resources are limited and few or no funds can be devoted to modern tabulation equipment. The danger is also present in countries where such equipment is available in large quantities, since in these countries the demands made for census tabulations are also correspondingly heavy. In most countries, canvassers prepare summaries referring to the number of persons enumerated and some other basic facts at the time of enumeration. Yet there is a definite need for early tabulations of some of the results, such as the distribution of the population by age groups, employment status, branches of industry, etc., which cannot be derived from the canvassers' preliminary returns. In addition to the speedy release of preliminary data, advance sample tabulations may also provide a convenient tool for internal checks and analysis. For example, one can use the sample to determine rates of non-response in the census for various characteristics and to decide on final tabulation specifications. In actual practice, the possible advantages of advance sample tabulations should be weighed against the consequent delay of the complete tabulations.

(ii) Tabulation of additional data. The complete tabulation of all the information collected in a population census requires considerable time and money. Consequently the tabulation programmes of most population censuses provide for the complete tabulation of only the legally prescribed items and of those elements of demographic and economic information which are regarded as of sufficient importance for small administrative areas. There are certain demographic data which are needed only for large areas and for the country as a whole. These tend to fall into two classes: (a) detailed tabulations of certain characteristics such as age by single years, or occupation or industry classifications by detailed codes, and (b) cross-tabulations of multiple factors such as

marital status by age and religion. Such tabulations are mostly intended for the use of specialists. Sampling enables one to obtain detailed tabulations for large areas with reasonably small sampling errors at a much reduced cost and in a shorter time than needed for tabulations on a complete basis. Even for cross-tabulations prepared on a sample basis, the cost is relatively not large.

The sampling considerations involved in the present case are somewhat different from those in the preparation of advance tabulations referred to above. Whereas advance sample tabulations are prepared for meeting urgent needs after a census and are published eventually on a complete basis, here one is concerned with sample tabulations and cross-tabulations which are not intended to be prepared on a complete basis at any later date. This important difference highlights the need to exercise the utmost care in the designing and executing of the sample with a view to attaining the highest accuracy within permissible costs. The use of sub-samples in reducing the processing errors was suggested by the Working Group of Experts on Sampling.

The Indian census of 1951 made extensive use of sample tabulations. The tables on age, literacy, civil condition and maternity were prepared only on the basis of a 10 per cent sample. The tables on household size and composition were obtained from a 4 per cent sample. In Japan, both in the 1950 and the 1955 censuses, sample tabulations were made for a number of characteristics.

III. Sample Inquiries Subsequent to the Census

In the foregoing section attention was given to the obtaining of additional information by using sampling methods as an integral part of the census programme either to broaden the scope of data collected at the census or to enlarge the range of tabulations made with the census data. There are a variety of circumstances where data in addition to those obtained in a census are required. Such requirements are frequent both in countries where the census enumerations cover a wide range of subjects and in others where the topics covered by the census have to be restricted because of cost, lack of trained canvassers, etc. Sample inquiries are extremely useful in meeting these requirements.

Some of the demographic requirements which call for the use of sample inquiries are detailed below:

- (1) The analysis and study of the census data may themselves suggest the formulation of hypotheses which would require testing by means of additional information obtained through sample surveys. Such, for instance, may be the investigation of factors responsible for the different rates of population growth in the different areas of the country. Sample inquiries relating to mortality, fertility or migration may explain such differences to a much greater extent than

would be possible only with the census data. The need for such inquiries will also be greater if birth and death registration data are incomplete or inaccurate, than if complete vital statistics are available.

(2) The working out of plans for economic and social development calls for detailed information on a number of subjects which cannot normally be covered by population censuses. Data on family budgets, levels of wages and earnings, under-employment, health and housing are constantly required and have to be provided by methods other than the regular censuses. Although such information has to be obtained through sample surveys, as will be shown below, the data obtained in the censuses have to be used in the planning of these surveys.

(3) Sample inquiries are often required to bring up to date the information obtained in a census. Censuses are usually conducted at intervals of ten years, and in the intervening period between two censuses some of the data might become grossly out of date. This is likely to be the case with respect to data on the labour force, information on which is obtained in the population censuses of many countries. Similarly, in areas subjected to intense public health programmes, the rates of population growth as estimated from previous census data may no longer be applicable and current estimates of population size and characteristics may be required. In some countries, as in Burma, the need for such data has resulted in the carrying out of a sample census, on the lines of a population census, but covering only a sample of the population.

A. The use of census data for the designing of subsequent sample surveys

(i) Use of sampling frames

The selection of sample units in conducting a sample survey has to be made from some kind of complete list of these units, namely the sampling frame. The sampling frame may be a list of small areas, structures, establishments, persons, households or groups of households. Normally such a frame is obtainable only from the main work of a census or from its by-products.

Unlike sampling in conjunction with the census programme, discussed earlier, the problem of ensuring satisfactory sampling frames becomes more difficult when sample surveys have to be carried out subsequent to the census. In the former case, whether sampling is undertaken for supplemental enumeration or for purposes of tabulation, the census itself provides directly a satisfactory sampling frame. When sample surveys are planned subsequent to the census, the frame provided by the census becomes increasingly obsolete as time elapses after the census. The comparative usefulness of different lists, viz. of small areas or households, available from the census, would to a large extent be governed by the relative stability of the units included in the list.

The limitations of the sampling frame provided by the census may not be due entirely to the changes that have occurred since the time of the census. Other considerations to be taken into account are such defects ^{3/} as (a) inaccuracy (arising from wrong information about the units listed or defined by it and also from listing of units which are in actual fact non-existent), (b) incompleteness (some of the units are omitted from the list), (c) duplications, and (d) inadequacy (exclusion of certain parts of the universe from the frame). The desirability of a careful examination of those population census data which are to be used as the frame for a proposed sample survey is apparent. Of even greater importance is the conclusion that census planners should give careful consideration to the fact that census records may be used as a frame for future sample surveys. This would call for the keeping of the census records in a form which can be readily used for sample survey work, with a clear indication of their deficiencies when used as a sampling frame. It is important to anticipate so far as possible the types of sampling inquiries that are likely to be undertaken and kinds of frames that will be required.

(ii) Efficiency in sampling design

Two principles are in general use for increasing the efficiency of the sampling design: "multi-stage" sampling and "stratification". The use of either of these principles in designing sample surveys depends to a large extent on the basic information provided by the censuses. If a multi-stage sample design is decided on, the specification of the sampling unit to be used at any stage depends not only on what is desirable but also on the maps and other information available from previous censuses. Similarly, the adoption of the principle of stratification presupposes the availability of information on economic or social characteristics such as type of housing, religion or literacy for smaller units which can be combined suitably to form a few large strata.

B. Recent experience in some countries.

A few examples of recent experience in some countries will help to bring out the variety of topics on which sample surveys are undertaken at the present time and the extent to which they depend on census data for their proper execution.

(i) In the Philippines Statistical Survey of Households, which is conducted periodically, about 6,500 households are enumerated by personal interviews. The types of data gathered have varied from one round of the

^{3/} These defects are discussed in the United Nations Handbook of Population Census Methods, vol. I. General Aspects of a Population Census. Statistical papers, Series F., No. 5, Rev. 1. The discussion is based primarily on Yates, F. Sampling methods for censuses and surveys. New York, 1953.

survey to another. Data on employment and unemployment and certain other characteristics of the labour force have been collected at each round. A multi-stage sampling design is used, the household being the unit of enumeration.

Bantegui and Makanas reported: "A major difficulty which was encountered in planning and carrying out the operations of the Philippines Statistical Survey of Households was the lack of maps. Only a few of the maps of municipalities and chartered cities and capitals of provinces which were used in the 1948 Censuses of Population and Agriculture were available at the start of survey operations in May 1956. Many of these maps, however, were out of date. Legislations and executive orders amending the boundaries of a number of provinces, municipalities and cities had been effected since these maps were made. In listing households within the selected poblaciones and barrios of sample municipalities as well as households in selected precincts in chartered cities and capitals of provinces, enumerators had to depend greatly on the information provided by the people residing in the aforementioned sample areas concerning the correct boundaries of these areas." 4/ The lack of maps resulted in considerable underlisting and overlisting of households, as shown by the comparison of the numbers of households listed in the same sample area in successive rounds.

(ii) In several recent sample studies in India, the principle of stratification has been applied, with the use of data provided by the previous censuses. The Mysore Population Study was one such undertaken jointly by the United Nations and the Government of India shortly after the 1951 Population Census as a pilot project to study the relationships between population and economic and social factors. 5/ The study covered a large part of the "urban" and "rural" areas of Mysore State, on a sampling basis. In Bangalore City, where the study of differentials of fertility and of factors affecting them was particularly important, the city population was stratified into five groups based on the religious composition and literacy of the population. Such stratification made it possible to include in the survey a larger number of persons of minority groups, such as Muslims and Christians than would have been possible if the sample households had been drawn indiscriminately from all parts of the city. The city had fifty municipal divisions, for each of which literacy data from the 1951 census were available when the survey was being planned. Since the 1951 data on religion had not yet been compiled at that time, the 1941 census figures on this subject were used instead.

4/ Bantegui, B.G. and Makanas, E. The Philippines Statistical Survey of Households - The problems met and their solution. Working paper presented to the meeting of the Working Group of Experts on Sampling Methods, Bangkok, 1959, pp. 2-3.

5/ Chandrasekaran, C. "Use of household sample in the United Nations - Government of India Population Study in the Mysore State". United Nations, Proceedings of the World Population Conference, Rome, 1954. Papers vol. VI (E/CONF.13/418).

Municipal division with 35 per cent or more Muslims formed one stratum, those with 35 per cent or more Christians formed another, and those with 35 per cent or more Scheduled-caste Hindus formed a third. The remaining municipal divisions were divided into two strata, according to whether the percentage of literates among males aged five years and over was sixty and above; or below sixty. In each of the five strata so formed a two-stage sampling procedure was adopted; the census block, as defined for the 1951 census, serving as the primary sampling unit and the household within the census block as the secondary sampling unit.

The National Sample Survey of India has since 1950 been obtaining data on a variety of subjects in the different rounds of the survey. These have included general demographic data on age, sex, marital status, economic and employment status, births, deaths, etc., and detailed economic data such as consumer expenditure, household enterprises, utilization of land, crop survey, survey of manufacturing establishments, etc. In dividing the country into strata for purposes of sampling, "natural divisions" have been used; these were formed by the Census Department at the time of the 1951 census to ensure homogeneity in geological, climatic and cropping patterns. This revision has been "an improvement over the earlier procedure in which the country was divided into compact areas by more or less arbitrary combinations of districts". ^{6/}

The officials of the 1951 census also arranged for the preparation of District Census Handbooks, each book relating to one district and containing the basic statistics relating to different local areas within the district as well as the basic totals of population (by eight separate livelihood classes) for every village, and every ward of every town in the district. The handbooks also include other items of information relating to the districts which could be readily assembled and used for purposes of reference, such as area and number of houses. This publication has been found of great value for use in multi-stage sampling, especially in using such devices as 'probability proportionate to size' for improvement of the efficiency of sampling.

Another effort made at the 1951 Indian census was to introduce the National Register of Citizens, in which important particulars of every person enumerated at the census were entered, arranged by households for each enumeration district. This register was expected to be kept up to date by making additions or deletions as persons or households moved into or out of the district. Experience has shown that the maintenance of these registers is not satisfactory, and that they cannot be completely relied upon to give a sampling frame of households or individuals in selected areas. Still, those registers, as well as the lists of households which are made in preparation for a census, are helpful in making fresh lists of households for sample surveys.

^{6/} Lahiri, D.B. Technical paper on some aspects of the development of the sample design. The National Sample Survey, No. 5, Government of India, March 1954.

In India also, the absence of detailed and up-to-date maps causes difficulty in the delineation of sampling units. In the rural areas, villages serve as administrative units and their boundaries are fairly well recognized by the village headman, who can be consulted on the spot. The use of a village as a sampling unit does not, therefore, cause much difficulty. In sampling towns or large cities the absence of well defined maps is a major handicap, although the use of "census blocks" as sampling units has been attempted in many urban surveys.

(iii) The difficulties experienced in Malaya in the selection of households for a Budget Survey are revealed by the following statement: "There were no proper frames for the selection of households in rural areas after the districts and mukims had been chosen, so investigators had to be sent to those mukims to get more detail about the locality and composition of the various villages and kampongs there. From the information thus obtained, villages and kampongs fairly close to one another were grouped and treated as one unit from which to select the ten households for one month of the survey." 7/ For an unemployment survey, "The sample in Malaya was drawn from the 1957 Census of Population household schedules. Many of the Census cards had been torn or removed and not all schedules bore the proper postal address. It must be remembered that this survey was carried out nearly one and a half years after the Census; some squatter units had disappeared since the Census, and houses in certain parts of Kuala Lumpur had been renumbered. In many cases, there were no proper Census block maps, so that the investigators, who were often unfamiliar with the areas allocated to them, had difficulty in identifying the households chosen in the sample". 8/

(iv) In Indonesia sample surveys have been carried out or proposed on family living, wages and earnings, employment, unemployment and labour force; and data to help in the application of multi-stage sampling appear to be readily available. "The successive territorial divisions in Indonesia are (i) Province, (ii) Kabupaten, (iii) Kawedanaan, (iv) Katjamatan and (v) Desa. Each of the major islands of Indonesia is composed of these divisions and they are all well defined in maps. Sampling work in Indonesia is very much facilitated by availability of lists of desas (villages), establishments, and households. For all the geographical divisions indicated above, information is available regarding area and population. For the smallest geographical division, namely the desa, information is, in addition, available regarding households and some of their significant characteristics, such as name of head of household, his address, family composition, extent of land cultivated, crops raised, etc." 9/ These data in Indonesia are obtained from sources other than the censuses.

7/ Leong, G.H., The Household Budget Survey and the Unemployment Survey in Malaya. Working Paper presented to the meeting of the Working Group of Experts on Sampling Methods, Bangkok, 1959, p. 2.

8/ Ibid., pp. 5-6.

9/ Adyanthaya, N.K. Sample surveys in Indonesia. Working Paper presented to the Working Group of Experts on Sampling Methods, Bangkok, 1959.

(v) In Japan, at every census, enumeration districts are formed. The entire area of Japan is divided into approximately 400,000 enumeration districts, each district being so established so as to include approximately fifty households, and to provide natural boundaries without crossing the existing political or community sub-divisions. The enumeration districts are revised at each quinquennial census. Information regarding each enumeration district is transferred to punch cards, and these punch cards provide a ready-made sampling frame for many inquiries. Similar information has been used in the Ryuku Islands to make a master sample. ^{10/}

C. Conclusions of the Working Group of Experts

The Working Group of Experts on Sampling Methods discussed the requirements of sample surveys which could be met by census operations and came to the following conclusions:

It is essential to divide the entire country into territorial divisions, sub-divisions and small area units. The main criterion in making these units is that they should have precise boundaries, identifiable on the spot and operationally convenient. It would be of further advantage if maps of these areas, even sketch maps prepared by non-specialists, were made available. Areas defined by such maps are more useful than those demarcated by lists of houses. In the formation of geographic units, it is advisable to consider both the rural and urban areas at the same time in order to deal effectively with the border problems.

The kind of classification of areas adopted must depend on the situation in the different countries, particularly the level of statistical development and demand for statistics. Villages may well form the units in rural areas. Large villages and towns will have to be further subdivided into enumeration districts which may serve as the area units. In making such sub-divisions the clarity of boundaries and operational convenience should be given greater weight than the desire to approximate uniformity of size in terms of population or area.

It is important not only to demarcate small areal units but also to make available the basic information in regard to population size, agricultural area, etc., for these units. The available information for the small areal units should be kept in one or more central places if it is to be conveniently available for use for sampling purposes.

It is doubtful if data for sampling units smaller than the enumeration districts can be provided. In particular, the preparation of household lists was not considered of great value by the Working Group of Experts, as it is difficult to keep these up to date.

^{10/} United Nations, Handbook of Population Census Methods. Vol. 1 (op. cit.), pp. 157-159.

The Working Group recommended that the programme of work in regard to the preparation of sampling frames should be built up as an integral component of the census programme. A beginning, even if it be modest, should be made in the 1960 census programme. The extra cost involved in the preparation of frames and maps and in assembling information for small units should be balanced against the continuous gains that would result when these are used for the planning and execution of numerous sample inquiries.

IV. Sample Surveys When a Census is Not Feasible

The sample survey serves as a useful method of obtaining information in countries where censuses have never been taken. Even in such situations, the sampling method should not be considered as a substitute for census-taking and, sooner or later, efforts should be made to take a complete census. A complete census has many important uses, not the least of which is that it provides a basis for numerous statistical inquiries, as has been pointed out above.

The problem of securing satisfactory sampling frames becomes extremely difficult in countries where censuses have never been taken. The very factors which have prevented the carrying out of a census, such as lack of experience or trained staff, also affect the carrying out of sample inquiries. It is advised that in such situations, short and simple questionnaires be used and the entire operation of the survey made extremely easy. As it would not be possible in this case to check the accuracy of the sample survey data by comparison with those of the census, the Working Group of Experts on Sampling Methods strongly advised the use of sub-samples and the indication of the magnitude of standard errors derived therefrom. The Group also felt that if the inquiry took the form of a sample census, it should be used to provide data for the carrying out of subsequent sample studies. Such a sample census has recently been carried out in Saigon. 11/

In conclusion it should be emphasized that sampling methods are potent instruments for obtaining demographic data in addition to what complete censuses normally provide, but the utility of the sample data will be influenced by their inherent inaccuracies in much the same way as the information obtained in the censuses. The inaccuracies that arise in census data have received consideration in other documents presented to this seminar, and the points raised therein have a bearing also on the conduct of sample surveys and the evaluation of their results. Normally, as sample survey undertakings are capable of greater control than those of the censuses, it is possible to ensure a higher degree of accuracy in sample data.

11/ National Institute of Statistics, Viet-Nam, Demographic Sampling Survey at Saigon in 1958. Working Paper presented to the meeting of the Working Group of Experts on Sampling Methods, Bangkok 1959.