UNITED NATIONS

ECONOMIC
AND
SOCIAL COUNCIL

LIMITED

ST/ECLA/Conf.32/L.19
17 May 1968

ORIGINAL: ENGLISH

SEMINAR ON THE ORGANIZATION AND
CONDUCT OF POPULATION AND HOUSING
CENSUSES FOR LATIN AMERICA

Santiago, Chile 20-31 May 1968

Organized by the United Nations Economic
Commission for Latin America, Statistical
Office and Office of Technical Co-operation,
with the collaboration of the Inter-American
Statistical Institute, the Latin American
Demographic Centre and the United States
Bureau of the Census

SOME APPLICATION OF SAMPLING TO POPULATION AND HOUSING CENSUSES

Prepared by ECLA Secretariat

TABLE OF CONTENTS

/I. INTRODUCTION

I. INTRODUCTION

1. Decisions on the specific role sampling is to have in the Census Programme have to take account of the details of the procedures to be employed in each proposed application of the sampling. While there are other even more important considerations, such as possible savings through the use of sampling, needs for personnel to see that various steps are done properly, effects of sampling on the timing of the census programme, and additional costs to be incurred in some cases, there is no escaping the need for examination of the detailed operational aspects of sampling applications. Finally then, it is only in the light of a fully developed plan, setting forth the precise steps to be followed, that a sound decision can be made to accept or reject a proposed use of sampling in the census programme.

2. The steps to be followed in most sampling applications in censuses are in considerable part operational, that is, operational as distinct from tasks performed in a central office only. These steps often involve use of forms and procedures requiring the co-operation of large numbers of persons. Therefore the successful use of sampling in the censuses depends first of all upon having a practicable plan and then being able to make sure that the plan is well understood and that it is being quite rigorously followed.

3. The principal purpose of this document is to describe and discuss important operational details which are suitable for some major uses of sampling in population and housing censuses. The objective not only is to indicate the full detail which is needed, but also in the process to clarify the question exactly of what the methods consist and thereby to show even more clearly than otherwise the advantages which may be derived from the uses of sampling in these ways. At the same time it will become apparent also, the points at which possible failures may occur and the need for staff with adequate training to assure that the different steps are in fact completed in accord with the instruction. It may be noted that in general each of the successive steps is in itself simple and that the successful use of sampling depends mainly upon obtaining a high degree of fulfillment of each small step along the way.

/4. The

4. The document ST/ECLA/Conf.32/L.12 prepared for this Seminar on the "Use of Sampling in Population and Housing Censuses" lists the possible applications of sampling methods in one or more of the following phases of a population and housing census:

(a) tests of census procedures;

(b) enumeration of items in addition to those for which universal coverage is required;

(c) post-enumeration field checks;

(d) quality control of data processing;

(e) tabulation of provisional results; and

(f) tabulation of additional data and data required for special studies.

The present document presents some practical examples on the method of application of sampling to each of the points mentioned earlier at working phase (b), (c), (e) and (f). Therefore the following uses of sampling are considered below:

(a) enumeration of a portion of the census topics by means of a sample of living quarters;

(b) post-enumeration field checks to study census errors;

(c) tabulation of samples of census questionnaires for advance estimates and for certain classes of final estimates.

/II. ENUMERATION

II. ENUMERATION OF A PORTION OF THE ITEMS IN THE CENSUSES BY MEANS OF A SAMPLE OF LIVING QUARTERS

A list sample selected systematically

5. When enumeration of some portion of the items in the censuses is to be done by means of sampling, an early decision to be reached is that of the method of sampling which is to be used. Experience of the past decade in various countries has confirmed the advantages of using living quarters as the unit of sampling in censuses, under conditions such as those existing in Latin America. The living quarters are normally selected systematically, hence the discussion here is based on the assumption that the method of sampling is to be a systematic selection of living quarters within enumeration districts. An exception is made for institutions and other group quarters in which more than, say, twenty persons reside, for which units the sample would be a systematic sample of individuals within each such institution and set of group quarters. To complete the sampling design, there is needed only the decision on the rate of sampling to be used, commonly expressed in terms of a sampling interval. For present purposes this interval will be taken as an example to be 4, hence, the sampling fraction is one-fourth and likewise the probability of selection of each set of living quarters and of each person is 1/4. Since within all countries, the results from such a sample would be published for relatively small geographical areas there is not necessarily any reason to vary the sampling fraction from country to country, assuming that the design is one of selecting living quarters systematically. Moreover, once the decision for a sample of 1/4 is contemplated it is noted that the reduction is already that of 75 per cent of the living quarters, in respect to the sample items, and therefore any further reduction in the size of sample would result in relatively small savings beyond those already visualized.

/6. The

6. The selection of the systematic sample has to depend upon the preparation of a complete list of living quarters, either in advance or at the time of the census enumeration.[1] The listing form must provide a column in which to enter "Sample Key Letter": A, B, C, or D. The system is that the letters are entered in sequence, line by line, on each line in which a set of living quarters is recorded. On the next line after a D has been entered the sequence is started all over again with A. In the event a line is used to list an institution or set of group quarters, no letter is entered on that line but a dash (-) is entered in the "Key Letter" column instead and the letter which would normally have been placed on that line is used on the line for the next set of living quarters. The sampling scheme could be that all living quarters on lines with an A are taken into the sample; hence the selection of the sample follows automatically once the letters have been entered.

When lists of living quarters are prepared in advance

7. In the event that the lists are prepared in advance of the Census the persons who prepare the lists are instructed to leave the "Key Letter" column blank. The task of entering the letter is then made the responsibility of the field supervisor and this work must be done before the lists are distributed to the census enumerators. The supervisor may be instructed to enter a D on the first line for the first enumeration district (E.D.) within each supervisor's area, continuing with A on the second line B, C, D, and so forth. Then when, the lettering of the first list has been completed the second list would be started using the next letter after the last one used on the first list. This method is continued, going from one list to the next until all E.D. lists in the supervisor's area have been given letters. Consequently, it is clear that the initial letter on the lists for all E.D.'s except the first may be any of the letters, being B, C, D, or A when the last letter on the preceding list was A, B, C, or D, respectively. One of the real advantages of the method is that since all sample living quarters are A-living quarters and vice versa, it very soon becomes known by everyone that A-residences are the sample residences i.e., the ones for which the extra questions are asked.

1/ The details on "Preparation and Use of Census Control Lists" see in document ST/ECLA/Conf.32/L.13 of this Seminar.

/Lists prepared

Lists prepared during census enumeration

8. In the event that the list of living quarters have not been prepared in advance but have to be prepared simultaneously with the census enumeration it is necessary to change the procedure just described slightly. In such a case it must be assumed that the first thing each enumerator does, as he arrives at a given set of living quarters for the first time, is to enter the identifying information for that set of living quarters on the next unused line on the listing form. Such preparation of a list of households is not to be regarded as an added task, arising from the use of sampling, but rather as an essential step in any event in order to make sure that no living quarters are later forgotten in cases when no one is found at home on the first visit. As the enumerator prepares his lists he enters the "Sample Key Letter" in the appropriate column, line by line, and again every A-household is a sample household and the enumerator immediately uses the more extensive "sample" questionnaires for all persons living in each such household; (in between A-households there would always be three households for which the shorter "general" questionnaires would suffice).

9. In order that proper entry of letters be made in the "Sample Key Letter" column simultaneously during the enumeration, suitable preparation must have been made. The preparation begins with the printing of listing forms with letters already entered on line 1 (or with the initial letter printed on the first page of listing books when more than one page is required per enumerator). A portion of the pages are printed with each of the four different letters. The supervisor then receives a mixture of forms of the four types for distribution to enumerators in his district. He is instructed not to re-arrange his supply but merely to hand them out as the needs arise in the normal manner. Enumerators then begin with the line already "lettered" and continue the entering of letters in sequence for other lines. Then when an A-line is reached a set of living quarters has been designated for the sample. In the event that the letter printed on the first line is an A the first set of living quarters visited is of course a sample case.

/Some procedural

Some procedural details

10. In the brief review of the processes just given a number of specific steps are specified without giving reasons or explanations for them. The full explanation on these points is beyond the scope of this document but some discussion of these details is needed. It is very important that the details of these methods not be changed.

11. The first point to be noted is that the letters are entered, line by line, for all living quarters as defined for the housing census, whether anyone is living there or not. There are several reasons for this, among them, the fact that the census is not only a census of population but also one of housing and often the characteristics of vacant living quarters are to be considered as well as those currently occupied. There are inevitably also occasional cases in which it is first reported that persons live in a given place and it is then learned on a second visit that no one lives there at present, and vice versa. To attempt at the listing stage to place the burden upon the enumerator always to specify whether a set of living quarters is occupied or not creates unnecessary difficulties, and in fact inevitably leads to problems with the sample.

12. In respect to institutions and group quarters the sampling process is carried through automatically by having the 2nd, 6th, 10th, etc., line on each institution questionnaire marked to show that the persons entered on these lines are in the sample. For all persons on these lines the "sample" questionnaires are used and a representation of the institutional population is therefore obtained.

13. In the case of lists of living quarters prepared in advance, the entering of letters after the lists have been prepared rather than simultaneously has marked advantages since in this way the persons preparing the list have no preoccupation over the matter of the sample. Experience has shown that when the letters are entered simultaneously there are inevitably a few enumerators who purposely re-arrange the sequence of listing households to some degree in order to get the "most representative" ones in the sample. Unfortunately, these efforts always result, over-all, in a bias and one which is easily detected - but not so easily corrected. The system of having the supervisor always use the letter D on the first

/line for

line for the first E.D. (when he enters these letters) has the advantage of simplicity and since lists for the other E.D.'s will start with various letters there is no appreciable bias from this cause, the practical advantage outweighing any theoretical bias.

Tendencies to go beyond instructions

14. Perhaps the most serious problems which arise in the use of this procedure are those due to the fact that well-intentioned persons perform steps which they have not been instructed to perform. They forget that the procedure as outlined gives each household and each person the correct probability of being chosen and that making changes in the letters entered after the enumerator has gone on to the next set of living quarters changes these probabilities, with the result that some of them will no longer be correct. Suppose for example, that in making his review, a supervisor discovers that an A-line was not even a dwelling-place (being a business place only) and that he therefore proceeds to change the letters entered for every household farther down on the list and sends the enumerator out to get long, sample questionnaires filled for those households now on the A-lines. Though not entirely certain it is highly probable that in addition to all the extra work which has now become necessary a bias has been introduced, because in all probability if the lettered line corresponding to the business place had not been an A-line the supervisor would not have made the corrections. The thing that the supervisor should do in a case of this sort is to do nothing at all. It is not intended here to do more than to indicate the subtle character of many errors and biases which may be introduced by making changes in the entries of sample key letters after the enumerator has done his work, but it is hoped that the argument on the importance of a rigid "no-changes" rule is convincing. The only cases in which changes should be permitted would be in the event of specific written instruction to do so from the Central Office. From the point of view of census directors and general technicians, the central point is that the occasional errors which occur unintentionally in the entering of key letters are not the causes of serious bias. Rather the troublesome errors are caused by deliberate changes in some details of the procedures, such as an intention on the part of enumerators or supervisors to "improve" the sample, or by an

/enumerator, to

enumerator, to reduce his work load, through a manipulation which places more small families on A-lines. Above all then, the objective is to see that a strong will is created to obey the simple instructions.

Four steps for sample control

15. Continuing once more with a description of the procedures, it is well established that despite all efforts of organization and instruction of enumerators, minor failures will arise in various ways and for this reason there has also to be introduced a detailed plan of control of the relevant census operations. These control procedures have to include as a minimun the four steps outlined below. Again the methods are not complicated and require only that each step shall be followed.

(a) The first step is a review of each enumerator's work by the supervisor. This review would normally be made in any event at the earliest opportunity after the enumerator has made a start in the enumeration of his E.D. - in order to make sure that he understands generally his work of census enumeration. As regards the sample, the supervisor merely observes that the enumerator has the sample key letters entered in sequence for households on the A-lines. The supervisor makes a written record of any failures on these matters in order that it may later be verified that these have been corrected.

(b) After the work has been completed for the enumeration districts in a census area, certain summaries are prepared and these data are submitted for clearance to the Central Office before the completed materials may be dispatched. The required data are totals of numbers of households and numbers of persons for which sample questionnaires have been completed. Only totals are submitted immediately but the data are summarized by E.D.'s in the District offices. In the event that the data indicated satisfactory consistency, clearance is given at once, probably by telegraph. In the event the totals are such that the area is "out of control", an analysis has to be made by E.D.'s to determine which are the E.D.'s largely responsible for the difficulty. The result may be that additional field work will have to be done for E.D.'s in which it is found that serious discrepancies have occurred, which would lead to some delay in the dispatch of the census materials. This second step is carried through during the finishing stages of census enumeration in the census area.

/(c) The

(c) The next step for control of the sample is carried through after the materials for minor civil divisions, or other areas for which data are to be published begin to be processed in the Central Office. Then for each "publication area" an analysis is made from the punchcards, by age groups, in order to detect cases in which the sample numbers for certain age group are "out of control". In such cases it is then necessary to make corrections in the office through a system of duplication and elimination of punchcards at random. For example, suppose that it is found that within a certain area, the proportion of persons of the age group 20-29 in the sample is only 12 per cent of the total whereas for the entirety of the population of the area the proportion in this age group is 15 per cent of all persons, the procedure would be to select at random a proportion of the sample punchcards for persons in this age group and to duplicate them mechanically and to add the reproduced card to the sample. In like manner, for an age group found to have been badly over-represented in the sample, the procedure would be to select a proportion of the sample punchcards at random and to eliminate them from the further processing of the sample data. The process here described is a technical one which is done by a small group of persons in due course of time, but the <u>need</u> for the process must be taken into account. Moreover, to the extent that there has been serious failure with any of the earlier control procedures, the amount of work at this stage is likely to be seriously increased with consequent reduced value of the sample data. It is known that these processes do not eliminate errors completely but only reduce them, hence it is not desired to carry through any more of this type of work than is found to be absolutely necessary.

(d) Finally, there are steps at the estimation stage, during the course of the electronic tabulation (if a computer is to be used) or after the "running" of the punchcards through the tabulating machine, by which further attempts are made to correct for bias still found to exist in the sample cases, such as if there is an under-representation or over-representation of single males for example. These processes are rather easily applied in the case of use of computers, but they are also feasible and practicable without the help of computers. Fundamentally, the estimation process consists of a system of weighting the sample data by factors which are determined on the basis of the complete data from the census for the area in question.

/<u>Acceptance of</u>

Acceptance of many small errors as inevitable in censuses

16. A fundamental concept in the use of sampling is that of the inevitability of some error and consequently the acceptance of results which are known to contain imperfections. The acceptance of this idea is necessary for the proper use of the method because the system becomes an impractical one if it is intended that no error should be left uncorrected. In the first place, a philosophy of correcting every error would make it necessary at numerous points in the operations to decide whether a given departure from an ideal quantity (from some point of view) was the result of an error of some kind or whether it was merely due to a normal variation in sampling.
The decisions to be taken rather are ones of examining whether differences of the order of magnitude found are acceptable or whether a correction of same kind must be made. Based on theory and on substantial experience with these systems, rules have been developed by which it can be seen at once whether a given deviation requires attention. Furthermore, it does not follow automatically, as in the case of the age group containing only 12 per cent instead of 15 per cent of the persons, that the correction to be made would be exactly that to bring the sample proportion to the 15 per cent level. Taking account of the larger territorial unit of which the area under consideration is only a part and to which other units belong where over-estimates might occur as well as some other under-estimates of minor importance, it may be found better to increase the proportion only to 14.2 per cent for example. This shows that notwithstanding the necessity for rigorous application of certain aspects of sampling processes a great deal of flexibility exists. It may be seen that this idea of acceptance of imperfections has to permeate the thinking of top staff in the census programme, as indeed the errors in censuses permeate all data to be published. From one point of view, the notions of errors found in sampling are helpful to census work generally because these ideas lead to balanced views on the magnitudes of errors which are acceptable, and on the extent of efforts which should be made to correct errors of all kinds.

/III. POST-ENUMERATION

III. POST-ENUMERATION FIELD CHECKS TO STUDY CENSUS ERRORS

17. The method of sampling for post-enumeration field checks is normally a multi-stage procedure in which minor civil divisions (M.C.D.'s) are selected at the first stage and enumeration districts (E.D.'s) at the second. In this way a sample might consist for example of a total of ninety E.D.'s selected, three per M.C.D., within thirty M.C.D.'s selected at the first stage. Under some conditions the E.D.'s of the sample might be sub-sampled and the post-enumeration, carried through only in the selected sub-parts of the sample E.D.'s.

Assignment of personnel and organization of post-enumeration studies

18. A major requirement in carrying through such studies is that separate office staff shall be assigned in advance to organize the undertaking and that these few persons shall not be taken away from this task at the stage of census enumeration, such as to assist in the census proper temporarily, because it is precisely at the enumeration stage that critical steps have to be taken to complete arrangements for, and to begin the post-enumeration check. The list of E.D.'s of the sample must of course be kept a closely guarded secret until the actual enumeration has been completed, and to the extent that those who are going to be supervising the post-enumeration study are in the "field" to begin the selection of enumerators for the study, they also should not at first know which are the E.D.'s of the sample. At the critical stage, names of prospective enumerators are accumulated within the sample M.C.D.'s and arrangements are made with them to do the further work, that is, the field check with the knowledge that the E.D. in which each will work will be other than the one in which he worked originally and that the identity of the E.D. will not be known to him until a few days before the work is to start.

19. As a practical matter, the exact E.D.'s to be included in the study may not be selected until the enumeration is under way because many times the division of M.C.D.'s into E.D.'s is not completed as far in advance as it should be and the exact number of E.D.'s is therefore not known. Consequently, the "pre-designation" of the sample might have to consist of giving a list of sample M.C.D.'s together with an instruction on how the

/sample E.D.'s

sample E.D.'s are to be selected. This instruction might be the 4th, 24th, 44th, etc., E.D. after a sequence of E.D.'s has been established, say those in the first sub-area according to number, followed by those in the second sub-area and so forth throughout the sample M.C.D. These details are included to indicate the processes which have to be used to assure that all is in readiness for the field check to begin immediately once the census enumeration itself has been finished within the particular M.C.D.

Timing for field check and subsequent visits

20. Operationally, the timing for the study has to be given the highest priority. It is essential that the re-enumeration shall have been completed within ten days after the date of the census itself. Such a timetable of course means that all the special instructions for the enumerators shall have been completed in advance in every detail and that copies will have arrived at the area offices containing the sample M.C.D.'s prior to the census date. A special training session is required for the enumerators selected for the field check and recognition has to be given from the start to the fact that discrepancies will be found which will require subsequent return visits to a portion of the households within sample E.D.'s from two to four weeks after the re-enumeration. Experience has shown clearly that the field check has to be carried through independently of the census itself, that is, without any reference whatsoever being made to the materials completed by the original enumerators for the E.D.'s, prior to the re-enumeration.

Distinction between coverage errors and content errors

21. Post-enumeration field checks normally include investigation of two kinds of errors which occur in censuses, namely:

(a) errors in coverage, and

(b) content errors.

Coverage errors are those due to leaving out or including erroneously certain individuals or even entire households in the census, while content errors are those due to the fact that answers obtained for persons enumerated in the census are nevertheless not entirely correct. In the case of content errors complications arise at the analytical stage because of inability to determine

/in many

in many cases which of two conflicting answers is the more correct. On the whole the priority has to be given to the study of errors in coverage and for this reason the details to follow relate only to the carrying through of this part of the investigation.

Comparisons of individuals enumerated in field check and in census

22. Immediately, that the sample re-enumeration has been completed there is the necessity for making the comparisons, household by household and person by person, of the field check against the census itself for the E.D. This work may have to be organized at Regional offices and it may necessitate sending Census materials to this office temporarily even though the final clearance of the materials has not been completed within the local area office. In such a case the comparisons would be carried through and the materials then returned to the area office. If then further steps are carried through within sample E.D.'s as a part of the census itself, it is at least clear that such changes were made subsequent to this step because the comparisons have already been recorded on the basis of the materials as they existed at the supposed completion date of the enumeration within the particular M.C.D. By this time, the fact that a given E.D. was in the field check will be known to all and if additional families are later added in the Census the reasons would be studied carefully to determine whether this additional work was or was not independent of the fact of the post-enumeration field check.

Reconciliation of differences between field check and census

23. The next and final step in the field work requires the greatest objectivity and persistence, and what is more, often contains real difficulties. This step is that of the reconciliation of major differences which have come to light from the individual comparisons, line by line, of field check with Census. In this document, it is possible only to touch upon a few of the matters to be resolved and perhaps to indicate the great variety of problems which are presented. One of the first objectives of the study is to find out if entire households and persons within households have been left out of the Census or have been included erroneously. However, the finding that there were differences between re-enumeration and Census by no means assures that such errors have been discovered. It may later be found that the family apparently omitted from the Census was in fact enumerated in a

/neighbouring E.D.

neighbouring E.D. or that the family was not actually living there on the census date. Similarly, further investigation may reveal that the family enumerated in the census and not found in the re-enumeration was merely missed in the field check. Likewise, the variety of explanations which may be found for differences between re-enumeration and Census on individuals within households is even greater. It may be found, for example, that a given person should have been enumerated in the family where found in the field check but that this person was nevertheless included in the family where he used to live in the census, such as at the home of his parents in another Province, and that it is therefore not true that he was omitted from the census. Similarly, a person may be found to have been erroneously included in the household within the sample E.D. but investigation shows that he was omitted from another household in which he should have been enumerated in the census. Studies on the quality of the census have to take all such factors into account if truly useful results are to be found. It is valuable to know about cases in which persons are enumerated within households other than the ones in which they should have been enumerated but it is necessary to make the proper classification of the error in the quality check and not to count cases as duplications or omissions in the census when such errors did not in fact occur in those particular cases. A tremendous number of complications come to light in the process of trying to resolve inconsistencies between field check and census.

Problems which still arise due to finding of inadequate information

24. Basically, many of the problems arise in the reconciliation of the data because it is later discovered that certain details of the troublesome cases had not in fact been obtained during the field check. Consequently, although only a few questions are verified in the study, it is necessary that related questions be asked on the questionnaire in order that a basis be provided for discovering the source of the differences in so far as possible without requiring unnecessary and costly re-visits. Obviously, it is also necessary that in the training of enumerators a great emphasis be placed upon asking all questions carefully and upon recording in full the information obtained during the re-interviews. Actually, the number of possible explanations for discrepancies is so great that it is not practicable to

/include sufficient

include sufficient questions in the re-enumeration form to take account of all of them. For this reason, great dependence has to be placed upon the enumerators in the field check to be on the alert to listen to unusual circumstances as described voluntarily by respondents and to enter these explanations in space provided on the field-check questionnaires. Even at best, however, a great many instances are certain to be found in which the only way to resolve an inconsistency is to make another visit to the household in question. What is more, unless great care is used in organizing the details of the case it will be found that even the third visit has not resolved the matter adequately and a fourth visit becomes necessary. The longer the matter remains unresolved the more frequently new visits merely produce new inconsistencies because with the passage of time the facts have become obscure. It is therefore imperative that the re-visits shall have been completed for 80 per cent or more of the cases within two months of the Census date.

Use of birth registration in a coverage check

25. Another approach by which a sample of cases is obtained for a Post-enumeration field check is to begin with a sample of registered births, say those shown to have occurred during the twelve months preceding the Census date within the sample of M.C.D.'s and then to seek to find the enumeration of the infant in the appropriate file of census questionnaires. For those not found in this manner, details are sent to local officials of the minor civil division who seek out the present residence of the family and child and the information obtained is used for further searching in the census files. Finally, a list is accumulated of infants missed in the Census enumeration and enumerators are sent to the present residences of the families to fill out a sample questionnaire for them, which in turn leads to further searching and a possible determination that the family was missed, or perhaps merely that the child was missed although the family was enumerated. Such a method serves to estimate the proportion of infants omitted from the Census, and even the proportion of families and persons in families in which infants are living, provided that it is known that a high percentage of births are registered within the country. This type of study has much to commend it, in part because it can be used simultaneously (by beginning with a

/parallel sample

parallel sample of infants enumerated in the Census) to estimate the proportion of births omitted from the registrations. However, the estimated percentages of error are likely to be under-estimates to a degree because a higher proportion of births not registered are likely to correspond to infants omitted from the Census (including cases of infants who lived only a few days) and a higher proportion of cases not enumerated in the census are likely not to have been registered.

26. The preceding suggests some of the difficulties in estimating proportions of omitted infants in the Census; however these difficulties are on the whole appreciably less than those of arriving at meaningful quantitative estimates of omissions and duplications in the census, based on the normal post-enumeration field checks. The difficulties of finding all omissions and of resolving all the discrepancies as required for the estimation process are often so great that the resulting figures clearly have little or no value. For example, in the event the formal estimate turns out to be a net omission of 3 per cent when other sources, such as a cohort analysis or a small check on enumeration of registered births, would indicate that a figure of 5 per cent to 8 per cent was more reasonable, it is difficult to say that the formal estimate has any value. It may be noted however that Post-enumeration field checks have great value in enabling those in charge to understand census problems and in teaching us all the importance of rigorous attention to details of definitions and instructions, whether or not the results are finally reduced to meaningful quantitative estimates.

/IV. TABULATION

IV. TABULATION OF SAMPLES OF CENSUS QUESTIONNAIRES FOR ADVANCE ESTIMATES AND FOR CERTAIN CLASSES OF FINAL ESTIMATES

27. Samples of census questionnaires selected in the Central Office can serve two important purposes in the census programme. Such samples may be used (1) to prepare advance estimates of census results as soon as possible after the Census enumeration, and (2) to prepare additional tables beyond those required for the 100 per cent tabulation programme, including cross-classifications of data. The first purpose is that of obtaining timely Census results, since the complete tabulation programme usually takes a relatively long time; the second purpose is principally that of reducing costs, but it may also have the effect of speeding up the tabulation programme. The second use of a sample may be seen to complement that of enumeration of a portion of the items in the Censuses by sampling (Part II), since in some instances, (rural areas for example), it may be judged better to ask all questions at all households in the enumeration stage and then to select a sample as presently discussed, at the tabulation stage.

Some practical considerations on selection and implementation of the sample

28. As a rule, the samples for the two purposes discussed here are different samples, the sample for the first purpose being much smaller than that for the second. When a sample is to be chosen for the purpose of advance estimates of census results it is necessary to organize the task in such a way as not to interfere unduly with the normal processing of the census. Even so, some additional costs are inevitable and these have to be budgeted and the necessary personnel assigned to the work. Despite the fact of some additional operations, the editing and coding of sample questionnaires and the cardpunching may be made to serve both the full census tabulations and the advance estimates through the duplication of sample punchcards and the retention of the original cards for inclusion in the normal census tabulations. The additional costs are then the costs of sampling and that of the preparation of the tabulation for the advance estimates.

/29. The

29. The sample design for advance estimates may be extremely simple, such as that of merely taking every kth household throughout the country and every kth individual within the institutions. More commonly however, there is the use of a two-stage design in which a sample of "work units" is first selected within M.C.D.'s and then the procedure above mentioned is used within the sample work units. The total size of sample might for instance be about 10,000 households, or even more if advance estimates are required for two or more Regions within the country, such as the Federal District and the rest of the country for example. Sometimes the plan is such that the M.C.D.'s within which the sample work units are to be selected are chosen in advance of the census and then procedures are instituted for having the census materials for the sample M.C.D.'s shipped in advance from the various regional and area census offices. Such a plan may require that priorities be established for the clearing of materials by M.C.D.'s, making sure that the M.C.D.'s needed for the advance estimates are always cleared for shipment before those for other localities, that is, provided that materials for these M.C.D.'s have been assembled and have been found to be in order. The matter of separate shipment of materials for certain M.C.D.'s, or even for certain sub-areas within M.C.D.'s, does not create serious difficulty provided that there is an adequate system of control of all census materials so that it is known at any time exactly where the questionnaires for a given M.C.D. and sub-area are to be found.

30. The use of simple sampling fractions becomes a great convenience in the selection and use of the sample for advance estimates. For example, in the event of a sampling fraction of 1/100 it may serve well to select every 10th work unit and then every 10th household within the selected work units. Variations of these rates are easily found to fit individual cases: every 20th unit and then every 10th within units if the over-all fraction is 1/200th, every 5th units and every 4th within the units for an over-all fraction of 1/20, and so on. In the estimation process it follows that there is merely required the multiplication of the sample totals by 100, 200, or 20 as the case may be. The process of selecting the sample households for the advance estimates is facilitated by the fact that lists of all households showing numbers of persons per household will have been

/prepared by

prepared by each enumerator or by the census district offices at some stage in the regular census work. In the office then the task is one of combining totals by E.D.'s within work units and preparing a list of these work units recording numbers of households within the unit. Suppose that in a particular case every kth work unit means that every 10th work unit is being selected (k=10) within all M.C.D.'s. The process is merely one of choosing a beginning point by means of a random number and then checking off the corresponding unit and every 10th one thereafter for the sample. Within the selected work units, the method is the same, suppose each 5th household, and the households for the sample are then checked off on the special lists of households for E.D.'s belonging to the work unit, continuing from one E.D. to the next as though the whole set of E.D.'s for the work unit constituted a single list. Great flexibility exists in the details of these processes, except that a decision must be made on exactly the process to be used in a given case and then adequate supervision is required to make sure that the process is followed without exception.

Processing of the sample

31. After the sample has been selected within a given work unit the process becomes one of finding the corresponding file of questionnaires and taking out the designated questionnaires replacing each with a small form to show that this one has been removed for purposes of the sample for advance estimates. Otherwise, the record of those removed may be written on the cover of the work unit, if such a procedure is more practicable in the given case. Files of sample questionnaires may be organized in which all those within a given sub-division of some kind are placed in a single new sample work unit, in order to facilitate the subsequent editing, coding, etc.

32. By the time parts of the sample are available for processing, the normal procedures of census editing and coding should have been started, and the system is then one of incorporating the files of sample questionnaires in the regular continuing operations, but giving the sample files priority over the regular census work units. However, there would be no variation whatever in the processes of editing and coding as between sample and non-sample work units, hence the total work to be done at this stage remains unaffected by the selection of the sample for advance estimates. Similarly, the normal

/operations of

operations of cardpunching and verification would be started as quickly as possible and again the files of sample questionnaires would be included simply as work units having the priority over the regular census work units, but not affecting the total quantity of punching to be done.

33. The eventual merging of the sample punchcards with those for the remainder of the census is facilitated by preparing duplicate copies (automatically) at the appropriate time, leaving the original sample cards for the complete census operations. The duplicated copies, clearly distinguishable in some way, are then the ones which are used to prepare the prior tabulations and the advance estimates. Needless to say, the complete set of tables to be run off with the special sample must have been agreed upon in every detail prior to the census date.

Calculation of sampling standard errors

34. In the event that the tabulating of data are to be done on a computer the entire plans have to be developed to conform with the system which is appropriate for the particular computer. In such a case it is probable that with little difficulty the computer programme can be prepared so as to make calculation of sampling errors simultaneously with the preparation of the tables of advance estimates. In the event that the tabulations are to be prepared by the older-type machines, the estimates of sampling errors will have to be approximated separately, very likely through development of a system of generalized tables for the benefit of users of the data. Such methods have been in use from time to time and do not cause particular difficulty.

Necessity for rigorous application of sampling process

35. As in the cases of the other uses of sampling, pitfalls are to be found in the application of the procedures because persons do not always follow instructions precisely and perform steps which may appear to be required but in fact generally lead to small biases. An example is the substitution of the next household for an institution, in the event the kth selection happens to be a collective group rather than a household. It has been mentioned that a separate procedure has to be established for selecting a sample of individuals within institutions, hence there is no selection of them simultaneously with the selection of households.

/The proper

The proper procedure then when the selection happens to be an institution is simply to omit it and go on to the next sample case. With a little care, the procedure may be carried through virtually free from bias, or at most the bias will be trivial. An advantage is that in due course of time the advance estimates can be compared with the figures finally obtained and published for the census as a whole, and it can then be seen whether the magnitudes of errors are about as anticipated, and to what extent, if any, a bias seems to have been introduced at some point along the way.

The sample for final tabulations

36. It is frequently the case that the sample for final tabulations is selected by a machine process from the punchcards themselves. Sometimes a relatively large sample may be required and sometimes one of quite modest size. The basic point to have in mind is that it is the absolute size of the sample which is important and not the proportion of the total which is taken for the sample. Often therefore, samples of adequate size may be obtained by selecting each 10th punchcard, each 20th, or perhaps as much as each 5th may be needed. It is often quite easy to calculate the savings to be achieved from the use of such samples since the machine time for processing will be reduced in proportion to the sampling fraction which was used, except that one run of all the cards may be necessary in order to select the sample. The discussion here is based on the assumption that the tabulations are being performed without the use of a computer and therefore that the runs for certain of the tables will require passing the cards through the machines several times, such as is required when cross-classifications are made on the basis of a number of variables. The savings may be calculated not only in terms of the reduced hours of machine time but also in the weeks or months gained in the completion of the Census tabulation programme.

A variation in design when a sample only is to be processed

37. In the event the sampling is done in the office as a substitute for having taken a sample at the stage of enumeration, such as if the rural areas were enumerated in full and are then to be sampled in the office, the sampling is again one of the questionnaires themselves; in such a case, the cards will not have been punched and it will be most important to make

/the saving

the saving on the costs of editing and coding as well as of cardpunching. The sampling process would then be somewhat similar to that for the advance estimates (and also in fact quite similar to that used by enumerators in the field), and the needs will probably require a very large sample, such as 1/5 or 1/4. It is clear that the work involved is time-consuming but that it will nevertheless result in a large saving as compared to the processes of complete editing, coding and cardpunching.

V. CONCLUDING REMARKS

38. As will have been noted from the foregoing, the use of sampling in population and housing censuses can have a major impact on the operational procedures in carrying through census programmes. In the case when all census topics which require lengthy coding operations, such as those on economic activity for example, are asked only for a sample of living quarters, the shorter "general" questionnaire may quickly be sent on for mechanical processing and the coding of the difficult topics even may be completed in only a fraction of the time that would have been required without the use of sampling. The effect then is to place great emphasis upon the data processing plan (Item H of the Draft Census Calendar) in order that all will be in readiness for the beginning of the census processing at the earliest possible time. On the other hand the use of the sampling may have permitted the inclusion of more questions in the census than would otherwise have been the case with the result that more punchcards are designed for the processing or at least a separate punchcard has to be used for the sample topics.

39. Likewise the possibility exists of introducing sample verification in the coding and punching operations (see paragraph 4, item d), a process which would have considerable effect on the over-all census plan as well as on the specific steps in the processing.

/40. The

40. The rapidly growing needs for extensive and reliable demographic and other social and economic data which have not been available in the past have also influenced census methodology in the sense that changes in methods are constantly being introduced. As is well known, complete census taking is very expensive and complete reliance on 100 per cent enumeration and 100 per cent processing prior to release of information results in a magnitude of census operation which would probably place a serious strain on the available resources. To lower the cost of census taking and to obtain the needed quantity and detail of information of satisfactory quality, it is more and more recognized that the solution in many instances is to include sampling as an integral part of the census and to develop the operational details accordingly.

41. It is inevitably found that planning and execution of census operations must be developed very fully and there is an ever-growing need for technical staff. There is evidence however that an increasing number of technicians can be and are being trained to perform the challenging tasks which are involved in modern census programmes.