UNITED NATIONS

ECONOMIC AND SOCIAL COUNCIL

GENERAL

E/CN.3/114
E/CN.3/Sub.1/30/Rev.1
28 September 1950
ORIGINAL: ENGLISH

STATISTICAL COMMISSION

SUB-COMMISSION ON STATISTICAL SAMPLING

Fourth session

REPORT TO THE STATISTICAL COMMISSION ON THE FOURTH SESSION OF THE SUB-COMMISSION ON STATISTICAL SAMPLING, 5-15 SEPTEMBER 1950

TABLE OF CONTENTS

INTRODUCTION

1. The Sub-Commission on Statistical Sampling held its fourth session from 5 to 15 September 1950. The Sub-Commission re-elected Mr. P.C. Mahalanobis as Chairman.

2. The following members were present:

Mr. P.C. Mahalanobis (Chairman)
Mr. W.E. Deming
Mr. F. Yates

The specialized agencies were represented by:

International Labour Organisation:	Mr. A.A. Evans
Food and Agriculture Organization:	Mr. P.C. Tang
World Health Organization:	Mr. G.E. Hill

3. The following representatives of non-governmental organizations attended:

Category A:

International Chamber of Commerce: Mr. G.L. Ridgeway

4. Mr. W.R. Leonard, Director of the Statistical Office of the United Nations, acted as the representative of the Assistant Secretary-General for Economic Affairs. Mr. W.J. Bruce, Assistant Director of the Statistical Office of the United Nations, served as Secretary to the Sub-Commission. Other members of the Secretariat who participated in the discussions of the Commission were:

Mr. J.B.D. Derksen
Mr. W.W. Flexner
Mr. S.B. Sen
Mr. C. Chandrasekar
Mr. J.D. Durand
Mr. M. Lacroix

5. The Sub-Commission wishes to express its appreciation to the following statisticians who took part in the discussions at the fourth session:

Mr. E.P. Billeter	Statistiches Amt der Stadt Zurich
Mr. R.H. Blythe	U.S. Department of the Air Force
Mr. S. Dutka	Dun and Bradstreet
Mr. R. Goodman	Survey Research Center, University of Michigan
Mr. M. Hansen	U.S. Bureau of the Census
Mr. W.G. Madow	University of Illinois
Mr. C.M. Mottley	U.S. Department of the Air Force
Mr. Y. Morita	Statistical Bureau, Japan
Mr. W.A. Shewhart	Bell Telephone Laboratories
Mr. R.T. Smith	U.S. Interstate Commerce Commission
Mr. R. Wilkinson	Bell Telephone Laboratories
Mr. S.S. Wilks	Princeton University
Mr. J. Van Yzeren	Central Statistical Office of Netherlands

/6. The Sub-

6. The Sub-Commission adopted the following agenda for its fourth session:

A. Constitution and membership of the Sub-Commission.
B. Comments on syllabuses drafted in the third session.
C. Operational research.
D. Sampling in tabulation.
E. Sampling to measure goods traffic by road.
F. Sample surveys of current interest.
G. Sampling methods for estimating distribution by size of individual and family income.
H. Methods to collect migration statistics.
I. Crop statistics.
J. Other subjects:
   a. Theory on extreme values and its technical application.
K. Report of the fourth session to the Statistical Commission.

I. CONSTITUTION AND FUTURE ACTIVITIES OF THE SUB-COMMISSION

7. The Sub-Commission had before it the request of the Statistical Commission to discuss the constitution and membership of the Sub-Commission (E/1696/Rev.1). This subject arose out of a discussion in the Statistical Commission on the future programme of the Sub-Commission.

8. In its report of the fifth session, the Statistical Commission observed that "the Sub-Commission has hitherto been most successful when the sampling problems on which its advice was sought have been submitted in detail and with full documentation"; and continued: "Arrangements are needed by which specialists in the subject fields in which the Sub-Commission is considering the use of sampling methods may participate more directly in the work".

9. In this connexion the Sub-Commission sought the views of the representatives of the specialized agencies and non-governmental organizations on the work of the Sub-Commission.

10. The representative of the International Labour Organisation stated that his organization found the recommendations of the Sub-Commission very useful. The Sub-Commission had already, at the request of his organization, given consideration to problems such as cost-of-living studies, family-living studies, and man-power statistics. He felt that sampling techniques would be increasingly used by labour statisticians in the future and the Sub-Commission's advice in regard to sampling methods would be an important help.

11. The representative of the World Health Organization expressed the view that his organization definitely hoped that the Sub-Commission would continue to function so that in the future, as it appeared very probable, problems might be brought before the Sub-Commission for its recommendations.

12. The representative of the Food and Agriculture Organization stated that his organization would like to see the Sub-Commission continue its work. The FAO from the very first session presented its problems before the Sub-Commission and found the recommendations useful. He felt that with the development of technical assistance programme for under-developed areas, the necessity for reliable statistics would be felt more than ever before. It is likely, therefore, that increasing help would be sought from the Sub-Commission in the future.

13. The representative of the International Chamber of Commerce stated that his organization had been engaged through its Distribution Committee in promotion of

/the use of

the use of sampling methods for the collection of data on distribution in the years between the distribution censuses. The International Chamber, therefore, had for this reason a continuing interest in the work of the Sub-Commission.

14. The Sub-Commission fully agrees with the Statistical Commission that sound advice on the application of sampling methods in specific fields can only be given when adequate information is provided on the material to be sampled and the problems involved. The Sub-Commission has always insisted on such information. It has also arranged, wherever possible, for the attendance of specialists in the particular subject field under discussion. Moreover, the Sub-Commission has frequently stated, that the details of the most appropriate sampling scheme to deal with any particular problem can usually be worked out only by a sampling expert who is in contact with the preliminary stages of the work, and who is in a position to carry out investigations into the accuracy of the various possible methods, and who if necessary can institute a pilot survey.

15. The Sub-Commission has recognized that its functions do not include those of detailed advice and experimentation; these therefore, have been regarded as the responsibility of the statisticians in charge of the survey. Indeed, the Sub-Commission has continually stressed the need for the training of statisticians in the methods of sampling.

16. In addition to its work on the application of sampling methods in specific fields, the Sub-Commission has concerned itself with the more general aspects of sampling methods, and as instructed in its terms of reference, has endeavoured to "promote the improvement of methodology in statistical sampling and its applications". To this end a standard terminology was drawn up which has not only served the immediate purpose of standardizing reports of sampling surveys but has also resulted in considerable clarification of the logic of sampling methodology. Two of the Sub-Commission's members have produced textbooks on sampling, one of which would certainly not have been written had it not been that the need for such a book was expressed by the Sub-Commission. In the opinion of the Sub-Commission these more general aspects have been as successful as the more specialized work referred to by the Commission, and indeed have laid an essential foundation for work in specialized fields.

17. In the opinion of the Sub-Commission, therefore, it would be a grave error to attempt to confine its work to specialized applications of sampling methods

/to the

to the exclusion of such general problems as appear to it to be of relevance. The Sub-Commission, in fact, has been considering, in conjunction with the discussion on operational research reported below, whether the scope of its work should not be widened somewhat so as to include as its subject matter other statistical techniques which in its opinion are likely to be of value to Members or organs of the United Nations and the specialized agencies. It has come to the conclusion that this would be advantageous for the following reasons:

(a) Some of the existing and developing statistical techniques used in scientific research are likely to be of value to Members or organs of the United Nations and the specialized agencies.

(b) The United Nations and the specialized agencies have at present no body other than the Sub-Commission which can advise on the application of such techniques.

(c) The constitution of the Sub-Commission is at present such that it is well qualified to deal with such problems, as it is composed of statistical experts whose collective experience covers a far wider field than that of sampling methods.

18. The Sub-Commission, therefore, wishes to recommend that it be authorized to undertake work along these lines. It is not anticipated that at the outset the amount of such work would be very large. If the volume increases to such an extent that sub-division of the work of the Sub-Commission appears desirable, this can be considered when the need arises.

19. If this recommendation is accepted the Sub-Commission suggests that the following revised terms of reference be adopted:

> The Sub-Commission on Statistical Sampling shall assist the Statistical Commission to promote the development and use of such statistical techniques as are likely to be of value to the Members or organs of the United Nations, the specialized agencies and non-governmental organizations brought into consultative status with the Economic and Social Council of the United Nations.
>
> In particular, it shall:
>
> (a) Advise on the use of sampling methods in the collection of information needed by Members or organs of the United Nations, the specialized agencies and non-governmental organizations, and make recommendations on the methods it considers most appropriate

/to specific

to specific fields under the varying circumstances existing in different countries;

(b) Keep under review the use that is being made of statistical sampling in different countries, and in different fields of subject matter;

(c) Determine in what circumstances sampling methods are preferable to complete enumeration in respect to reliability, speed, cost, continuity and other factors and make recommendations regarding the use of sampling methods in conjunction with attempted complete enumeration in order to assess the reliability of that enumeration, or to obtain early returns and supplementary information;

(d) Bring to the notice of Members developments in other statistical techniques which appear likely to the Sub-Commission to be of use in the solution of problems which are directly the concern of Members or organs of the United Nations, the specialized agencies and non-governmental organizations, and make recommendations regarding their use in specific fields;

(e) Promote the improvement of the methodology of statistical sampling and other statistical techniques which are the concern of the Sub-Commission, and their applications.

20. In establishing the Sub-Commission, the Statistical Commission selected the experts from different countries to serve in their individual capacities and the experience of the last three years has confirmed the wisdom of this decision. Because the members were serving in their individual capacities they felt free to express their views on technical questions without in any way committing the official policies of their governments. Moreover, the Sub-Commission has encouraged the active participation of individual experts in different subject fields and has profited greatly by such participation. The Sub-Commission strongly recommends that these practices be retained .

21. The Sub-Commission's membership was originally set at nine members, but no more than five members have been invited to serve. In addition, on the recommendation of the Statistical Commission the Secretary-General requested one consultant to serve with the Sub-Commission. In view of the present small membership of the Sub-Commission and in view of the fact that members are

/appointed

appointed in their personal capacities and no alternates can be recognized, the Sub-Commission believes that advantage should now be taken of the authorization for nine members. In the opinion of the Sub-Commission, however, expansion could best be accomplished at this time by the appointment of one to three additional consultants. The initial appointments should be for a period of one year.

22. To make provision for the participation of subject-experts, as recommended by the Statistical Commission, it is recommended that the Secretary-General, in cases in which this appears advisable, be authorized to appoint experts to assist the Statistical Office and to attend meetings of the Sub-Commission.

/II. COMMENTS ON THE

II. COMMENTS ON THE SYLLABUSES DRAFTED IN THE THIRD SESSION

23. Because of shortage of adequately trained statisticians in sampling, the Sub-Commission in its reports had stressed the need for training statisticians in sampling. Requests had been received from specialized agencies in regard to the development of a training programme. It was felt that the Sub-Commission could broadly indicate the level of training which should be expected of a statistician qualified in the field of sampling. The Sub-Commission also wished to stimulate discussion on suitable syllabuses for courses in sampling. With this object in view two members at the third session of the Sub-Commission prepared syllabuses and the Statistical Office was requested to circulate them to appropriate organizations and persons for their comment. It was realized that while the primary responsibility for drawing up definite syllabuses or conducting courses on suggested lines must lie with universities, institutes and government and private organizations, the Sub-Commission might encourage further consideration and activity in this direction, especially when no generally recognized suggestions regarding syllabuses had hitherto been made. It may also be noted that the Sub-Commission did not endorse any of the syllabuses but only requested their circulation to obtain views from persons interested in education and training in sampling.

24. The large number of comments received by the Statistical Office clearly indicates that the suggested syllabuses had been successful in creating widespread interest. Thirty-one statements have been received from statisticians or groups of statisticians from seventeen countries and most of these are useful and interesting. As the Sub-Commission does not propose to draw up sets of specific syllabuses, it will not pursue the matter further at this time.

25. The Sub-Commission, therefore, requests the Secretary-General to transmit the suggested syllabuses and the comments to the International Statistical Institute and to express the hope that the Institute will give further attention to this subject. The Sub-Commission would be pleased to receive from the Institute reports of further progress.

/III. OPERATIONAL

III. OPERATIONAL RESEARCH

26. The Sub-Commission had a general discussion on the subject of operational research (E/CN.3/Sub.1/25, E/CN.3/Sub.1/28, E/CN.3/Sub.1/L.5). The object of the discussion was to ascertain what activities are currently covered by the term, and to determine whether it appeared advisable for the Sub-Commission to concern itself with the subject, or with particular aspects of it.*

27. It was represented in the course of the discussion that the methods of operational research as now developed are capable of making useful contributions to the welfare of the peoples of the world through the more effective solution of the problems arising in governmental administration and planning. In addition to their use in fully developed countries these methods are likely to be of particular value in the development of countries which are at present under-developed and technically backward. This is because the uncritical adoption of technical processes which have been found satisfactory in more fully developed countries is always dangerous, and particularly so if the rate of development is rapid, since with rapid development there is little time to remedy defects which only become apparent after the processes have been in use for a number of years. Even if the processes are in themselves satisfactory, their use frequently gives rise to economic and social problems which cannot be properly solved without scientific investigation of the operational research type.

28. The discussion, however, revealed that there is at present no clear or agreed definition of what constitutes operational research (or, as it is alternatively called, operations research, or operations analysis). Moreover, there is a tendency to stretch the term to embrace so many types of investigation that it tends to lose much of its value. It was reported that a National Research Council Committee had been set up in the United States to define the subject more clearly, and that the Operational Research Club of the United Kingdom was also concerning itself with this problem. In the circumstances, the Sub-Commission believes that nothing would be gained by any attempt on its part at a definition, but requests the Secretary-General to keep it informed of any progress made.

---

* The Sub-Commission wishes to express its thanks to the following statisticians who participated in the discussion: Messrs. Blythe, Goodman, Madow, Mottley, Shewhart, Wilkinson and Wilks.

/29. On the other

29. On the other hand the Sub-Commission recognizes that it is generally agreed that operational research requires the application of scientific research methods to the problems involved, and that such research could only be conducted properly if there were close liaison at a sufficiently high level between the research workers and the administration or management.

30. The Sub-Commission also decided that it does not consider it appropriate at the present time to concern itself with the general field of operational research. It recognizes, however, that certain types of investigation, which at least in some quarters would be classed as operational research, are likely to contribute to the more effective solution of problems arising in governmental administration and planning. In particular it has in mind the possibility of investigations into agriculture and small scale industry by means of surveys and planned experiments. Inasmuch as the statistical techniques used in these types of investigation consist in large part of sampling methods, and moreover as various members have had considerable experience of such investigations, the Sub-Commission considers that it should give consideration to methods appropriate to these problems when the need arises and be prepared to advise on them.

/IV. EARLY CENSUS RESULTS

IV. EARLY CENSUS RESULTS BY ADVANCE PROCESSING OF A SAMPLE

31. Some of the important census results for a country or for a large sub-division thereof may be obtained by advance processing and tabulating of a sample of the returns. The reduction in time may be considerable - for example, the reduction in time may be in the ratio of two or three years to four months. An advance tabulation programme of a population census could be used to produce statistics, for example, on characteristics of the following type:

The population count, by sex and age classes, urban and rural areas and other simple categories;

The number of households, by sex of head, size of household and other simple categories;

Broad industry or occupation groups (with the possible exception of groups that are relatively small or highly concentrated in area, not easily limitable in advance).

32. The report of the third session of the Sub-Commission (E/CN.3/83) contained a chapter entitled "Sampling methods in relation to population censuses", which pointed out some of the advantages and limitations of the uses of sampling in connexion with censuses. Economy and speed may be gained by using sampling in various ways:

(a) Sampling is used to obtain advance estimates of some of the important census results for the whole country and large sub-divisions thereof;

(b) Sampling is used as an integral part of a complete census to obtain supplementary data;

(c) Sample surveys carried out independently at or about the same time as a complete census are used to assess the incompleteness of enumeration in the complete count, and other errors, for example, in age reports;

(d) Sample surveys are used instead of complete censuses where funds, physical facilities, and suitable personnel are inadequate for a complete count;

(e) Sample surveys are used in inter-censal years for the study of demographic changes;

(f) Samples of the completed returns (e.g., after punching) are used to obtain many cross-tabulations which would be very expensive and time-consuming on a 100 per cent basis.

(g) Sampling is used to control the errors of editing, coding, punching, and tabulating the returns of censuses and sample surveys.

/33. It is the

33. It is the first use named above which the Sub-Commission wishes to comment on and emphasize at this time. A great many countries will take censuses next year, and their plans will by now be pretty well settled. Indeed, some countries have already taken their censuses. Hence, at this time it would be useful to speak only of simple changes in the tabulation programmes that might be made in order to obtain early returns.

34. The full advantages of speed in early returns may be attained by taking a sample of returns from the complete count, and processing them in advance, before the remainder of the returns are edited and coded. Such a sample is, of course, speedier than the tabulation of a sample of the complete returns of a census, subsequent to editing, coding and possibly punching.

35. The problem of choosing a suitable sampling unit arises. To be processed ahead, and in order not to interfere with the regular processing of the census, the sampling units must be extractable physically from the remainder of the returns, so that they can be sent through the operations of processing (which would possibly be preceded by a second stage of sampling). They must therefore be the fundamental building blocks that are ordinarily processed as inseparable units. Usually they will be definable areas. When the early tabulations are planned in advance, the sample areas may be designed in advance with the aim of achieving uniformity in the number of inhabitants, or the number of acres, to increase the precision of the estimates of totals that are obtainable from a sample of a given size. (The precision of a proportion may not be seriously impaired by inequality in size.)

36. In many countries the building block for processing the census is the "enumeration district", ordinarily the area covered by a single enumerator (although two or more Enumeration Districts, if very small, may be assigned to one enumerator). The Enumerator District will, in some countries, form a suitable sampling unit. A sample of Enumerator Districts, drawn in a proper manner, is then to be put into the stream of processing, for advance tabulation. Two possible sampling plans may be considered:

Plan A:

A single-stage plan of sampling, whereby the whole of every Enumeration District in the sample is processed, i.e., every person or household or farm therein.

/Plan B:

Plan B:

A two-stage plan of sampling, whereby a sample of names or of households or farms is drawn from each selected Enumeration District, and is processed.

37. Plan A has the advantage of being simple and easy to carry out, as the sample can be drawn by one person. In centralized operations, it is practically fool-proof, as no further sampling is required. It has the added advantage in hardly disturbing or slowing the regular census operations, because, as no further sampling is required, no extra training of the workers is required. Moreover, practically no additional cost is involved, and the tabulation of the census on a 100 per cent basis is achieved in its regular time with no retardation.

38. Plan B is desirable when the sizes of the Enumeration Districts are far from uniform and it is impossible to stratify them effectively. A larger sample of Enumeration Districts will then be required. A second stage of sampling can be introduced to decrease the cost, and to recover some of the speed that would be lost were the Enumeration Districts processed completely (Plan A). The efficiency of this two-stage sampling will be increased if sampling is with probability proportional to size at the first stage and inversely proportional to size at the second stage.

39. The size of sample, and the procedure of drawing it, will depend on the precision that is desired in the total number of inhabitants (where no alternative estimate is available) and in the proportions. The precision in most of the important proportions will undoubtedly be much greater than that in the total population. A simple calculation on the size of sample may be helpful. Ordinarily, something is known in advance concerning the variability in the sizes of Enumeration Districts. If nothing is known, some exploratory work will often elicit sufficient information. If, for example, the coefficient of variation in the sizes of the sampling units were 70 per cent, a sample of 3,600 units would give a standard error of just over 1 per cent. Bigger coefficients of variation will require bigger samples, or a relaxation of the precision desired, or the introduction of a second stage of sampling (Plan B).

40. The calculation of the most economical sizes of the samples in Plan B is more difficult. Texts on the theory of sampling should be consulted therefor.

/41. In some

41. In some countries, Plan A may be preferred because of its convenience. If a count or other estimate of the total population can be obtained quickly, then this plan can supply proportions (e.g., the proportion of males in the age-class 25 to 34) with considerable accuracy. It can also supply totals with less accuracy.

42. Before publication, the standard error in the total number of inhabitants, and in any of the proportions as well, may be estimated quickly by methods that have been described in textbooks on sampling. As advised in previous reports of the Sub-Commission, estimates of the standard errors of the characteristics of chief interest should certainly be made and published along with the results.

43. The extent to which the sample may be sub-classified will depend on the standard errors of the frequencies in the cells and in the marginal totals that are involved.

44. Differences between the sample and the subsequent complete count may be expected from differences in the procedures of editing and coding. The sample flows through first, before uniform procedures for handling unforeseen problems of editing and coding have all been finalized, or indeed before they have all been recognized. Moreover, no visual or mechanical verification by small areas is possible. These differences for most population characteristics, and for some agricultural characteristics as well, may not be serious, but publication of the sample for any characteristic that is in doubt might well be withheld.

45. Annex I contains an example of an actual tabulation plan for a population census and embodies some detailed procedures for drawing the samples.

V. THE MEASUREMENT OF GOODS TRAFFIC CARRIED BY ROAD

46. Most countries collect and publish statistics on the movement of goods by rail, by inland waterway, by sea and by air; the statistics range from total figures like total tonnage loaded in the country to detailed statistics which analyze the total figures by commodity, by origin and destination and so forth. But the carriage of goods by road is performed by so many different types of carrier, including many small carriers, that it has in most countries so far not been systematically measured despite the fact that in many places road traffic is seriously competing with the railways.

47. The Economic Commission for Europe of the United Nations (ECE) has drawn attention to the importance of securing data on road traffic. The Inland Transport Committee of the ECE at its fourth session

> "Noting
>
> (a) the increased importance of road transport in Europe,
>
> (b) the almost complete lack of statistics relating to this form of transport particularly in respect of the volume of freight traffic, and
>
> (c) the fact that the absence of road statistics renders it increasingly difficult to obtain any knowledge either of the total volume of transport as a whole or of the manner or direction in which transport is developing;
>
> Invited Governments...to forward their views as to what may be considered the best methods of collecting the information necessary for the compilation of road traffic estimates, with particular reference to spot checks and samplings, giving as much information as possible on past experience in this field." (E/ECE/TRANS/166, resolution No. 64)

48. The Transport Division of the ECE has co-operated with the Statistical Office in bringing the question of the use of sampling to measure goods traffic by road before the Sub-Commission on Statistical Sampling and in preparing documentation for the Sub-Commission (E/CN.3/Sub.1/24). The Sub-Commission also had available to it the documents listed in paragraph 1 of annex II to this report and ECE document W/TRANS/WP6/12 and its annexes.

49. At the suggestion of the Working Party on Transport Statistics of the Inland Transport Committee of the ECE the problem thus put before the Sub-Commission, in order to reduce it to a more manageable size by treating only the most important aspect, was restricted to the study of obtaining the following total figures for goods traffic by road motor vehicles:

/(a) Vehicle-

(a) Vehicle-kilometres, loaded and empty separately;

(b) Capacity in ton-kilometres on the basis of kilometres actually run;

(c) Tonnage of goods loaded;

(d) Freight ton-kilometres.

50. On considering the matter the Sub-Commission decided:

(a) To treat in detail the restricted problem of obtaining the figures listed above in countries where the road network is dense, and to make a formal recommendation upon it;

(b) To make less detailed remarks on other aspects of the problem of measuring road traffic.

A detailed study of the restricted problem is appended to this report as Annex II; the Sub-Commission's recommendation appears in paragraph 51 below; the less detailed remarks appear in paragraphs 53 and 54 below.

51. The Sub-Commission on Statistical Sampling notes the desire of the Governments of many countries to obtain figures for road-motor-vehicle transport comparable to those which are available for railways, inland waterways and air transport. The Sub-Commission therefore requests the Secretary-General:

(i) To recommend to interested Governments that they undertake in 1951 experimental sample surveys to obtain, during some conveniently chosen period such as a month, for road motor-vehicle goods traffic the following total figures:

(a) Vehicle-kilometres, loaded and empty separately;

(b) Capacity in ton-kilometres on the basis of kilometres actually run;

(c) Tonnage of goods loaded;

(d) Freight ton-kilometres.

(ii) To draw to the attention of the Governments the technical suggestions for making such surveys which are contained in Annex II to this report and to ask governments to communicate to him their experience in conducting such experimental surveys for consideration by the Sub-Commission.

(iii) To provide, within the limits of his resources, advice on the details of sampling technique when asked to do so by a Government conducting an experimental survey of the sort recommended by the Sub-Commission.

/52. Annex II to this

52. Annex II to this report is based on an exploratory paper submitted by the Statistical Office which was revised and approved by the Sub-Commission. Annex I provides in some detail an outline of how the sample survey mentioned in the recommendation in paragraph 51 might be conducted, discusses the most economical way to select a sample, lists the information required to enable the quantities referred to in the recommendation to be computed, and describes how their final computation can be made. In Annex II emphasis is placed on the importance of enlisting the co-operation of carriers and carriers' associations. The representative of the International Chamber of Commerce expressed the view that the Chamber might through its Transport Commission be able to assist in assuring the co-operation of carriers' associations in countries undertaking surveys.

53. The Sub-Commission considered the broader aspects of road transport statistics going beyond the four fundamental figures it studied in detail. It noted that for simplicity the following matters had been omitted from detailed consideration:

(a) The measurement of passenger traffic by road;

(b) The measurement of the movement of goods in road vehicles drawn by animals;

(c) The determination, as is required for road planning, of the number and type of vehicles using certain stretches of road.

The Sub-Commission also noted the importance of analyzing by commodity and by origin and destination total figures on the movement of goods in order to be able thoroughly to study the relations between road and rail transport. The representative of the International Labour Organisation emphasized the importance of seeking information on the cost including the labour cost of carriage of goods by various forms of transport.

54. The Sub-Commission considered the uses of road traffic counts, with and without stopping of vehicles to interrogate drivers. It is of the opinion that in countries where the road network is not dense, stopping of vehicles and interrogating drivers at a number of highway points might be the easiest way to obtain the total figures discussed in paragraphs 49-52 above and in Annex II. The Sub-Commission emphasizes the importance of traffic counts in planning improvements of the road network.

VI. SAMPLE SURVEYS OF CURRENT INTEREST

55. The Sub-Commission had before it for examination the third report from the Statistical Office on Sample Surveys of Current Interest containing descriptions of surveys carried out in different countries and in different subject fields (E/CN.3/Sub.1/23).

56. The Sub-Commission once again would like to express its appreciation to the Governments and research organizations for their co-operation in supplying information on the various surveys. It is apparent from the available information that better sampling methods are being used in different countries and that increasing interest is being shown in the recommendations of the Sub-Commission in the methods of preparing sampling survey reports.

57. The Sub-Commission is pleased to note the wide interest shown in the reports of the sample surveys and is of the opinion that the collection of particulars on sample surveys should be continued. It is hoped that the reports on individual surveys collected by the Statistical Office from the different countries will, to an increasing extent, contain details on the lines recommended by the Sub-Commission and that sufficient particulars will be available in the near future for comparative studies to be made on various technical aspects such as sampling procedures, designs, costs or speed of operation. Such studies should serve a useful purpose in encouraging the adoption of better technical and organizational methods.

58. The Sub-Commission considered in some detail several of the technical aspects of the National Sample Survey (India) (E/CN.3/Sub.1/L.7) and expressed considerable interest both in the statistics to be made available and in the opportunity afforded by the Survey to gain more knowledge about many technical sampling problems connected with a survey of this magnitude, which is intended to be continued from year to year. The National Sample Survey is a country-wide multi-purpose survey developed in the first instance to obtain data on rural national income but ultimately to provide statistics for a number of administrative and planning purposes. There will be 1,800 sample villages distributed into about 600 strata by population or area. The number of sample villages will be selected in each stratum with probability proportional to the area of the village (or population), or at random when neither population or area figures are available. At the outset, about sixteen household schedules will be obtained from each village. Enumeration is

/expected to

expected to begin at the end of September 1950. It is intended to make four successive surveys in the first year.

59. The Sub-Commission appreciates the administrative difficulties encountered in the initial establishment of a country-wide survey and believes that after the initial stages of organization and operation are completed it would be possible to improve the efficiency of the sample design. This will be particularly desirable if it is intended ultimately to provide figures by states, or other areas smaller than national. At that time it will be necessary to increase the number of villages in the sample. Moreover, the results of the forthcoming census can probably be used to increase the efficiency of the sample design.

60. It is emphasized that the magnitude and continuing nature of the survey offers an excellent opportunity for experimental work on a large scale, and the Sub-Commission requests that full accounts of the work in 1950-1951 be made available to it at its next session.

VII. SAMPLING METHODS FOR ESTIMATING DISTRIBUTION BY SIZE OF INDIVIDUAL AND FAMILY INCOME

61. At its fourth session the Sub-Commission gave consideration to the use of sampling methods for estimating distribution by size of individual and family income. It had before it a report on the usefulness of these statistics for the study of various economic and social problems, including a description of conceptual and sampling problems arising in their compilation (E/CN.3/Sub.1/27).

62. From a social point of view, it is desirable to secure information on the proportion of the family or household units whose incomes fall into lower income brackets, in particular those whose incomes are below standards which are considered to be the minimum compatible with a decent standard of living. From an economic point of view, it is also desirable that the breakdown should show the above information. Such data can be used for certain purposes in conjunction with the results of family living studies, a subject on which the Sub-Commission reported at its second session (E/CN.3/52). Data giving such information can in most countries only be secured from sample surveys representing all family or household units.

63. Information on the distribution of income by size may also be secured from income-tax forms, though in many countries these do not give adequate information on the distribution of income in the lower brackets. The Sub-Commission dealt mainly with the use of sampling methods in the compilation of income tax statistics as the general problem in household surveys had been considered at its previous sessions.

64. In general, reliance on sampling methods in compiling income tax statistics has several advantages. In addition to considerably reducing the costs of compilation, it yields results more quickly and permits the individual returns to be treated with greater care and by more highly qualified personnel. The Sub-Commission, therefore, wishes to draw the attention of Governments now publishing income tax statistics to the usefulness of sampling procedures in this particular field. Moreover, since income tax data are often not very reliable for certain purposes owing to under-reporting and are subject to other limitations (for example, differences in the time-periods to which the individual returns refer, the treatment by the fiscal authorities of certain items such as capital gains and losses, depreciation allowances and allowances for dependents and life insurance), statistics obtained by sampling, instead of by complete enumeration, may be sufficient for most purposes.

/65. Because of

35. Because of the confidential nature of the income tax returns, and also for administrative reasons, fiscal authorities may not be prepared to make the basic material available for statistical purposes. Sometimes the original returns are accompanied by carbon copies containing the main figures. This is a better procedure than using different forms to be filled in by the taxpayers for statistical purposes. However, either method gives more than is necessary if sampling methods are to be used. If a sample is to be used, a practicable method may be to have the fiscal administration copy the significant data from the returns included in the sample on special forms which thereupon are passed to the statistical authorities.

36. The statistics should be compiled by random sampling, and if possible stratification by geographical areas should be used. Information with respect to the small proportion of the population in the high-income brackets usually weighs heavily in the means, aggregates and distribution of aggregates collected in the survey. To improve the precision of the results, a large proportion of the high-income groups may be included. Detailed instructions will be necessary to avoid that in taking the sample its basic principles are unintentionally violated. For example, in the case of persons who were late in sending in their returns, or whose returns have not yet been approved by the revenue inspectors, substitution by other returns should not be permitted, but replacement by the return of the previous year may be possible under appropriate circumstances. Where the returns contain information, for example, on the number of dependent children, or other important characteristics, such information should also be included in the inquiries.

37. Statistics of income tax may include joint returns of husband and wife but usually refer to the incomes of individuals. Therefore, for obtaining information on the distribution of family income other methods than those described above must be used. Here also the Sub-Commission wishes to draw attention to the usefulness of sampling procedures.

38. Very often surveys of family incomes will be combined with inquiries on other subjects, such as expenditure on goods and services and ownership of assets. The definitions of "family" and "income" to be used will depend on the purpose of the survey. In many surveys the "household" and not the "family" will be chosen as the unit. It is also possible to confine the spending unit to those family members who pool their incomes for all major expenses. If this is done, the family members

/who do not

who do not pool their incomes will be treated as separate units. Family membership may be determined as of the interview date, or each person who was part of the family during a portion of the year may be included. If the first method is used, only the income of persons being members of the family as of the interview date will be included. Persons who started earning an income during the year, newly-formed families, and other partial-year spending units may be included in the survey, or they may be excluded altogether.

69. In defining the income concept to be used, attention will have to be given to the treatment of such items as income in kind, farmers production for own consumption, net rental values of owner-occupied houses, contributions to social insurance and pension funds, family allowances of members of the armed forces, gifts, capital gains, and proceeds from change in ownership of assets. Understatements of income in sample surveys may be expected to occur, owing to inaccurate reporting. Another source of possible bias in the average incomes of families included in the sample may arise from absence or from refusal of the families to co-operate. Wherever possible, estimates of total income derived from sample surveys should be compared with independent estimates of total personal income, obtained as part of the evaluation of national income, after duly allowing for differences in the underlying income concepts. Such comparisons may be carried out separately for the various income components, such as total wages and salaries, interest, rent and dividends. The accuracy of the income data may, under appropriate circumstances, also be improved through other means, for example, by examining both income and expenditure of the families concerned, and also the changes in total assets between the beginning and the end of the year.

70. The usefulness of statistics of the distribution by size of family incomes for purposes of economic analysis would be greatly increased if they are classified by composition and size, occupation, and other family characteristics. If population census data on characteristics of families are available, the classifications selected for the sample should correspond with the census classifications. This permits the sample data to be brought into relationship with the census data.

VIII. STATISTICS ON INTERNATIONAL MIGRANTS AND TRAVELLERS

71. The Population Commission at its fifth session expressed the desire to have the Sub-Commission's views on the possibility of using sampling methods to obtain migration statistics (E/1711). While considering this question the Sub-Commission had before it the 1949 draft recommendations of the Population Commission for the improvement of migration statistics (E/1313, annex 3) and an analysis of the comments of Governments on these draft recommendations (E/CN.3/90 and Add.1). The Sub-Commission also considered two papers prepared by the Secretariat and the International Labour Office (E/CN.3/Sub.1/26 and E/CN.3/Sub.1/L.6), which contained, among other things, an account of experiments in sampling methods made in this field since 1947.

72. According to the draft recommendations of the Population Commission, statistics should be obtained on all arrivals and, if possible, on all departures of international travellers (with the exception of frontier traffic),* classified into the following major categories:

(a) Arrivals from other countries

(i) Visitors (for transit, holiday, education, business, and other purposes);

(ii) Residents (nationals and aliens) returning after less than one year abroad;

(iii) Temporary** immigrants: i.e., persons intending to exercise temporarily an occupation, and their dependants (but not refugees);

(iv) Permanent** immigrants (excluding refugees or transferred populations);

(v) Refugees (at times and places where applicable);

(vi) Transferred populations (at times and places where applicable).

(b) Departures to other countries

(i) Visitors departing on completion of visit;

(ii) Residents (nationals and aliens) departing for less than one year;

(iii) Temporary** immigrants departing;

---

* Frontier traffic is the movement of persons residing in frontier zones, moving frequently across the border and often authorized to use simplified travel documents (frontier cards).

** Removals for one year or more are to be regarded as permanent; those for less than one year as temporary.

/(iv) Emigrants,

(iv) Emigrants, i.e. residents (nationals and aliens) departing for one year or more:

(a) Emigrants, excluding persons deported or transferred populations;

(b) Persons deported (at times and places where applicable);

(v) Refugees, departing from countries in which they have been residing temporarily (at times and places where applicable);

(vi) Transferred populations (at times and places where applicable).

For categories (i) and (ii) of both arrivals and departures, only simple counts of total numbers are required; for categories (iii) to (vi) of both arrivals and departures, the draft recommendations call for classifications by certain characteristics. In what follows, the term "travellers" refers to arrivals and departures in any of the major categories while the term "migrants" refers to categories (iii) to (vi).

73. Since sampling has not yet been used extensively in this field, experimentation is desirable to determine what methods are best suited to various conditions. It is requested that the Secretary-General report to the Sub-Commission at a later session on the results of any such experiments which may be carried out in any country.

74. The following remarks refer primarily to the use of sampling methods in the collection of statistics at the time when the travellers cross the frontier. On that occasion the usual frontier formalities could generally be so organized as to permit the identification of travellers in the major categories of arrivals and departures through simple verbal questions; thus it would be possible to obtain at least the totals of the major categories by complete enumeration. A statistical form containing appropriate questions could then be handed to travellers recognized as migrants; in order that no bias be introduced by non-response, administrative methods can be devised to ensure the return of all statistical forms by migrants. In many circumstances, however, it is not possible to identify the major categories without considerable extra expense, inconvenience to the travellers and additional work for the personnel engaged in frontier control. In that case, it may be possible to devise sampling methods which, without excessive expense or inconvenience, will yield the totals of the major categories and the desired data on characteristics of migrants. Experimentation with such methods is particularly desirable in countries where the existing methods provide statistics on only certain types of migrants, such as aliens,

nationals, or persons subject to military obligations, or where migration statistics are obtained only from population registers. By the application of sampling methods at the occasion of frontier control, it may be possible for such countries to obtain comprehensive, internationally comparable statistics at a reasonable cost and without causing excessive inconvenience to travellers.

75. Even in those countries where the total numbers of arrivals and departures in the various major categories are obtained by complete enumeration, it may be desirable to use sampling methods in collecting the desired statistics on characteristics of migrants. Sampling for characteristics might be advantageously used on occasions when large numbers of migrants arrive or depart through the same frontier post on the same day (e.g., arrival of a ship carrying many immigrants). In many such countries, however, this application of sampling may not be advantageous under ordinary conditions; since the migrants ordinarily represent only a small proportion of all travellers, their number may be small enough, under most conditions, to permit complete enumeration of their characteristics with little expense or difficulty.

76. Three kinds of sampling plans can be considered for obtaining data on the total numbers of arrivals and departures in various major categories and of the characteristics of migrants:

(a) Collection of data from one traveller (or migrant) selected at random out of every n travellers (or migrants);

(b) Collection of data from all travellers (or migrants) at suitable intervals of time, for example, on every eighth day or thirteenth day, etc., with equivalent rates of selection in the case of non-daily transport services;

(c) Collection of data from all travellers (or migrants) arriving or departing in certain ships, airplanes, railway trains, buses, motorcars, and other conveyances, selected so as to constitute a properly designed sample of conveyances.

These methods can be combined; for example, samples might be taken on certain days or in certain conveyances, comprising a certain fraction of persons travelling on those days or in those conveyances.

77. If the method used is to collect data from every nth traveller (or migrant) there should be an easy, automatic way of selecting the travellers at the specified intervals. One possible method which could be tested is to give each

traveller, at an early stage of the journey across the frontier, a card taken from a package in which a distinctive mark would have been placed, in advance, on every nth card. Thus the travellers bearing the distinctively marked cards could subsequently be required, at the frontier post, to give the information needed for statistical purposes. A variation of this method, which might be effective under some conditions, is to make the identifying card itself bear the questions for statistical purposes, which would not appear on the cards given to travellers not selected for the sample. The latter variation might be especially effective where some card (e.g., a landing card) must be filled out in any case by all travellers; then the cards given to those selected for the sample could merely bear more questions than the others. By any of these methods it might be possible, under certain conditions, to expedite operations at the frontier post by forming one or more special queues of the travellers selected for the sample, so that they could answer the special questions, fill out the statistical forms, or have their filled-out forms inspected, while the other travellers go through the usual formalities.

78. The methods of collecting data from all travellers (or migrants) at selected intervals of time or in selected conveyances have the disadvantage that, for a given number of travellers or migrants, they produce estimates with increased sampling variance resulting from variations between days or conveyances. On the other hand, they may, under some circumstances, involve fewer administrative difficulties. Experimental surveys may give some indication of the loss of efficiency resulting from the use of these methods.

79. The Sub-Commission stresses the importance of applying, in any experiments with sampling for the collection of statistics on migrants and travellers, the general principles that are valid in any sampling work, in particular the following:

(a) The selection of elements in the sample should be effectively random, since the selection of elements that are considered a priori as "representative" introduces unknowable biases.

(b) Schemes providing for voluntary response are hazardous: if many persons fail to respond the results may be seriously biased.

(c) If the travellers or migrants included in the sample are asked to fill out special statistical forms, they should be given ample time to do so, and the forms should be collected before they leave the frontier post.

/(d) If, for

(d) If, for any reason, there are considerable numbers of persons who do not respond, the resulting biases should be investigated from time to time, for example by obtaining data from small numbers of travellers at various frontier posts in such a way that there can be no non-response.

80. The Sub-Commission calls attention to the advantages of collaboration between the authorities of neighbouring countries in the operation of any sampling experiments which may be undertaken in this field, and notes the suggestion concerning regional arrangements made by the Statistical Commission (E/1696/Rev.1) and by the Population Commission (E/1711). Such collaboration would be helpful in improving the comparability of results obtained in the countries concerned and might make it possible to achieve substantial reductions in cost. Regional arrangements may prove easier to achieve if they bear on the application of sampling schemes, which are by nature purely statistical, than if they were to embrace the operations of administrative frontier control applied to all travellers. It may in particular prove advantageous to have statistical forms distributed to travellers by the authorities of the country of departure and collected by the authorities of the country of arrival, due care being taken to formulate the question so as to provide the statistics desired by both countries.

81. Where data on the total numbers in various categories of arrivals and departures, or on the characteristics of migrants, are obtained by complete enumeration, it may nevertheless be useful to apply sampling methods in tabulating the results. The use of sampling in tabulations on these subjects could follow fairly well standardized methods and would offer several special advantages:

(a) It may prove easier for the agency responsible for the statistical tabulations to get access to a sample of the documents used by the agencies controlling the arrivals and departures of international travellers, or a sample of the population registers or other records, than to survey the entire materials. Care should be taken that such sample is selected so as to be unbiased.

(b) Sampling makes it possible to obtain tabulations of which the cost would be prohibitive if they were carried out on the basis of complete enumeration. This should facilitate the improvement of the international comparability of the statistics in question, which are regarded as being gravely defective in this respect as well as from the point of view of coverage (E/805).

IX. DEMOGRAPHIC STUDY IN INDIA

82. The plans for a study of the inter-relationship of demographic, economic and social factors in India were presented to the Sub-Commission for its consideration (E/CN.3/Sub.1/L.3). The study will be undertaken jointly by the United Nations and the Government of India and was considered by the Population Commission at its fifth session and by the Economic and Social Council at its eleventh session. The Sub-Commission noted the recommendation of the Population Commission that "this pilot survey be considered as only the first step in a continuing programme of field surveys designed to obtain data on various aspects of the inter-relationship of demographic, economic and social factors in selected areas of India" (E/1711). Because of the lack of experience in this type of survey and because valid conclusions can be built up only by means of long-term observations, the Sub-Commission completely endorses this view of the Population Commission.

83. The Sub-Commission noted that in India, the routine registration system had failed to cover a large fraction (of the order of 40 per cent or more) of births and deaths, and that attempts at obtaining data on births and deaths by household surveys had resulted in omissions of a similar order of magnitude. The Sub-Commission is strongly of the opinion that a considerably higher level of ascertainment in the household survey will be necessary in order to obtain reliable and trustworthy results in a study on the varying aspects of economic and social conditions on birth and death rates. To this end, the Sub-Commission recommends the undertaking of preliminary studies to compare the efficiency of different survey techniques. The Sub-Commission feels that such studies will also be helpful to the local administration in establishing suitable methods for recording births and deaths.

84. The plans for the study of inter-relationships indicate that three or four areas of contrasting economic development may be selected within a geographical region. The Sub-Commission suggests that within the limits of budgetary resources the groups of contrasting areas should be selected from a few strongly contrasting geographical regions. The results of such replicated studies will be very valuable in revealing the behaviour of populations with different social and ethnic backgrounds. It has been noted that a National Sample Survey is being

organized on a continuing basis under the auspices of the Government of India and the costs of replications in the study of inter-relationships may be reduced substantially if some co-ordination can be worked out between the two surveys.

85. The Sub-Commission feels that the studies will offer methodological guidance in certain types of statistical sampling problems and requests that a report on their progress be submitted to the fifth session of the Sub-Commission.

X. FUTURE PROGRAMME OF WORK

86. At the request of the Food and Agriculture Organization, the Sub-Commission at two of its past sessions dealt with the problem of agricultural statistics and made certain recommendations for improving the methods of data collection. It is noted that the FAO is now making a study of the methods of agricultural data collection in different countries, particularly in under-developed areas (E/CN.3/Sub.1/29). The Sub-Commission wishes to examine the results of this study at its next session and also the results of the censuses of agriculture, in so far as available, to give further consideration to improvement of statistics on acreage, yield and forecasts.

87. The Sub-Commission suggests the following new items for consideration at its next session:

(a) Statistical techniques for the development of methods for testing standards for international trade commodities.

(b) The use of survey and experimental techniques in the development and improvement of agriculture and small-scale industry.

(c) Nutrition studies.

88. The Sub-Commission suggests also a few other items to be taken up at a future session but not necessarily at its next session. These are:

(a) Theory of extreme values and its technical applications. The Sub-Commission expressed general interest in this subject but felt that it was in no position at present to evaluate the theory or to consider its application. It was decided that further discussion should be deferred until a fuller exposition was possible.

(b) Sampling methods for collecting data on distribution in the field of trade and commerce.

(c) Methods for sampling forms already completed for administrative or similar purposes.

89. In connexion with some of the problems discussed in the preceding chapters, the Sub-Commission indicated the desirability of experimental surveys for developing suitable methods of data collection. Such suggestions have been made in regard to measuring of goods traffic by road; study on inter-relationships of demographic, economic and social factors; and collection of statistics of

international migrants and travellers. The Sub-Commission requests that the results of these experiments should be made available to enable it to give further consideration to these problems at its next session.

90. The Sub-Commission would also like to examine the results of application of sampling methods in studies such as, national income and income distribution, family living inquiries and other subjects in which it has expressed its continuing interest. It would also like to examine the results of the application of sampling methods in the recent population censuses if particulars are available.

91. The participation of observers in its discussions has been an important feature in all sessions and the Sub-Commission wishes such participation to continue. It has been observed that the experience and outlook of experts who can arrange to attend meetings in New York and those who can do so in Geneva varies, and the Sub-Commission finds this variety of experience very useful. Since the fourth session was held in New York, the Sub-Commission requests that the fifth session be held elsewhere.

92. The Sub Commission noted the fact that the 27th biennial session of the International Statistical Institute will be held in India in December 1951. This session will be attended by a large number of the outstanding statisticians from a great many countries. It is the Sub-Commission's view that it would be very advantageous if the fifth session of the Sub-Commission could be held in India either just before or just after the 27th session of the International Statistical Institute. This would permit the direct participation of a number of additional experts and regional participants, which would add both to the technical accomplishments of the fifth session and to the speed with which Sub-Commission recommendations could be implemented in the different countries and particularly in the under-developed areas. The Sub-Commission therefore urges strongly that every effort be made to set the date and place of the fifth session to coincide with the 27th session of the Institute. In making this recommendation, the Sub-Commission took account of the fact that additional financial implications might be involved.

93. In the event that it proves not possible to hold the session in India in conjunction with the Institute, the Sub-Commission recommends that the fifth session be held in Geneva. It also recommended that the session, if not held in December in India, be held in June or July 1951 instead of in September because of commitments which members have for other scientific meetings in September.

/ANNEX I

ANNEX I

AN EXAMPLE OF A TABULATION PLAN

There is appended here an actual sampling plan for a census of population. It is a single stage plan, and it possesses a feature not mentioned before, namely, a 10 per cent sample as well as a 1 per cent sample. The 10 per cent sample is intended to follow the 1 per cent sample, and to give much information that is impossible to obtain with the 1 per cent sample. The 10 per cent sample should nevertheless be ready months ahead of the full 100 per cent tabulations.

The samples

Sample 1: A 1 per cent sample of E.D's (an E.D. is an enumeration district). Each E.D. to be processed completely (Plan A).

Samples 2-10: Each of these samples is an independent 1 per cent sample of E.D's.

Procedure for drawing the samples

1. List the E.D's. in any convenient order.
2. Divide them into groups of 100, in the order listed. Number the E.D's. within each group from 01 to 00.
3. By the use of random numbers draw 10 E.D's. at random from each group of 100.
4. The first E.D. drawn from each group will be a number of a 1 per cent sample of E.D's. This is Sample No. 1, and it will be called "the 1 per cent sample".
5. Sample No. 1 will be tabulated for the purposes of the 1 per cent sample with all possible speed, and the results will be published, after consideration of the standard errors of the small or doubtful frequencies. Some of the results should be ready within a few months following the date of the census.
6. There will be nine more 1 per cent samples. The second E.D. drawn from each group will be a member of Sample No. 2. The third E.D. drawn from each group will be a member of Sample No. 3, etc.
7. Sample No. 1 will then be tabulated to make the tables that are planned for the 10 per cent sample.
8. Samples 2-10 will be processed on the heels of Sample 1, and each will be tabulated separately (a) for the tables that are planned for the 10 per cent sample, and also (b) for the tables that will already have been published for Sample 1.
9. The average in any cell will be the result of the 10 per cent sample for that cell.

/10. Each cell

10. Each cell will have ten independent estimates from the ten sets of tables and these estimates will provide nine degrees of freedom for the estimate of error. These estimates of error will provide a basis for decisions as to what tables are sufficiently accurate for publication.* They will also provide a basis for:

    (a) Publishing a statement in regard to the precision of the 1 per cent sample (Sample No. 1, already published);

    (b) Publishing a statement in regard to the precision of the 10 per cent sample.

11. For deciding what tables to publish it will not be necessary in most cells, perhaps 95 per cent of them, to calculate the standard error or probability limits. It will be obvious, in most cases, that the cell is strong enough for publication, or too weak. A calculation can be made for any cell for which an actual estimate of precision is desired. The range may be used for this estimate, if desired, with the advantage of simplicity, as the calculation is very simple.

12. Naturally, the decision on which to publish a certain table, and on how much detail to publish, will not depend on any hard and fixed rules of precision. The possible uses and misuses of the data should be kept in mind. Often the user of a table will be interested in the approximate ratio of one cell to another, and not on the magnitude of any one cell.

13. The above plans will not delay the 100 per cent tabulations. Actually, the 10 per cent sample, being a sample of whole E.D's. will be 10 per cent of the 100 per cent job. The remaining 90 per cent will be completed and the results published, in the usual time. However, the tabulation programme for the 100 per cent job will be simplified because, for many if not most purposes, the 10 per cent sample will suffice.

14. If a two-stage plan is required (Plan B), the estimation of the standard errors is unaffected: the standard errors are estimated by the examination of the variability of the ten samples, just as in the single-stage plan.

---

* Instead of calculating an estimate of the standard error of the frequency in any cell, as estimated from the ten samples, some statisticians may prefer to make use of the fact that the probability is only 1/1024 (obtained as $(1/2)^{10}$ that the lowest (or highest) of the ten estimates of any cell will be found to lie above (or below) the median of the 100 per cent count.

ANNEX II

SAMPLING TO MEASURE THE GOODS TRAFFIC CARRIED BY ROAD MOTOR VEHICLES

1. References

(a) Economic Commission for Europe: Report of the Working Party on Transport Statistics on its third session (E/ECE/TRANS/224) referred to below as WP.

(b) Economic Commission for Europe: Report of the meeting of statisticians held at The Hague on 24 to 29 July 1950 (TRANS/WP/4) referred to below as S.

(c) United Nations Statistical Office: "International Standard Definitions for Transport Statistics" (Statistical Paper M 8) referred to below as M 8. Transport statistical terms are used where possible in the sense of these definitions.

(d) United Nations Statistical Office: "The Preparation of Sampling Survey Reports" (Statistical Paper C 1, revised) referred to below as C 1 rev. Sampling theory terms are used in the sense of this paper.

2. The object of this study

As explained in paragraph 51 of the report to which this study is annexed, it is desired to measure approximately the following quantities for goods traffic carried by road motor-vehicles:

(a) Vehicle-kilometres, loaded and empty separately (WP, paragraph 24);

(b) Capacity in ton-kilometres on the basis of kilometres actually run (WP, 24);

(c) Tons loaded (M 8, 23a; WP, 24);

(d) Freight ton-kilometres (M 8, 23e; WP, 24).

This study outlines a method of measuring these quantities by sampling and will take the view that tons loaded in commercial traffic (M 8, 17) and freight ton-kilometres in commercial traffic are the fundamental figures required for economic purposes. It suggests that vehicle-kilometres and capacity ton-kilometres can be obtained as a by-product of the process used to obtain figures for tons loaded and ton-kilometres.

3. The "frame" on which a sample is based

The frame (C 1 rev., 2a) is the set of documents on which the choice and utilization of a sampling method is based. If possible, the frame should be based

/on documentary

on documentary material already available in the country in question. This material may refer to:

(a) The carrier (permits to carry, permits to buy petrol, reports to supervisory authorities, insurance documents etc.);

(b) The vehicle (census, registration, insurance documents etc.);

(c) The goods (freight documents);

(d) The locale where the carriage takes place (road maps, etc.).

The nature of the available documentation must determine the sampling procedure including the choice of elementary unit, sample-unit, sample etc.
(C 1 rev., 2b, c, d).

4. The choice of a sampling method

In this paragraph the choice of a sample based on each of the four types of material listed at paragraph 3 above will briefly be discussed. It is not intended that the discussion shall exhaust all possible uses of the material which may be available.

(a) Sample of vehicles. Because of national registration laws, a frame based on individual vehicles is probably easiest to construct and a stratified (C 1 rev., 2h) sample then easiest to set up. The operator of each individual vehicle, v, selected would be asked to report on the performance (loads handled, kilometres run etc.) of v in the time period P (say a month) covered by the survey. But in most cases the operator of v will not be able to report on the performance of v during P unless he knows at the beginning of P that he will be asked to do so. Such knowledge would, of course, introduce bias into the sample in a way which, while probably not serious in the case of operators of single vehicles, might be serious where the vehicle in question was one of a fleet operated as a unit. For instance, if an operator owning three identical vehicles was told at the beginning of P that he would be required to report on a specific one of them, and if, for a part of P he had only work enough for two vehicles, it would embarrass him to know to which of his three vehicles to assign the work. The objection just described is here considered sufficiently grave to make it undesirable to base the survey on a straightforward sample of vehicles. It is, however, recommended in (d) below that in certain cases a sample of carriers be based on a sample of vehicles.

/(b) Highway

(b) Highway traffic counts. The frame for a highway traffic count can probably be constructed in any country but in countries where the road network is at all dense a traffic count is so cumbersome and expensive that it should not be undertaken unless the paucity of documentary information of other sorts makes a traffic count necessary. It is here assumed that this case will not arise in any European country. The situation in countries where traffic is not dense is briefly discussed in paragraph 94 of the report to which this study is an annex.

(c) Sample of freight documents. Freight documents may be in the hands either of shippers, carriers or consignees. The multiplicity of shippers and consignees is so great that it is probably impossible to obtain lists of either of these from which samples can be drawn. So if freight documents are to be used they must be obtained from carriers. It is probable that operators of large fleets use freight documents from which can be obtained most of the information required about goods handled, and advantage will be taken of this fact in paragraph 8 (e) below. But small operators, particularly those who operate single vehicles, are unlikely to make systematic use of freight documents and it therefore appears to be undesirable to attempt to sample freight documents directly (compare S, 28).

(d) Sample of carriers. A sample of carriers, including carriers for hire (M 8, 15a) and private carriers (M 8, 15b) seems to offer the fewest objections. Each carrier sampled would, as described in paragraph 8 below, be asked to report the total tonnage of goods he has loaded in a period, P, of about a month inside the country in question and to supply some additional data. It is recognized that carriers, particularly large carriers, are likely to resist providing information from which their total revenues could be estimated. But because of varying rates for different commodities the figure for total loadings would not, in general, make it possible to approximate revenue. Where serious objection is nonetheless feared, the suggestion made in paragraph 8 (e) below may be of some use. Paragraphs 5-11 below discuss in some detail the sampling of carriers.

5. Listing and sampling carriers

(a) It is unlikely that in any country a complete list of carriers is available which includes all carriers for hire and private carriers no matter how small their operations. But partial lists may be accessible in

the hands of regulatory authorities, carriers' associations, insurance companies etc. As many as possible of these should be assembled and the co-operation of the authorities, associations and companies should be enlisted (S, 29e). If a sufficiently complete coverage of carriers results from these lists, the carriers can forthwith be numbered serially and sampled by a random method. If information is available on the nature of each carrier's operations (S, 18) and the number of his vehicles, the sample can be stratified according to this information.

(b) If, as will usually be the case (S, 28), the lists discussed at (a) omit a significant number of carriers, the registers of the numbers (on license number plates) used to identify the lorries can be used to construct an unstratified sample of carriers (S, 27). The sampling procedure might be as follows:

(i) From the registers mentioned above the requisite number of lorries could be drawn, perhaps simply by picking out licence numbers at random.

(ii) The sample of carriers would then be the set of those carriers who operate the vehicles selected at (i).

(iii) In computing the final figures allowance must be made arithmetically for the fact that the probability of selecting a carrier operating several vehicles is greater than the probability of selecting a carrier operating a single vehicle and that some carriers may be "selected" more than once. If every gth vehicle is selected, and n_p vehicles, belonging to n'_p carriers, are selected for carriers having exactly p vehicles the raising or weighting factor for the totals (loadings, ton-kilometres, etc.) returned by the carriers having exactly p vehicles will be

$$\frac{g\ n_p}{p\ n'_p}$$

Thus if T_{pi} is the tonnage loaded (or ton-kilometres run etc.) by the ith carrier who owns exactly p vehicles and $\bar{t}_{pi}$ is the corresponding means per vehicle, the total tonnage loaded (etc.) may be estimated by

$$T = \sum_p \left(\frac{g\ n_p}{p\ n'_p} \sum_i T_{pi} \right)$$

/and the

and the variance of T may be computed from the formula

$$V(T) = s^2 c^2 \sum \frac{(n_p)^2}{n'_p}$$

where s^2 is the variance of the $\bar{x}_{pi}$ taken over all vehicles of all carriers at once.

(iv) In deciding on the size of the sample it should be remembered that a small sample, chosen truly at random, a large proportion of whose members co-operate in the survey, produces more accurate results than a larger sample which is biased either in its initial selection or by a large amount of non-response from the carriers sampled. Methods of avoiding non-response are discussed at 8 (f) below.

6. Relative accuracy of the various methods

If the carriers can be stratified by nature and number of vehicles operated (5 (a)), a sample of carriers, stratified according to this information with a variable sampling fraction (C 1, rev. 2), approximately proportional to the size of operations, will provide the most accurate results. On the other hand, if the information on carriers is not adequate to effect this stratification a sample of carriers based on vehicles (5 (b)) may be expected to give more accurate results than would a random sample of carriers, since the procedure of 5 (b) will sample carriers with probability approximately proportional to the number of vehicles operated. In cases in which the required information is available only for certain categories of carriers, a stratified sample of these categories can be taken, the remaining carriers being sampled by means of vehicle registrations. Only carriers not occuring in the categories on which the stratified sample is based will be retained in the latter sample.

7. Analysis of the final figures by class of carrier

The final figures can of course be analysed, for instance by class of carrier (S, 18), even if they were not obtained from a stratified sample. To maintain a given accuracy in the figures for different categories the size of sample must increase with the number of categories; these must therefore be known before the sampling fraction is fixed. In particular it can be arranged that the survey shall yield to a desired degree of accuracy, the average performance per vehicle (over-all or by class of carrier) thus relating the

/result of

result of the survey to data on the number of vehicles in use, a type of data which is normally kept up to date in most countries (S, 29f).

8. The information to be collected from the carriers sampled

(a) If P is the time period for which data are to be collected, the carriers to be sampled should be notified before the beginning of P of the questions they will, at the end of P, be asked to answer. This appears to be a necessary step if adequate replies are to be expected even though it may introduce bias into the sample. Since a report on total operations will be asked for, the bias introduced will be negligible compared to that referred to at 4 (a) above.

(b) The principal datum each carrier would be asked to give is the total gross weight (i.e. weight of goods plus their packing (M 8, 22b) of the commercial goods loaded into his vehicles inside the country in question in period P.

(c) For a relatively short sub-period, Q, of P, say a day or two if P is a month, each carrier sampled should be asked to keep a "complete record" showing the following facts for each vehicle which he operates in the country during Q. The complete record should include:

(i) The license number of the vehicle;

(ii) The distance it has run in the country in period Q and the distance it has run empty;

(iii) For each consignment loaded within the country during Q, its gross weight, place of loading, the place at which it is either to be unloaded or to cross the frontier for unloading abroad;

(iv) For each consignment entering the country loaded (M 8, 23c) in the vehicle during Q, its gross weight, point of entry, the point at which it is to be unloaded or, if it is in transit, the point where it is to leave the country.

As operators may be expected to be reluctant to show overloads, some bias may be expected in the complete records so obtained. If, under (i) above, the rated capacity of the vehicle is asked for the bias will be increased, so rated capacity should not be asked for if it can be obtained from other sources (see 6 (5) below).

/(d) The periods

(d) The periods Q should not run concurrently for all carriers sampled but should vary from carrier to carrier so that the whole of P is adequately covered.

(e) If some carriers in the sample regularly make use of freight documents, arrangement can be made through a carriers' association or by an interview (see (f) (i) and (ii) below) to permit them to submit a sample of these documents instead of the reports described at (a), (b) and (c) of this paragraph, thus saving them clerical work and avoiding the necessity of any one of them divulging the extent of his operations. The carrier would have to number serially the documents he used in period P and submit the ones with specific serial numbers selected at random.

(f) Some suggestions follow for minimizing the number of carriers who do not co-operate with the survey.

(i) It should be determined from the lists mentioned at 5 (a) above which of the carriers in the sample belong to associations, and as much help as possible should be obtained from the associations in getting the information described above from those of its members who are in the sample. It would be helpful if the association itself would provide information on the number, license numbers and capacities of the lorries operated by the carriers.

(ii) It is desirable that carriers in the sample who do not belong to associations be interviewed with the object of enlisting their co-operation and obtaining the information mentioned in the last sentence of (i) above (S, 29a). The extent to which this can be done depends, of course, on the number of carriers in question and the funds available.

(iii) The number of carriers who do not co-operate can be substantially reduced by the methods described at (i) and (ii) but there are in any case likely to be some carriers who do not co-operate. Though some adjustments can be introduced into the final figures to allow for non-co-operation, such adjustments cannot be relied on to remove all bias arising from this cause.

9. The data obtained from the survey

The process described at paragraph 8 above will make possible the estimation of the quantities listed below.

(a) The total gross weight, L, of commercial goods loaded inside the country by domestic road motor vehicles.

(b) If "local traffic" is defined as the traffic hauled less than x kilometres, x being determined in each country to include in local traffic short-distance traffic not competing with other means of inland transport, the complete records make possible the estimation of the ratio, f, of local to total traffic.

(c) From the complete records the average length of haul, h, of a ton of goods moving in domestic traffic can be estimated.

$h = \frac{\sum d_i x_i}{\sum x_i}$ where x_i is the weight of the ith consignment of goods and d_i is the distance it was carried.

(d) From the complete records the average length of haul, h', of a ton of goods moving in international traffic in domestic vehicles can be estimated.

(e) From the complete records an analysis can be made showing the quantity of goods hauled for each of a representative range of distances (S, 19).

(f) From the complete records vehicle-kilometres and the degree of loading can be estimated.

(g) From the complete records and the data on which the sample is based capacity ton-kilometres can be estimated.

10. International traffic

If special arrangements can be made for a check of all vehicles at frontier points during the period P (see WP, 4-5), the following quantities can be accurately determined.

I_1 the gross weight of commercial goods imported in domestic vehicles;

I_2 the gross weight of commercial goods imported in foreign vehicles;

E_1 the gross weight of commercial goods exported in domestic vehicles;

E_2 the gross weight of commercial goods exported in foreign vehicles;

T_1 the gross weight of commercial goods moving in transit in domestic vehicles;

T_2 the gross weight of commercial goods moving in transit in foreign vehicles.

/Where

Where there is an appreciable amount of local traffic crossing the frontier it should, if possible, be excluded from the figures.

11. The computation of commercial traffic

The quantities determined in paragraphs 9 and 10 can be combined as follows.

(1) Weight of goods loaded in long distance traffic:

$(1 - f) L + E_2$.

(2) Weight of goods unloaded in long distance traffic:

$(1 - f) L - E_1 + I_1 + I_2$.

(3) Weight of goods carried in long distance traffic:

$(1 - f) L + E_2 + I_1 + I_2 + T_1 + T_2$.

(4) Freight net ton-kilometres performed in long distance traffic:

$h ((1 - f) L - E_1) + h' (E_1 + E_2 + I_1 + I_2 + T_1 + T_2)$.

These formulas are subject to the following inherent errors in addition to the sampling errors.

(a) Duplications. Goods transloaded from one long-distance lorry to another will, in formulas (1), (2) and (3), be counted each time they are transloaded. (Transloading from vehicles engaged in short-haul collection or distribution to long-distance vehicles does not give rise to duplication because of the use of the factor 1 - f.)

(b) Omissions. Consignments carried in the domestic traffic of the country (i.e., both loaded and unloaded in the country) by vehicles of foreign registry will not appear in the figures.

Neither of the errors just listed is likely to be appreciable.
