

ON DESIGNING AND CONDUCTING MULTI-SUBJECT HOUSEHOLD ENQUIRIES WITH REFERENCE TO A PERMANENT SURVEY ORGANIZATION*

By M. N. MURTHY

Indian Statistical Institute

SUMMARY. In this paper the advantages of and problems involved in designing and carrying out integrated multi-subject surveys in the developing countries are discussed with special reference to household enquiries. It is pointed out that integrated multi-subject surveys are generally found to be more economical and operationally convenient than a series of uni-subject surveys. Further, integrated surveys enable us to study more intensively the inter-relationships between various social and economic characteristics through tabulation of data by deeper classificatory breakdowns. The advantages of having a permanent survey organisation for planning and carrying out social, economic and agricultural surveys involving difficulties in concepts, design, etc. are stressed. Suggestions for having suitable designs in case of multi-subject surveys are given. Some aspects of data collection and processing of data also have been mentioned.

1. INTRODUCTION

In recent years the need for and the problems involved in designing and conducting multi-subject sample surveys have received considerable attention, especially with reference to carrying out household enquiries in the developing countries to meet the ever-increasing demand for statistical information in the context of their efforts to achieve rapid social and economic development (United Nations, 1964a). This matter is of much importance to the developing countries as many of them either do not have any established statistical survey system or are still at the initial stages of building up a suitable system. Some of the countries are, however, already conducting multi-subject surveys with varying degrees of integration of the subjects of enquiry.

In this paper, it is proposed to discuss the implications of adopting a multi-subject survey system with special reference to household surveys, such as the integration of subjects of enquiry, use of sampling frames, choice of survey design, etc. As a prelude to this discussion, the general data requirements of developing countries are briefly described together with the problems involved in their detailed specification. It may be mentioned that the discussion is a rather general one and that the approach and the points made may need suitable modifications to accord with the nature and extent of statistical developments in the different countries.

2. DATA REQUIREMENTS

2.1. *Need for data.* In the developing countries, there is a growing need for statistical information on a variety of subjects, as such information is essential to planning for economic and social development and to assessing the progress achieved as a result of the general policies of and specific efforts made by the governments. This is mainly due to the fact that the concept of the role of the state in the countries of

*Based on a paper prepared by the author for ECAFE in connection with the Seminar on Sampling Methods held at Tokyo in 1965.

this region is slowly changing; the state is no longer interested just in the maintenance of law and order, but has begun to participate increasingly in the improvement of the social and economic conditions of the people. In fact, some countries have already resorted to planned development, which is necessary for the rational use of the resources especially since, in practice, the countries' resources are limited and their needs are considerable. To execute its various responsibilities, a country requires statistical information regarding its available resources, such as manpower, cultivable land, forests, minerals and oil, and the needs to be satisfied—an indication of which is given by the consumption pattern, rate of population growth, educational level etc., of the people.

2.2. *Types of data.* The data required by a country for the proper carrying out of its welfare activities may be classified into three types : (i) information on physical features, climatic conditions, soil type, etc., which do not change much over time; (ii) information more related to time, such as data on population, births, deaths, consumer expenditure, crop statistics, etc.; and (iii) a knowledge of the relationships between characteristics and of trends over time for purposes of projections and forecasts based on (i) or/and (ii). In this paper, attention will be mainly confined to data relating to type (ii), with special reference to data amenable for collection through households.

The data on various types of social and economic characteristics are to be obtained for the different sectors of the economy, sections of the people and geographical regions. The total economic complex of a country may be broadly divided into two parts—(i) the organized sector consisting of organizations and establishments recognized, registered or/and controlled by the Government and (ii) the unorganized sector consisting of the residual part. The latter can be further sub-divided into two parts—(a) household and (b) non-household. In this paper, only (ii) is considered with special reference to category (ii)(a).

The first task in specifying data requirements consists in visualizing in a comprehensive manner the nature of the statistical information required to satisfy the current and the future needs. An indication of items on which statistical information would be of considerable interest is given by the following :

- (i) population, births and deaths, migration, employment and unemployment;
- (ii) income and expenditure, consumer preferences and demands;
- (iii) housing, health and education statistics;
- (iv) small, medium and large scale manufacture, trade, transport, building construction and services;
- (v) village/block (*bario*/enumeration district) statistics relating to the existence of certain amenities such as railway stations, electricity, post offices, etc.;
- (vi) land utilization and crop yield statistics, land holdings and cost of cultivation; and
- (vii) wholesale and retail prices.

In this connexion, reference may be made to the document "Basic statistics

ON DESIGNING AND CONDUCTING MULTI-SUBJECT SURVEYS

for formulating and implementing plans of economic and social development in countries of Asia and the Far East" (United Nations, 1964b). It is necessary to get information not only on the aggregates of these items at one point of time, but also on the trend exhibited by them over time.

2.3. *Tabulation requirements.* The next step in specification of data requirements consists in specifying the tables with the details of types of characteristics to be estimated, as well as the geographical and classificatory breakdowns for which data are required. At this stage, if quick results are to be obtained with limited resources, it is necessary to check the tendency to require data by very detailed geographical and classificatory breakdowns. This aspect of data specification requires very careful consideration and should be oriented to the purposes in view.

As a part of specification of data requirements, it is necessary to specify the items or groups of items for which (i) aggregate estimates, (ii) estimates of trend over time or (iii) both aggregate and trend estimates are required together with the frequency (monthly, quarterly, yearly, etc.) with which they are required. Here again, it is desirable to be less ambitious, as the utility of obtaining reliable estimates in general and those of changes in particular over short intervals of time would not usually be commensurate with their cost.

2.4. *Permissible error.* The third step in specification of data requirements consists in giving careful consideration to the question of specifying the permissible errors, that is, the margins of error that would be permissible in the estimates of the various characteristics in the light of the action to be taken on the survey results. Usually, this is not an easy task because of the complex nature of the decision-making process. But it is very desirable that efforts be made to specify the permissible errors in a rational manner on the basis of an examination of the action taken on past survey results and the loss or gain resulting therefrom. This is a matter of considerable importance as over- or under-specification of the permissible errors would lead to loss due to wrong decisions or to an increase in cost of the survey. However, in the initial stages when very few data on any particular subject are available, it may be worthwhile even to get estimates subject to moderately large errors and then to reduce the margin of error in subsequent surveys, if found desirable from the joint consideration of cost of survey and possible loss due to wrong decisions based on survey results.

3. MULTI-SUBJECT SURVEY

In a multi-subject survey, data on two or more subjects, not necessarily very closely related, are collected in a single joint operation for the sake of economy and convenience (United Nations, 1964c; Lahiri, 1963; Murthy, 1967, ch. 14). Experience in some countries has shown that multi-subject surveys, carried out on a continuing basis, are generally more economical than a series of *ad hoc* uni-subject surveys, provided that the subjects of enquiry are not so numerous and diversified as to affect the quality of data. The conditions obtaining in the developing countries regarding the sampling frame, educational standards, transport and communications, availability of trained personnel, etc., favour the use of multi-subject surveys. Since a large section

of the population are illiterate, postal enquiry is not feasible in most cases in these countries and hence we shall consider only the method of interview and of direct physical observation. In this section, the concept and the advantages of multi-subject surveys are briefly discussed and, in the later sections, the implications of carrying out multi-subject surveys are considered in greater detail with special reference to household enquiries.

3.1. *Extent of integration.* The extent of integration of the different subjects of enquiry would depend on a number of factors, such as the nature of subjects, method of enquiry, type of ultimate sampling unit and sampling frame. For instance, (i) data on population and employment may be collected from a common set of sample households, since the sampling unit is the same and the method of enquiry is the interview method for both these subjects; and (ii) when data on consumer expenditure and land utilization are required, one may consider the possibility of having a common set of sample penultimate stage units (such as villages, *barios*, etc) and collecting data on the former from a sample of households by enquiry and on the latter from a sample of fields or farms by actual physical observation.

3.2. *Less cost or greater precision.* The economy in a multi-subject survey, integrating two or more subjects of enquiry, arises mainly from savings in overhead cost, which consists of setting up a survey organization with adequate and suitable administrative and supervisory staff, and in survey cost, which comprises travel to the penultimate stage sample units (villages, enumeration districts, etc.), camp-setting, and listing and contacting the ultimate stage units in the sample penultimate stage units. Economy in overhead cost is not only important in terms of money but also because of the limited number of trained personnel in the developing countries. Economy in survey cost may not be considerable if the survey is carried out using *ad hoc* local staff. But use of *ad hoc* local staff may not be desirable except in a few special situations, in view of the advantages of having a permanent staff which are described in the next section. Usually in large-scale surveys in the developing countries, one has to resort to multi-stage and multi-phase sampling designs and, when the survey is performed by a wholtime staff, there is considerable advantage in integrating two or more subjects to form a multi-subject survey.

The savings achieved in integrating two or more subjects may be utilised for increasing the number of sample units in the first stage (and in subsequent stages up to the penultimate stage, if necessary), thereby improving the efficiency of the survey estimates, since the contribution to the total variance of the variation between first stage units is generally more important than that between the ultimate stage units. Alternatively, for the same precision, increasing the sample size at the first few stages would generally lead to a reduction in the total number of ultimate stage sample units, which in its turn would reduce the cost of tabulation. Further, integration of two or more subjects by having the same sample penultimate stage units reduces the total cost of the survey, including the overhead costs. The increase in the sample first stage units may also facilitate deeper stratification, taking into account more than

ON DESIGNING AND CONDUCTING MULTI-SUBJECT SURVEYS

one auxiliary characteristic when available, with a view to increasing the overall efficiency of the survey design.

Integration of enquiries makes the work of data collection in the penultimate stage sample units more worthwhile in terms of the overhead costs, since the investigator is able to stay in those units canvassing data on different subjects for a longer time than is possible in a uni-subject survey. This aspect is important from the point of view of the quality of data, as a longer stay enables the investigator to become familiar with the local conditions and to establish cordial relations with the local people. Furthermore, less time is lost in idleness owing to the flexibility the investigator has in adjusting the work of collecting data from different sample households or other units in case some units are temporarily not available for data collection.

4. PERMANENT SURVEY ORGANIZATION

4.1. *Advantages.* An essential pre-requisite for successfully implementing a multi-subject sample survey is the existence of a permanent sample survey organization with trained personnel to do the work of survey designing, data collection and processing. In fact, the report of the Working Group of Experts on Sampling Methods, which met at Bangkok in 1969, emphasized the need for a permanent survey organization (United Nations, 1969). A permanent sample survey organization with a whole-time staff experienced in different aspects of surveys would be in a very sound position to meet adequately the growing needs for statistical information in the developing countries.

As the work of designing and carrying out multi-subject surveys efficiently requires considerable skill derived from experience in this field, one of the primary objects of the permanent organization should be the building up of a competent and experienced nucleus staff able not only to conduct large scale surveys but also to evolve a suitable programme of training the primary staff in sampling methods, methods of data collection and processing work. One of the main advantages of conducting a multi-subject survey through a permanent survey organization with whole-time staff as compared to carrying out of a series of uni-subject surveys with part-time or temporary staff is that this requires fewer trained personnel, which is a very important consideration in the developing countries.

The preliminary work involved in conducting a sample survey (evolving a suitable sampling design, sample selection, planning the work programme, preparation of schedules and instructions and training the field staff) is considerable; it requires that information be obtained on cost and variance functions for different sampling designs and on the question of feasibility of collecting the required statistical data. A permanent survey organization is able to carry on studies relating to sampling design, methods of data collection, analysis and presentation of data as part of its normal activities with a view to evolving suitable survey procedures best suited to the local conditions. Moreover, in such an organization, the experience of the survey personnel is not lost, so they can continue to improve themselves through a process of trial and error; this is not possible if *ad hoc* staff are employed, as their experience gets lost.

4.2. *Control of non-sampling errors.* A point which needs special mention is that a permanent organization is in a better position to assess and control non-sampling errors in its surveys. This is important, as considerable experience in survey work is required even to recognize the existence of non-sampling errors and then to realize that they may be frequent and considerable unless special steps are taken to control them. In fact, it would be advisable for the permanent organization to set apart some of its resources for studying the problems involved in assessment and control of errors on a continuing basis, since experience has shown that non-sampling errors arising from various sources (defective sampling frame, faulty methods of data collection and compilation) may, under certain circumstances, be considerably more than the sampling error.

4.3. *Meeting of urgent demands.* A permanent survey organization with whole-time field staff in the different parts of the country will obviously be in a position to meet urgent demands for statistical information on a particular topic of current interest to the government. This is an important feature, as it is not uncommon for the governments of developing countries to require statistical data on some topic or other at short notice in view of the lack of objective data in many fields, which are essential for national planning for economic development and plan assessment. In this connexion, it would be desirable for the survey organization to make its survey plan flexible by a planned phasing of the work of data collection over shorter periods of time (such as a month, two months, etc.) in such a way as to be able to obtain valid estimates periodically even while the survey is in progress.

4.4. *Sampling frame.* One of the main difficulties in designing and conducting multi-subject surveys in the developing countries is the lack of suitable sampling frames. Though the decennial census is expected to provide a sampling frame as a by-product of its activities, in some cases this frame proves inadequate for sampling purposes owing to the difficulties of identifying the units after the lapse of some time and also due to possible changes in the boundaries of those units, (Murthy 1968). A permanent survey organization, since it will have a direct interest in the evolution of a suitable frame, can exert its influence in preparing well-defined, identifiable, compact geographical area units at the time of the census, and it should also set apart some of its resources for bringing the frame up to date periodically. This work would be of much help not only to the survey organization, but also to the census agency.

4.5. *Deterrents to having a permanent organization.* Possibly the reasons for not setting up a permanent sample survey organization in some countries are the considerable expenditure involved, especially at the initial stages, and the existence of some surveys (uni-subject or multi-subject, continuing or *ad hoc*) conducted by different departments or provinces, states, prefectures, etc. As to the first point, the long-range benefits that would accrue from a permanent survey organization would more than offset the initial expenditure. One way adopted by some countries to meet the second point is to encourage the departments and the provinces, states, prefectures, etc., to participate in the work of the permanent survey organization on a partial

ON DESIGNING AND CONDUCTING MULTI-SUBJECT SURVEYS

or full matching basis, by doing the field and tabulation work in respect of an additional sample using the same concepts, definitions and procedures. This type of participation would help in providing a broad check on the survey results and in getting better estimates by pooling the estimates obtained by the different agencies.

5. INTEGRATION OF HOUSEHOLD ENQUIRIES

Integration of household enquiries relating to different subjects is important not only because it results in economy, but also on account of considerations of quality and usefulness of data in the case of the developing countries, whose economies, though mainly agricultural, are complex in nature owing to the households being engaged in more than one activity. The report of the Group of Sampling Experts which met in Geneva in 1963 defined the "integrated survey" as a survey, where data on several subjects (or items, or topics) are collected for the same set of sampling units for studying the relationship among items belonging to different subject fields (United Nations, 1964c). It points out that such surveys are of special importance in studies on levels of living and that integrated surveys of consumption and productive enterprises are also of special importance in developing countries, where the related activities are frequently undertaken in an integrated manner in the household.

5.1. *Types of integration.* In principle, it should be possible to obtain all the information that is required about the household part of the unorganized sector of the economy through an integrated multi-subject household survey. A question that needs careful consideration relates to the nature and the extent of integration of the different subject fields. The data requirements, the purposes for which they are to be collected and the question of feasibility should be constantly kept in view in answering it. For instance, if we are interested in getting data on :

- (i) population, births and deaths;
- (ii) income and expenditure;
- (iii) land holdings and agricultural practices; and
- (iv) household enterprises (manufacture, trade, transport, etc.).

the point needing attention is whether :

- (a) to integrate all the enquiries completely by collecting all the required data from the *same set* of sample households having a common sampling design;
- (b) to integrate these subjects in the sense of having a *common set* of sample villages/barrios/enumeration districts, and to have *separate* sample of households for the different subjects; or
- (c) to have *different designs* whose use will depend on the available supplementary information.

5.2. *Complete integration.* The main advantage of alternative (a) apart from the economy achievable through complete integration, is that it introduces flexibility into the tabulation of results through suitable cross-classification using meaningful items belonging to different subjects. For instance, if subjects (i) and

(ii) are related, then one can study the birth and death rates and other demographic characteristics in relation to the level of living of the household as measured by *per capita* income or expenditure or some other suitable indicator. Another advantage which is equally important, concerns the quality of data. As mentioned earlier, the economy of developing countries being a mixed one, especially in rural areas, collection of data on income and expenditure and on enterprises when carried out separately may be subject to greater response error due to unconscious mixing up of the household consumption and expenditure in relation to the enterprise than would occur in an integrated survey, as, in the latter, both aspects are covered for the same household and at the same time.

Alternative (a), though it has considerable advantages in the context of the conditions obtaining in the developing countries, has not yet gained wide acceptance. The lack of a permanent organization specializing in sample survey methods makes it difficult to implement this approach in many countries, for considerable expertise is required in planning and conducting a comprehensive survey of this nature. Further, there may be genuine doubts about the advisability of collecting all the data from the same sample households, lest the enlargement of the scope of data collection lead to the data being of poor quality on account of respondent and investigator fatigue. This aspect needs careful study; it appears that, with suitable pre-tests of different types of integrated schedule under field conditions, it should be possible to evolve a workable integrated schedule. If necessary, the possibility of collecting data in more than one sitting through repeated visits may also be kept in view.

5.3. *Partial integration.* A second approach, which is a *partial* integration of the surveys, is based on the expectation that the sampling efficiency can be improved by using possibly different frames in the first (or penultimate) stage sample units. For instance, in the case of a land holding survey, it may be desirable to stratify the households in the sample first stage units on the basis of a rough idea of the holding size collected at the listing stage and to have disproportionate sampling fractions. Similarly, in the case of a household manufacturing survey, it may be advisable to compile a frame of households having this activity at least in a subsidiary capacity and to select sample households using this sub-frame.

5.3. *Augmented sampling.* A third approach is usually adopted when there is no central agency responsible for collecting statistical data. For instance, the different departments of the government may undertake isolated surveys for collecting data in which they are interested. If surveys of a similar nature with the sampling unit being the same (household) are to be conducted, then an attempt should be made at integration, as otherwise there would be a wastage of effort and resources. However, there may be situations where the sampling efficiency for a given total sample size is more in a uni-subject survey with a separate sampling design than would be the case in an integrated design. But in such cases a comparison should be made of the efficiency per unit of cost, as this would be more realistic in practice. Still, if there is a situation where the characteristic under consideration is concentrated in a few first

ON DESIGNING AND CONDUCTING MULTI-SUBJECT SURVEYS

stage units (household manufacturing enterprise in some regions, for example), then the general purpose sample may not be adequate for estimating this characteristic; in this situation, one may explore the possibility of supplementing the general sample by a relatively small additional sample selected from the first stage units exhibiting a concentration of that characteristic.

5.4. *Combined approach.* Taking an over-all view of the requirements of data for the household sector as a whole and for the different sectors of the economy such as household manufacture, trade, etc., sections of the population and geographical regions, it appears desirable to adopt approach (b) in addition to (a), with the difference that, even in (b), an integrated approach (possibly with an abridged version of the general integrated schedule) be used in collection of data with emphasis on the subject matter of direct interest. This combined approach is being suggested because, from the point of sampling efficiency, it may be necessary to have different subsampling fractions for the different subjects. If necessary, the possibility of supplementing the sample of first stage units mentioned earlier may also be considered.

6. PROBLEM OF SAMPLING FRAMES

6.1. *Need for area units.* One of the main requirements for efficiently designing a multi-subject survey sample is a set of well-constructed up-to-date sampling frames with information on suitably selected auxiliary characteristics. In developing countries, this constitutes a basic problem as usable sampling frames are either not available at all or have to be compiled from frames, constructed mainly for administrative purposes at different places. Even when a frame is available, steps should be taken to ensure that it is free from omissions, duplications and other inaccuracies and that the units are clearly identifiable in the field. The report of the first Working Group of Experts on Sampling suggests the building up of a frame with small, compact and identifiable area units (United Nations, 1959, p. 11).

6.2. *Master frame of area units.* After having ensured a *minimal frame* which may consist of a list of villages or enumeration districts with identification particulars, one should explore the possibility of getting information on suitable auxiliary characteristics for using in stratification, allocation, arrangement, selection or/and estimation in designing a multi-subject survey. For this purpose, data collected by different agencies or departments as a part of their normal activity may have to be compiled and collected. This work is not likely to be easy if data are to be compiled from local agencies, and there may be difficulties in matching the information obtained from different agencies with the basic frame. It may be noted that a permanent survey organization is best suited to undertake this task and that, once a *master frame* is built up, it can be of considerable use in designing multi-subject surveys. Moreover, if the basic frame is well constructed, the information on auxiliary variables obtained through compilation and collation need not be very accurate, as only broad information on them would be needed for improving the sampling design.

6.3. *Frame of housing units and households.* In most of the countries, a stratified multi-stage sampling design is usually adopted for household surveys, as an up-to-date list of households is rarely readily available and it would be uneconomical

to survey a direct sample of households due to the cost of travel, making contact, etc. Thus, the master frame of villages, enumeration districts or other area units is only used for obtaining an efficient sample of the first stage units. Then the question of an appropriate sampling frame within the selected first stage units needs to be examined. Here again it is possible that no frame at all is available, or else it may be possible to obtain lists of persons on electoral or other registers. In the former case, a fresh frame of households has obviously to be prepared within the selected first stage units. In the latter case, the experience has been that usually these lists are not kept up-to-date; hence it is desirable to prepare a list of the households currently residing in those units by updating and checking the available list.

While listing the households within the selected first stage units, it is very desirable to list also all the houses and housing units, as this will help in locating vacant housing units, small scale enterprises, etc., besides reducing the possibility of omission of some households. A frame of housing units is useful in a housing survey, where a housing unit may be a suitable sampling unit; and a frame of enterprises, obtained from a list of housing units, can be used in an enterprise survey. Further, if up-to-date lists of houses or housing units are available, they may be used as frames for household enquiries by considering the canvassing of all households or a sample of households residing in a selected house or housing unit. Incidentally, it may be noted that even when household is used as the sampling unit, it is possible to collect and analyse data for housing units by noting the number of households residing in the housing unit in which the sample household is situated.

At the stage of preparing a sampling frame within the selected first stage units, it is also desirable that information on some auxiliary characteristics with a bearing on the subjects of enquiry be collected for each household, for use in increasing the efficiency of sampling of households at the second stage with a view to reducing the within village variability. Such procedures make the contribution of the within village variation to total variation less important, thereby requiring a smaller sample in the first stage units to ensure a specified precision for the estimator.

7. CHOICE OF RATIONAL SURVEY DESIGN

Having ascertained the data requirements in terms of the types of data, their geographical breakdowns and the frequency with which they are required, one should proceed to evolve a rational survey design to meet the requirements of the users of statistical data satisfactorily at a reasonable cost. It has already been pointed out that it is far better to conduct integrated multi-subject surveys through a permanent survey organization than to have a series of uni-subject surveys undertaken by *ad hoc* or/and part-time staff. The survey design should take into account the need for periodic estimates of both levels and trends (changes), noting that the periodicity will depend on the type of estimates required.

7.1. *Ad hoc versus continuing survey.* Periodic estimates may be obtained either by conducting *ad hoc* surveys at the two end points of the specified period or by conducting continuing surveys and obtaining estimates at the two end points of the

ON DESIGNING AND CONDUCTING MULTI-SUBJECT SURVEYS

specified period by pooling the relevant data. The latter scheme, besides permitting integrated multi-subject surveys to be carried out on a continuing basis (which has considerable technical and operational advantages) has the added advantage of eliminating the risk that either of the two end points of the period may be a freak in the sense of having an unusual economic fluctuation; it also permits the study of trend by considering the estimates for the intervening periods also. Furthermore, this scheme makes it possible to obtain usable data for different types of geographical breakdown by pooling the estimates over a period of time, provided that the geographical areas are not too small.

7.2. *Flexibility of design.* As was pointed out earlier, it is not generally possible to collect all the data required in an integrated manner, especially in continuing surveys. The nature and the extent of integration in a continuing survey should be decided, taking into account the local conditions obtaining in the country in respect of sampling frame, type of personnel available, respondent cooperation, etc. In fact, it is desirable to have only a few subjects of enquiry in the initial stages and to increase its scope in a phased manner as survey experience accumulates. The survey design should be flexible enough to allow the addition of some *ad hoc* enquiries on topics of special interest, to be conducted as a part of the multi-subject survey. In fact, it may be desirable to set apart some of the resources every year for this purpose. The possibility of marginal expansion of the staff to permit the taking up of a special survey together with the regular surveys may also be kept in view; but this expansion is possible only when a decision about the special survey has been taken well ahead of the survey period, in which case there is enough time to recruit and train the additional staff.

7.3. *Programme of work.* In evolving a rational survey design, it is necessary to consider carefully the question of programming the work in the field and at the tabulation stage. The field programme of work should be so drawn up as to obtain the maximum information per unit of cost consistent with the required standards regarding the quality of data. Since, in a continuing survey, there should be a balance between the workload in the field and the tabulation load, the tabulation programme of work should be so planned as to avoid bottlenecks and delays in the publication of results.

7.4. *Need for a rational design.* The evolution of a suitable sampling design for a multi-subject survey utilizing all the available relevant information in the sampling frame presents many problems. Even the formulation of the concept of optimum design in the case of multi-subject surveys is quite difficult, hence one has to be satisfied with the choice of a *rational* sampling design from among those possible. The principle to be adopted in this choice should be the reduction of the overall cost including the field and the tabulation costs when the permissible errors are pre-specified, or the reduction of the margins of error in the estimates when the total cost is fixed. In both the cases, it is extremely important to take adequate steps to control the non-sampling errors, which may sometimes be large enough to vitiate the survey results.

In this context, special attention should be paid to the feasibility of implementing a particular sampling design under the operational conditions obtaining in the country.

7.6. *Self-weighting design.* Another point needing attention is the question of making the sampling design self-weighting, which is of considerable importance from the view point of tabulation of data. In the developing countries, where data are processed manually or through the use of conventional tabulating equipment, weighting the sample observations with the appropriate inflation factors, which is necessitated by the use of a non-self-weighting design, may create a bottleneck at the tabulation stage and may hold up quick tabulation. Hence, efforts should be made to make the sampling design completely self-weighting with one single inflation factor; if this is not possible due to operational and other considerations, the design should, at least, be made partially self-weighting in the sense of having a few inflation factors.

7.6. *Stratification and allocation.* As mentioned earlier, generally a stratified multi-stage design (two stage or three stage usually) is adopted for multi-subject household surveys. If information is available on two or more auxiliary variables connected with the subjects of enquiry under consideration, multiple or deep stratification may be resorted to using such information. This is usually possible because experience has shown that the formation of a few strata in an optimum or near optimum manner on the basis of a stratification variable results in the extraction of most of the benefit of stratification, and sub-stratification of these few strata into sub-strata formed on the basis of another stratification variable leads to deep stratification. To facilitate the field work, it is desirable to make the strata geographically compact so that each stratum can be conveniently considered as the investigation zone for an investigator. The question of allocation may be tackled by considering the allocations arrived at on the basis of the data on auxiliary variables and by accepting a compromise allocation. Here again, it is generally found that even moderate deviations from a specific allocation do not affect the overall variability considerably. However, if the allocation for a particular subject turns out to be very different from the other allocations, special steps such as supplementing the sample may be considered to avoid deviating substantially from the specific allocation.

7.7. *Selection of area units and households.* The selection of the first stage units (generally area units such as villages, enumeration districts, *barrios*, etc.) within the strata may conveniently be done by systematic sampling, or by sampling, with varying probabilities systematically after arranging the units in a suitable order. In varying probability sampling the question of choice of the measure of size arises. If the interest lies mainly in the whole population, the previous census population may be a suitable size for household surveys. However, if the size measure has already been considered as a stratification variable or if the sizes are almost equal (as in case of census enumeration districts or blocks), sampling may be done with equal probability systematically. The possibility of selection with probability proportional to other suitable measures of size through an appropriate integrated method of selection may also be considered. Alternatively, the other measures of size may be used to

ON DESIGNING AND CONDUCTING MULTI-SUBJECT SURVEYS

improve the estimates at the estimation stage through the use of ratio or regression estimators. In the selected first stage units, the households are to be listed preferably with some information on auxiliary variables having a bearing on the subjects of interest, and selection may be done systematically from the whole frame or from a suitably constructed sub-frame (in the case of special enquiries) with a suitable arrangement or sub-stratification.

7.8. *Interpenetrating sub-samples.* It would be useful to select the sample in the form of two or more interpenetrating sub-samples, each of which is capable of providing a valid estimate of the population parameter (Mahalanobis, 1946; Lahiri, 1954; United Nations, 1964c). The main advantage of this technique is that it permits easy calculation of the estimates of the variances of the survey results irrespective of the complications involved in selecting the sample and in the estimation procedure used. This is important in the case of sampling designs where the first stage units are selected with varying probabilities without replacement or systematically or/and where complicated estimators such as ratio and regression estimators are used; since, in these cases, the expressions for the sampling variances and their estimators are mathematically complicated and are difficult to compute numerically on the basis of a single sample. When these sub-samples are surveyed and processed by different sets of field and processing staff, a comparison of the results based on the sub-samples provides a broad check on the quality of the survey operations. This technique can also be used to study the non-sampling errors, such as the differential investigator bias (that is, the difference between the biases of investigators or groups of investigators), the differential effects of different methods of investigation, etc., by getting the different sub-samples canvassed by different investigators (or groups of investigators) or by subjecting the sub-samples to different methods of investigation, etc.

8. SURVEY AND REPORTING PERIODS

8.1. *Staggering of the survey.* In a multi-subject household survey, it is advisable to take one complete year as the period of survey and to spread the work of data collection evenly so as to be able to collect data for different parts of the year. This is especially important in the developing countries, where the economy is still mainly agricultural and hence subject to pronounced seasonal fluctuations within a year. This staggering of the survey, which is in a way necessitated by a set of permanent survey staff, is particularly desirable since the reporting period for many of the items has to be less than a year (a week, month, etc.) due to the possibility of recall lapse.

It may be mentioned that a permanent survey organization is not well suited to carry out point surveys, where the interest lies in getting reliable estimates at a particular point of time, as such surveys require the canvassing of a large sample at a particular point of time or during a very short period so as to avoid recall lapse. However, this may not be a serious disadvantage, for the average picture of the economy obtained through a staggered survey is likely to be more useful in a seasonal economy than a point survey with a short reference period.

8.2. *Moving reporting period.* Another point needing attention is that, in a multi-subject survey conducted through a fixed field staff by staggering the survey, it is difficult to use fixed reporting periods owing to recall errors resulting from the time lag between the reporting period and the survey date; hence it may be desirable to use a moving reporting period, where the reporting period is one of fixed duration prior to the date of survey, such as the week or the month preceding the date of survey. A large moving reporting period such as a year may also be used in such cases as the collection of vital statistics, if found feasible. In such cases, it may be useful to collect some auxiliary information to permit working with a smaller reporting period, as this helps in studying the nature and extent of recall lapse and in obtaining data for a specified fixed reporting period also (Som, 1968).

The problem of reporting periods requires careful consideration, especially in an integrated household survey, as operational convenience requires the use of the same reporting period for all items. This is also necessary for providing some internal checks so as to obtain a total picture of the economic activity of the household. This common reporting period has of necessity, to be a short one (a week or a month); though, for certain items where recall lapse may not be considerable, a larger reporting period may be desirable, and in such cases it may be preferable to collect data for both the common and specific reporting periods.

While staggering the survey, it is advisable to divide the survey period of one year into shorter periods (of one, two or three months) and to canvass a representative sub-sample of the total sample in each such shorter period with a view to obtaining valid estimates periodically even when the survey is in progress. This point has been previously mentioned in connexion with the possibility of meeting expeditiously an urgent demand for certain data of current interest.

9. FIELD WORK

9.1. *Need for trained investigators.* All aspects of field work, such as recruitment and training of investigators, inspection and supervision, etc., should be given careful consideration in the light of the operational conditions obtaining in the developing countries, since the final survey results can at best be only as accurate as the basic data collected in the field. The need for highly trained and experienced investigators for carrying out a multi-subject household survey becomes obvious when one recalls that, since a large section of the people in the developing countries is illiterate and not familiar with the concepts and terminology used in the study of the economic conditions, there is considerable scope for ascertainment errors. The investigators should be trained well not only in the concepts and definitions of the terms to be used in collecting the required data, but also in the art of eliciting accurate (or at least reasonably accurate) information through asking probing questions in a tactful manner. In recruiting the investigators, special attention should be paid to their ability to withstand long and arduous travel, possibly in inhospitable terrain, as transport facilities may not be well developed in many parts of the developing countries. It is desirable that the investigators be trained by a single team of instructors instead

ON DESIGNING AND CONDUCTING MULTI-SUBJECT SURVEYS

of through a chain of intermediaries. The latter course should be adopted only when necessitated by cost consideration and, even then the number of intermediate stages of training should be kept to a minimum.

9.2. *Inspection and supervision.* Inspection and supervision are indispensable in assessing and controlling non-sampling errors, especially in a large scale multi-subject survey and suitable arrangements should be made for carrying out this work efficiently. To facilitate this task, the country may be divided into certain divisions in each of which the field work is entrusted to a team of investigators and inspectors headed by a supervisory officer; and then these divisions may be sub-divided into inspection zones, each comprising a group of investigation zones. Inspection work would consist of different types of check on the work of the investigators, such as spot-checks, pre- and post-survey checks, scrutiny of filled-in schedules, etc. Besides the regular inspection work oriented towards improving the standard of the weak investigators, it may also be useful to check a sample of the work of all the investigators selected in such a manner as to get overall estimates of the quality of the data collected in the field. The supervisory staff should be in constant touch with the staff designing the survey and processing the data, so that they can obtain clarifications and instructions on the concepts, definitions and procedures used in the field and apprise those designing the survey of the difficulties encountered in carrying out their instructions in the field under the actual operational conditions.

The ratio of the supervisory staff to the primary staff should be fairly high in the developing countries, as, on account of inadequate transport and communication facilities, they will have to spend a considerable part of their time on travelling and contacting the investigators, which will reduce the time available for the actual inspection work. It is advisable to undertake a preliminary scrutiny of the data collected in each field division before sending them to the survey organization for processing, as any errors detected may conveniently be referred back to the investigators concerned for rectification and explanation.

10. PROCESSING OF DATA

10.1. *Control of errors.* Processing of survey data in a multi-subject household survey involving detailed quality and routine scrutiny of the data, mechanical and manual computations, etc., needs such skill; hence a well-trained and experienced staff will have to be built up for this purpose. It is extremely important that there should be built-in checks and cross-checks on the computations, so that the non-sampling errors introduced at this stage of the survey operation are minimized. Suitable sampling methods may be used to assess and control the errors involved in different stages of the tabulation work. Considerable planning of the data processing work is necessary so as to permit a smooth flow of data through the various stages of the tabulation operation and ensure that the final results of the survey will have a fair degree of accuracy. While one should be careful about and, as far as possible, avoid the introduction of errors at this stage, one should nevertheless not be too meticulous

about producing error-free results, as the effort involved in eliminating the last few traces of error is likely to be so substantial as not to be worthwhile.

10.2. *Study of inter-relationships.* In a multi-subject survey, there is considerable scope for studying various types of inter-relationships among different socio-economic characteristics. The extent of the classificatory breakdowns is likely to be limited only by the inadequacy of sample size for very detailed cross-classifications. This limitation is mainly due to cost and operational considerations. However, in a continuing survey, it can be overcome, at least to some extent, by pooling the estimates over time. It may be advisable to carry out the work of processing of data also in a phased manner giving priority to certain basic tabulations of immediate interest. A detailed analysis of the survey data to study the different types of inter-relationships of the variables may be taken up later; but it is very desirable that the work of processing one year's field data be completed within about a year so as to avoid a possible lag in tabulation work.

10.3. *Fractile graphical analysis.* In tabulating the data from a multi-subject survey, it would be useful to adopt the fractile graphical method of analysis. This consists of dividing the sample observations into a number of groups after arranging them in increasing order of a suitable basic variable (such as *per capita* consumer expenditure in a family budget survey) in such a way that the groups are of equal content with respect to the estimated number of units in them, and then calculating the estimates of certain meaningful averages and ratios for each of the *fractile* groups. If the estimates of the variable under study for the different fractile groups are plotted and if the successive points are joined by straight lines, then a visual (geometric) picture of the relationship between the study and the basic variables is obtained and this graph is termed a *fractile graph*. By obtaining such fractile graphs separately for two interpenetrating sub-samples and for the combined sample, it is possible to get an idea of the error of the combined graph by considering the area between the sub-sample graphs and such an analysis is termed *fractile graphical analysis*. (Mahalanobis, 1960; United Nations, 1964c). This procedure is of considerable help in studying effectively the inter-relationship between different characteristics with reference to the basic variables and in permitting realistic comparisons over time and space.

REFERENCES

- LABIRI, D. B. (1954): Technical paper on some aspects of the development of the sample design. *National Sample Survey Report No. 5*, Government of India, reprinted in *Sankhyā*, 14, 204-316.
- (1955): Multi-subject sample survey system—some thoughts based on Indian experience. *Contributions to Statistics*, 115-220, presented to Professor P. C. Mahalanobis on the occasion of his 70th Birthday, Pergamon Press, London and Statistical Publishing Society, Calcutta.
- MAHALANOBIS, P. C. (1946): Recent experiments in statistical sampling in the Indian Statistical Institute. *Journal of Royal Statistical Society, (A)*, 109, 325-378, reprinted in *Sankhyā*, 20, (1958), 1-68.
- (1960): A method of fractile graphical analysis. *Econometrica*, 28(2), 325-351; reprinted in *Sankhyā*, 23(A), 1961, 41-64.
- MURTHY, M. N. (1967): *Sampling Theory and Methods*, Statistical Publishing Society, Calcutta.
- (1968): Population census as the source of sampling frame in India; to be presented to the 56th Session of the Indian Science Congress to be held at Bombay.
- UNITED NATIONS (1959): *Report of the First Working Group of Experts on Sampling*, ECAFE document E/CN.11/517.
- (1964a): *Handbook of Household Surveys*, Studies in Methods Series F, No. 10, New York.
- (1964b): *Basic Statistics for Formulating and Implementing Plans of Economic and Social Development in Countries of Asia and the Far East*, ECAFE document E/CN.11/ASTAT/Con.6/L.4.
- (1964c): *Recommendations for the Preparation of Sample Survey Report*, Statistical Paper Series C, No. 1, Rev. 2, New York.
- SOM, R. K. (1968): *Recall Laps in Demographic Enquiries*, Asia Publishing House, Bombay.

Paper received : October, 1968.