*This report contains the collective views of an international group of experts and does not necessarily represent the decisions or the stated policy of the World Health Organization.*

*WORLD HEALTH ORGANIZATION*

*TECHNICAL REPORT SERIES*

No. 336

# SAMPLING METHODS IN MORBIDITY SURVEYS AND PUBLIC HEALTH INVESTIGATIONS

## Tenth Report of the WHO Expert Committee on Health Statistics

Page



WORLD HEALTH ORGANIZATION

GENEVA

1966

## WHO EXPERT COMMITTEE ON HEALTH STATISTICS

*Geneva, 23-29 November 1965*

*Members :* *

Dr P. Armitage, Professor of Medical Statistics, Department of Medical Statistics and Epidemiology, London School of Hygiene and Tropical Medicine, London, England (*Rapporteur*)

Dr F. E. Linder, Director, National Center for Health Statistics, Department of Health, Education and Welfare, Public Health Service, Washington, D.C., USA (*Chairman*)

Dr J. N. Manceau, Director, Special Service for Public Health, Rio de Janeiro, Brazil

Dr K. K. Mathen, Professor of Statistics, All India Institute of Hygiene and Public Health, Calcutta, India

Dr V. K. Ovčarov, Head, Health Statistics Department, Semashko Institute of Health Administration and the History of Medicine, Moscow, USSR

Dr M. Vacek, Statistical Department, Institute for Research into the Organization of the Health Services, Prague, Czechoslovakia (*Vice-Chairman*)

*Representatives of Other Organizations :*

United Nations : Mr C. Scott, Regional Statistical Adviser, UN Economic Commission for Africa, Office of the Resident Representative, Accra, Ghana

International Labour Organisation : Mr G. Tamburi, Actuarial Section of the Social Security Branch, Geneva, Switzerland

*Secretariat :*

Mr W. Z. Billewicz, Reproduction and Growth Research Unit, Princess Mary Maternity Hospital, Newcastle upon Tyne, England (*Temporary Adviser*)

Preceptor T. Dalenius, Institute of Statistics, University of Stockholm, Sweden (*Temporary Adviser*)

Dr W. P. D. Logan, Director, Division of Health Statistics, WHO

Dr C. Rouquette, Chief, Pediatrics Section, National Institute of Health and Medical Research, Paris, France (*Temporary Adviser*)

Dr C. C. Spicer, Chief Medical Statistician, General Register Office, London, England (*Temporary Adviser*)

Mr K. Uemura, Chief Statistician, Health Statistical Methodology, Division of Health Statistics, WHO (*Secretary*)

---

* Unable to attend : Professor D. Schwartz, Director, Statistical Research Unit of the National Institute of Hygiene, Gustave Roussy Institute, Villejuif, Seine, France.

---



PRINTED IN SWITZERLAND

*Wld Hlth Org. techn. Rep. Ser.*, 1966, **336**

# SAMPLING METHODS IN MORBIDITY SURVEYS AND PUBLIC HEALTH INVESTIGATIONS

## Tenth Report of the WHO Expert Committee on Health Statistics

The WHO Expert Committee on Health Statistics met in Geneva from 23 to 29 November 1965 to discuss the role of sampling methods in morbidity surveys and public health investigations. Dr P. Dorolle, Deputy Director-General, opened the meeting on behalf of the Director-General. Dr F. E. Linder was elected Chairman and Dr M. Vacek, Vice-Chairman. Professor P. Armitage was appointed Rapporteur.

### 1. INTRODUCTION

Much of the statistical information on which the administration of health services depends is provided by large permanent systems of reporting or registration, usually covering the whole population under study. Such systems will always be necessary to some extent, but they are often expensive to operate and not easy to modify or improve to meet changing needs for information.

Furthermore, it is clear that the functioning, effectiveness and utilization of the health services depend on a great number of complex interactions between sociological, biological, administrative and other factors. To study these factors, especially when resources are limited, requires a flexible and economical tool that can be used as an adjunct to registration and the other permanent statistical systems.

Sampling inquiries provide such a tool and should take their place as an essential part of the health statistical services. Much progress has been made in the practical and theoretical aspects of sampling techniques and, especially since the introduction of electronic computers, in their application to the solution of complicated statistical problems. Their main limitation, the necessity to argue from a small fraction of the population to the whole, is offset by their adaptability and speed of evaluation and, often, by the better quality of information that can be obtained, since the smaller numbers allow the use of better-trained personnel, more accurate technical methods and closer supervision. The estimates of population characteristics derived from a properly chosen sample are adequate under most circumstances.

It should be emphasized that sampling inquiries are not only suited to developed countries but are of equal or even greater value in developing areas. In the latter, quite simple sampling investigations may often provide information unobtainable in any other way. In the former, they may be used to supplement the large masses of rather crude information accumulated for other purposes, such as hospital records or death registrations, or to extract from them an almost equally informative sample of more manageable size.

## 2. THE PURPOSE AND USE OF SAMPLING TECHNIQUES IN MORBIDITY SURVEYS AND PUBLIC HEALTH INVESTIGATIONS

### 2.1 Public health situations in which sampling may be useful

In the subsequent discussion it will be convenient to refer to the whole collection of units from which a sample may be drawn as a population. This term will not necessarily mean a population of persons ; for example, the units may be institutions, records or events. The sample is intended to give results that are representative of the whole population ; one of the main purposes of the theory of sampling methods is to ensure that this is so.

There are many instances in public health work of situations in which results of acceptable accuracy may be obtained by sampling methods with substantial savings in cost, time and labour, as compared with what would be required in complete enumeration.

The reasons for choosing a sample survey in any particular situation will usually be one or more of the following :

(*a*) A sample survey may be the only feasible method of collecting the relevant data.

(*b*) The lower cost of a sample survey and the smaller demands on personnel may be critical.

(*c*) The speed with which results become available may be important.

(*d*) It is often possible to collect more reliable observations when the size of the study is limited.

It is not feasible to describe here all possible applications of statistical sampling in public health. Nevertheless, it may be useful to refer to some of the principal types of investigation in which sampling may play an important role.

2.1.1 *Evaluation of the health status of a population*

In many countries, complete information is available about basic characteristics of the population, certain vital events, notifiable diseases and other conditions of special public health interest. More generally, however, information about morbidity, nutritional status, etc. must be collected by special sample surveys. These surveys may cover a wide range of health conditions or they may relate only to one disease or to a group of related diseases. Information may be sought on physical or mental impairments, on bodily measurements or on such physiological measurements as blood pressure or tests of respiratory function.

In some developing countries, basic demographic information may not be available. Sampling methods then provide an invaluable tool for the collection of information about population structure and vital rates.

Extensive medical records exist in many countries, but the cost of a complete central processing of all the records would be prohibitive. Sample surveys of records are a valuable means of extracting essential information, particularly as this restriction in the scale of the investigation permits a critical study of the quality of the information contained in the records. In some countries, medical and other records pertaining to a single individual are the responsibility of several different administrative authorities, and it may not be easy to assemble them. Although centralized record-keeping is to be encouraged, recent developments in methods for linking separately kept records are likely to be useful in such sample surveys.

2.1.2 *Investigation of factors influencing health*

In studying the etiology of disease, it is important to collect information about the occurrence of disease in various sub-groups of the population. For example, it is useful to know how the incidence or prevalence of a particular disease varies with home environment, occupation, nutrition, personal habits or genetic background of the patients. To obtain this information would require the collection of socio-economic as well as medical data. In etiological studies of this type, the emphasis often is placed on the comparison of different sub-groups; for example, a comparison of the incidence of malignant disease in individuals with different personal habits. Socio-economic data may also be needed in studying the effect of ill-health on living conditions.

2.1.3 *Studies in the administration of health services*

In planning the development of health services, it is essential to have reliable information on such features as the need for particular services, their availability, the extent to which they are utilized, the deficiencies from which they suffer, their cost and their manpower requirements. Sample surveys of administrative and medical records may be used to provide more detailed information about health services than would otherwise be available, as, for example, in a study of the characteristics of prescriptions issued by medical practitioners. Similarly, surveys of households can provide useful information about the utilization of health services. Sampling methods may be used also for field investigations in operational research ; for example, in studying appointment systems for patients.

2.1.4 *Evaluation of the effectiveness of health measures*

The effect of changes in therapeutic or prophylactic measures should, whenever possible, be studied by controlled experimentation in which the individuals or other units to which the changes may be applied are allocated at random to alternative treatments. However, this course of action is sometimes impracticable. In this situation, the new measures may be introduced, perhaps in a limited area, and surveys may be carried out to measure characteristics of the population before and after the change, similar observations being made in a control area in which the measures have not been introduced. For example, the progress of a malaria eradication programme may be assessed by measuring the proportions of individuals carrying the malaria parasite before, during, and after the attack phase of the campaign. Such measurements may be obtained conveniently on samples of the whole population. Similarly, the effectiveness of a programme of health education may be assessed from samples of the population before and after the instruction.

2.1.5 *Studies in environmental hygiene*

Sampling plays an important part in the continuous surveillance of standards of hygiene. Frequently, the population to be sampled consists of specimens of food or water, the object of the investigation being to verify that levels of contamination are sufficiently low. This use of sampling is closely related to sampling inspection and quality control in industry and gives rise to certain special considerations that are not discussed in detail in the present report.

2.1.6 *The rapid processing of data*

When information is collected by complete enumeration, the tabulation of all of the raw data may take a long time. It may then be useful to process rapidly a small fraction of the total data so as to obtain useful results quickly. This course is now adopted in many countries for advance tabulations of national census data.

2.1.7 *Checking the quality of information*

Sampling is also useful for checking the quality of data. For example, a small proportion of death certificates may be scrutinized, in conjunction with clinical records and autopsy reports, to investigate the accuracy of the statement of cause of death. In any sample survey, it will usually be profitable to scrutinize carefully a sub-sample of the results for possible errors or deficiencies.

## 2.2 Situations in which sampling may be inadvisable

Although sampling occupies a central role in epidemiological and public health investigations, there are certain situations where sampling is not appropriate. Three such situations are as follows :

(*a*) The relevant information for every member of the population may be required for non-statistical purposes. For example, medical records must be maintained for each patient in a hospital ; records of all births, deaths and marriages are normally required for legal purposes ; in tracing the source of an epidemic or the residual foci of an endemic disease, details of all known cases will be required. In certain situations where complete information of this type is available, it may still be advantageous to carry out certain inquiries by studying a sample of all available records. In other situations, however, it may be better to take advantage of the fact that complete data exist and to analyse the entire collection.

(*b*) Data may be required for certain sub-divisions of the whole population containing few individuals. Sampling may then fail to provide sufficiently precise information. For example, vital- and health-statistical data for very small administrative areas cannot be adequately obtained by sampling ; death rates for certain rare causes of death can be obtained adequately only by complete enumeration.

(*c*) There are occasions when sampling in a community will create a feeling of discrimination. In such situations, it may be better to include in a survey all the inhabitants rather than only a sample of them.

The use of information relating to the whole population can be regarded as sampling 100% of it, and many of the considerations mentioned in this report in relation to sample surveys apply equally well to complete surveys (see, for example, section 5 on non-sampling errors). There is, however, a qualitative difference, in that a complete investigation avoids the need to introduce a mechanism for the choice of the sample. If, therefore, considerations of accuracy lead to a decision that a relatively large sample is required (say, 75% of the population), it would probably be preferable to perform a complete survey instead.

The investigator should bear in mind that sample surveys should always be undertaken with a definite aim in view and with a careful assessment of the benefits to be derived and the costs to be incurred. Such assessment may lead to the view that, in a particular situation, a sample survey would be an unprofitable exercise, or that it should be carried out on a reduced scale.

## 2.3 Some types of sample surveys

Sample surveys may be used to provide information on a wide range of branches of morbidity and public health. Investigations in different subject-matter fields will make use of different sources of information and different forms of data. In any one subject-matter field, considerable diversity of sources and forms of data may be available.

It would be impossible to make an exhaustive list of all the relevant ways in which the sources and forms of data could be categorized, but it may be useful to mention some of the more important features.

### 2.3.1 *Source of data*

(*a*) *Data relating to individual persons.* In some surveys, the health data are supplied by the individual himself without direct contact with an interviewer. Self-administered questionnaires of this type may be used in postal inquiries; they are also used in some countries when patients attend for certain medical services, and records so obtained may form the basis of sample investigations.

In other surveys, interviews are conducted by trained (but not necessarily medically qualified) staff. Here, as in the use of self-administered questionnaires, the information is usually restricted to the reporting of symptoms and the utilization of health services and may be limited in quality.

In some surveys, the data relating to the individual consist of observations or judgements made by another person. Such data may be derived from medical examinations, either of a general character or for specific conditions.

A large class of sample inquiries makes use of existing medical records. These may relate to patients making use of particular medical services, such as hospitals, health centres, general practice or occupational health services or they may be comprehensive records kept for social security or other purposes.

(*b*) *Data relating to institutions.* In studies of the functioning and administration of health services, it may be desirable to obtain information about the characteristics of particular institutions as well as about individuals receiving medical care. Such information may be available from existing administrative records, or it may have to be obtained by direct approach to the institutions.

2.3.2 *Coverage of data*

(*a*) *Geographical coverage.* Some surveys may involve comparisons between conditions in different countries; for example, international collaborative investigations are an important source of epidemiological information. In such surveys, the responsibility for collection of data in particular countries will normally rest with national agencies, and greater care is needed to ensure uniformity of definitions and procedures. Other sample surveys may be national, regional or local in scale.

(*b*) *Scope.* A sample survey may refer to the whole of a population in a chosen area or it may be confined to specific sub-populations as, for example, people in different occupations or age groups or with specific medical conditions. In a general survey of a human population, it may be necessary to undertake a specially designed survey of certain sub-groups such as nomads.

2.3.3 *Temporal relationships*

(*a*) *Nature of observations.* The health information (conditions of an individual, characteristics of a medical institution, degree of contamination of food, etc.) may relate to a particular instant of time. In morbidity surveys, this would be true of studies to estimate the prevalence of some conditions at a certain point in time. On the other hand, it may relate to events taking place over a period of time. In a single survey of individuals, inquiry may be made about retrospective health experience, or similar information may be obtained from appropriate records. Surveys may involve observations relating to various points in time; information about events occuring after the start of the survey may be obtained either from records (as, for example, by examining death certificates) or by repeated surveys. Such repeated surveys may be either on different individuals on each occasion, or they may be longitudinal, involving repeated observations on the same individuals (see section 3.3).

(*b*) *Repeated surveys.* Certain surveys, particularly national health surveys, may be organized on a continuing basis; that is, surveys of essentially the same type are carried out by the same organization at regular intervals to provide a continuous picture of the health of the population. The samples used in such surveys may be entirely or partially replaced on successive occasions.

## 3. THE PLANNING OF SAMPLE SURVEYS IN PUBLIC HEALTH

Like any other large-scale scientific investigation, a major survey will take time, cost money and pose administrative problems. Even a simple study is wasteful if it fails to produce the desired information. It is therefore extremely important to devote careful thought to the planning of the whole operation before any steps are taken to execute it. There should, at all times, be close collaboration between the statistical advisers and the authorities concerned directly with the subject matter of the survey. This collaboration should start at the earliest stages of planning and continue until the publication of the final report and the completion of the post-survey evaluation.

It is not possible, in the present report, to describe all the relevant considerations, but some of the more important points are noted.

### 3.1 Objectives, conditions and resources

The following should normally be included :

(*a*) Specification of the health problem for which the sample survey is to be designed, and consideration of all available evidence on this topic.

(*b*) Specification of the target population. This is the population about which information is required.

(*c*) The variables and characteristics to be investigated. Due consideration should be given to alternative information that may be available and to its reliability and cost.

(*d*) Methods of classification. For example, in a sample survey to measure the prevalence of a certain condition, it may be desirable to tabulate results in different age- or sex-groups or for populations of different regions, or for combinations of such factors. It is often useful to prepare skeleton tables to specify precisely the results that are being sought.

(*e*) The degree of accuracy required. The term " accuracy " refers to the discrepancy between the true value of some characteristic of the target population and its estimate from the sample. It is affected both by the sampling errors (that is, errors attributable to the choice of one possible sample rather than another) and non-sampling errors (that is, errors that would also be present in a complete investigation ; for example, observational and processing errors). These errors may be due to non-systematic variations, to which the term precision refers, and may include systematic biases as well (see sections 4.5 and 5).

(*f*) The cost that may be incurred and the labour resources needed. Estimates must be made of the cost of all parts of the investigation. In particular, provision must be made in the budget for the cost of statistical tabulation and analysis and, if possible, for special studies referred to in section 7.

(*g*) The provision of procedures for evaluation of technical aspects of the inquiry that may be relevant to the planning of future work (see section 6.2).

(*h*) A critical study of the uses likely to be made of the results.

## 3.2 Specification of the survey design and procedures

The following points should be considered :

(*a*) The types of data to be collected (see section 2.3) and the necessary forms of documentation (record forms, questionnaires, etc.).

(*b*) The unit of inquiry (person, household, medical record, specimen of milk, etc.).

(*c*) The sampling unit and frame. The sampling unit is the basic unit to which a selective procedure may be applied. It may be, but is not necessarily, the same as the unit of inquiry mentioned above ; for example, the unit of inquiry might be the individual person, while the sampling unit was the household. The sampling frame is the documentation that permits the selection of sampling units. This may be a list of sampling units (for example, an electoral list provides a useful sampling frame for adults in some countries) or it may be a map from which sampling units can be constructed.

(*d*) The sample design. This is a specification of the method by which sampling units are chosen and is discussed in considerable detail in section 4.

(*e*) The size of the sample. This question is discussed in section 4.5.

(*f*) The specification of supporting inquiries. These may be pilot surveys, conducted before the main inquiry. Pilot surveys may provide information useful in the choice of sample design for the main survey ; more importantly, perhaps, they provide the opportunity for careful scrutiny of the whole operating procedure. Other possible supporting inquiries are follow-up studies after the main survey and special investigations to check the quality and accuracy of the information collected.

(*g*) The timetable of the whole operation, including both the duration of the study and the particular time of the year to which the information is related (see section 3.3).

(*h*) The source and training of the personnel (see section 3.4).

(*i*) The methods of processing the data.

Many of the points raised in this section and section 3.1 above are closely interrelated. For example, the choice of objectives of a study may be modified by a knowledge of the availability of sampling frames. The point to be emphasized is that time devoted to preliminary considerations of this sort is time well spent ; a sampling inquiry that has been inadequately planned may be worse than no inquiry at all.

The rules of procedure decided upon initially should be incorporated into a written document and made available to the personnel engaged in the study.

### 3.3 Sampling in time

The prevalence of many diseases undergoes both seasonal and secular changes, besides random variation in time. It is essential that these should be taken into account in the design of morbidity surveys.

Most morbidity surveys have as their purpose the determination of some kind of current rate, presumably with an implicit assumption that this is to represent the position over a fairly long period. It follows that, if seasonal changes are believed to be important, the sample must cover the seasons systematically and in such a way that there is no confusion (or interaction) between geographical changes observed as the team moves about the country and true seasonal variation. Further, if changes from year to year are important, the survey must either extend over a long period or be repeated at intervals.

Another type of survey is designed to measure the effect on morbidity of some kind of official action or campaign. Here again, care is needed to eliminate seasonal effects. For example, if the survey is to be repeated at annual intervals, then the field work must be performed at the same period each year.

In studying changes in time, important gains in precision may be achieved by longitudinal surveys in which each individual included is observed on more than one occasion. It will often not be possible to ensure that every individual is observed on every occasion ; even if this were possible, it would not necessarily be desirable, because a sample that gives a satisfactory representation of the population on one occasion will probably become less representative with the lapse of time. A useful device is that of partial replacement, in which a fraction of the sample is replaced on each successive occasion. The replacement may be arranged on a rotational basis, so that individuals re-enter the sample after a certain period.

In a study of growth over an extensive age-range, say over the first 20 years of life, it is unnecessary to extend the actual survey over the same number of years. In this example, it would be possible to observe a sample of children of all ages at a few successive yearly intervals, keeping each child in the study for perhaps two or three years.

### 3.4 The choice of personnel

In planning a sample survey or a continuing programme of surveys, consideration must be given to the selection of personnel. The type of qualification possessed by the personnel must depend on the objectives of the surveys and special characteristics of the information sought. In most countries, there is an acute scarcity of medically qualified personnel. Nevertheless, good use may often be made of investigators who are not medically qualified. Considerable experience is available of the successful use of lay investigators for the performance of technical measurements ; in tuberculosis surveys, for example, such personnel have, after suitable training, been usefully employed for the screening of X-ray films. Similarly, lay interviewers are often used successfully to obtain data on morbidity and absenteeism from work.

Thorough training of the staff is essential, both in basic instruction in the methods by which the observations are made and also in the refinements of technique by which observer error can be reduced. There should be some mechanism for keeping a constant control over the quality of work done, with particular reference to its consistency and accuracy.

The personnel employed in the field operation of a sample survey should be able to communicate easily with the individuals being interviewed or otherwise observed. Similarly, in record-based surveys, the investigators should be familiar with the content and potential deficiencies of the recorded data.

### 3.5 The national organization of sample surveys in public health

In many countries there are permanent organizations for the design and conduct of sample surveys in a variety of subjects. These organizations contain experts in the theory and practice of sampling. Although such organizations are usually mainly concerned with sampling for socio-economic and/or agricultural purposes, the survey organization and the expertise available could profitably be used for morbidity surveys or other public health investigations involving the use of sampling techniques. The possibility of using, for health surveys, frames constructed for other purposes should be explored.

Sample surveys in public health range in scale from continuing national surveys to smaller studies of special diseases, localities, sections of the population or institutional health facilities. Large-scale sample surveys require careful planning and administration, involve extensive resources and call for considerable experience in their conduct. In a country with little previous experience of health surveys, it would be desirable to start on a modest scale. After the necessary experience has been gained, the scope and scale of surveys can be expanded gradually and in keeping with the needs of public health programmes. For measuring morbidity, small-scale surveys are to be recommended for gaining knowledge of the types of information that could be usefully obtained without the aid of large numbers of highly trained personnel or of elaborate laboratory or other expensive diagnostic facilities.

It may be useful for some central national organization to assume responsibility for the co-ordination of sample surveys in morbidity and public health. This responsibility could include the dissemination of general information about recent and current surveys, the co-ordination of information derived from post-survey evaluations of the type described in section 6.2 and perhaps the training of personnel.

### 3.6 Multi-subject surveys

Many sample surveys in the health field have more than one purpose. A distinction may be made between special health surveys, which deal with one aspect of health (although a number of different variables may be observed) and general health surveys, which cover a number of different topics. Both types of health survey are of value.

A third type of survey, which may be termed the multi-subject survey, may attempt to deal simultaneously with questions of health and with matters in quite different fields, such as economics or education. Unfortunately, the differing design-needs of the variety of subjects on which information is collected in some multi-subject surveys sometimes make them unsatisfactory. Nevertheless, they have undoubted advantages in economy of effort, particularly in developing countries where travel is slow and difficult and household-sampling frames may not exist; furthermore, it may be valuable to have information about the same individual in relation to health as well as other factors such as socio-economic conditions.

These advantages should be balanced against the possible disadvantages, such as the difficulty in training investigators, the length of questionnaire that may be required and the slowness of processing results. If individual surveys are carried out for each subject, each survey can be processed as it is completed. The administrative aspects also could cause difficulties, since the subjects covered might be the responsibilities of a number of different departments.

From an analytical viewpoint, the optimum size of the sample may be difficult to determine, since different subjects require different numbers of observations. In such circumstances, it might be useful to collect information on certain subjects from a sub-group of the whole sample (see section 4.5).

### 3.7 Information about the planning of sample surveys in public health

The Committee was particularly anxious that problems encountered in the planning of sample surveys in morbidity and public health, and the solutions adopted to overcome such difficulties, should be made known as widely as possible to workers in these fields. The Committee therefore recommends that WHO take steps to collect and disseminate such information. This effort should cover both the general topics discussed above and the more technical aspects of sample design discussed in section 4.

## 4. TECHNICAL ASPECTS OF SAMPLE DESIGN

### 4.1 The selection of a sample

The sample should be chosen in such a way as to permit satisfactory estimation of the characteristics of the population and, at the same time, determination of the precision of these estimates. This can be accomplished by the proper use of probability sampling, in which each sampling unit has a known probability of being selected. In simple random sampling, each unit has the same probability of being selected, and every possible sample of a certain size is equally likely to be selected. Other forms of probability sampling include random sampling with different (but known) probabilities of selection for different units and the methods of stratification and multi-stage selection described below. All these methods use random selection, which can be regarded as equivalent to the drawing of lots, although the actual operation can be effected more easily by the use of random sampling numbers. Random selection does not ensure that the characteristics of the sample and the population coincide exactly; it does ensure a known probability that their divergence lies within given limits.

An alternative form of probability sampling, which is usually acceptable, is systematic sampling, in which the units selected in any one sample occupy related positions in the sampling frame, and the first unit to be selected from the frame is chosen at random. For example, in a list of households, individuals or areas, one might choose every hundredth entry in the list, choosing the starting point at random. Sometimes the whole sample may comprise a number of independent systematic samples. Except in the last-mentioned method, it is usually difficult to know exactly how precise a systematic sample is. It may be rather more precise than a simple random sample of the same size, because if there is any regular trend in the characteristics of the units as the frame is traversed, this variation is likely to be more faithfully represented in a systematic than in a simple random sample.

Another method occasionally used because of its administrative convenience is purposive selection, whereby the sample is chosen because it is for some reason believed to be typical of the population. Certain characteristics of the sample may deliberately be made to agree with the same characteristics of the population, when these are known. For example, the age and sex distribution of a sample of individuals may be made to agree with those of the population, but, within these quotas, selection is left to the discretion of the investigator. The difficulty with this method is that the sample and population may differ to a large extent in some characteristic that has not been used in the matching. This method provides no real guide to the likely effect of such discrepancies and is not to be recommended.

### 4.2 Stratified sampling

In the selection of a simple random sample, no use is made of any prior information about the distribution of certain relevant characteristics of the sampling units, yet some such information is frequently available. One important device which uses such information is known as stratification. The population is divided into groups or strata according to one or more characteristics, and each stratum is sampled independently with a known sampling fraction (which may be constant or may vary from sample to sample). Stratification may be used with one or more of the following aims : (*a*) an increase in precision (over that of a simple random sample of equivalent size) for the over-all estimates of the population value of the characteristic under study ; (*b*) the spreading out of the sample over the field of the survey ; and (*c*) the provision of reasonably accurate estimates for various subsections of the sampling frame.

Qualitative characteristics, such as geographical areas and occupational groups, as well as numerical variables, such as income, can be used for stratification, though the latter must be categorized before stratification can be effected.

The objectives (*b*) and (*c*) can be achieved by the choice of appropriate sampling fractions in different strata, with perhaps a relatively high sampling fraction in any important sub-group of the population that is too small to be represented adequately in a simple random sample. The increase in precision noted under (*a*) will be achieved in practice with almost any form of stratification, but it will be most marked when the strata are rather homogeneous with respect to the variable under study but vary considerably between themselves.

The two principal difficulties of stratification are, first, that a stratifying characteristic suitable for one aspect of multi-subject survey (see section 3.6) may be almost totally irrelevant to another aspect of the same survey ; and second, that if within-stratum variability is rather large in comparison to between-strata variability, the gain in precision is small and may be outweighed by costs of carrying out and analysing a stratified sample.

### 4.3 Multi-stage sampling

In multi-stage sampling, the sampling frame is first divided into a set of first-stage sampling units, and a sample of such units is selected. Next, each first-stage unit selected is further sub-divided into second-stage sampling units, and so on. Thus, in multi-stage sampling, a complete frame of first-stage units is required, but frames for second-stage sampling are required only for the first-stage units that have been selected. Devices of stratification or variable sampling fraction may be used at any stage, if desired.

Multi-stage sampling has two principal advantages. One is that it permits the available resources to be concentrated on a limited number of sections (e.g., areas) of the frame. This results in a lower cost per unit of inquiry, since some of the overheads of a survey are proportional to the number of investigation centres that must be set up, rather than to the number of units of inquiry obtained. The second is that multi-stage surveys are more economical in situations in which a good sampling frame is difficult to obtain, because the compilation of frames of second-stage units is restricted to sampling units selected at the preceding stage.

On the other hand, multi-stage sampling has two primary disadvantages. One of these is that the sampling error is usually increased, since the variability between ultimate sampling units belonging to the same section of the frame will generally be rather smaller than that between sections. For a given sample size, it would be more informative, but also more costly, to spread the total sample over a greater number of sections of the frame. The second of these disadvantages is that some analytical difficulties may be created by the fact that the sampling units will usually be of unequal size at various stages with respect to the number included in the ultimate sampling units.

In this connexion, a useful device in two-stage sampling is to choose a first-stage unit that has a probability proportional to the number of second-stage units that it contains, then to choose second-stage samples of constant size. This procedure, which is called self-weighting and which leads to simplification at the processing stage, ensures that each second-stage unit has an equal chance of being included in the sample. Similar devices are available for statistical designs with more than two stages.

Multi-stage sampling involves the use of samples that are, to some extent, clustered rather than being spread more or less evenly over the population. For this reason, sampling of this kind is sometimes referred to as cluster sampling. The same term is sometimes used more specifically to refer to any sample design in which the population is divided into groups, a sample of groups is drawn either by single-stage or multi-stage sampling, and the selected groups are investigated exhaustively (complete enumeration). Cluster sampling in this sense has several advantages :

(*a*) It avoids the need for a definition of sampling units, the preparation of a sampling frame and the selection of a sample, within clusters.

(*b*) The cost of field operations per unit of inquiry is reduced. Thus, it may be almost as easy to obtain information for the whole cluster as for a sub-sample. For example, when the units at any one stage are families, it would often not be worth while to carry out further sub-sampling within the family.

(*c*) It may be more acceptable to the population surveyed; for example, a cluster sample of villages avoids the embarrassment of interviewing some persons and not others.

(*d*) It may lead to more accurate observations; for example, in a survey of mortality, the same death is sometimes reported by more than one household in a village. This error can be eliminated more easily if the whole village is covered.

The main disadvantage of this form of cluster sampling is the decrease in the efficiency of the sampling that generally results from clustering. This factor depends on the variability both between and within clusters, which in turn may vary according to the cluster size. If variability and costs are known, an optimum cluster size may be found that will achieve the highest precision for a given cost.

### 4.4 Domains of study

Any subdivision of the sampling frame about which information is to be reported may be termed a domain of study. In etiological investigations, it is often desirable to compare the characteristics of individuals in different domains of study. In a case-control study, for example, the aim may be to compare the characteristics of individuals with and without a certain medical condition in order to throw light on possible associated factors. If the condition is rare, a much higher sampling fraction may be used for the affected than for the non-affected individuals, and stratification may be introduced by pairing an affected and a non-affected individual in a particular stratum (for example, a particular age group and a particular sex group). In other studies, a sample of the population may be sub-divided according to one or more social or personal characteristics, and medical information (for example, the incidence of deaths from certain causes) collected over some ensuing period of time.

Special problems in the design and interpretation of etiological surveys are discussed in a number of publications on epidemiology and will not be considered in detail here.

### 4.5 Determination of sample size

The accuracy with which any characteristic of the population can be estimated from a sample survey will depend on the number of observations in the sample, on the variability between different sampling units, and on the magnitude of non-sampling errors of the type discussed in section 5.

It is therefore impossible to determine the sample size required without some knowledge of, or assumption about, those other factors.

If there are no non-sampling errors, and the variability is known or can be estimated in advance, it will be possible to determine the sample size required in order to reduce the sampling error (and hence increase the precision) to any desired level. It may be important to specify precision, not only for estimation of the whole population, but also for particular domains of study.

An important point is that the variability and precision normally determine the sample size as an absolute number, not as a fraction of the population. Thus, a sample of size 100 from a population of 100 000 (i.e., a 0.1% sample) will provide very nearly the same information about prevalence, for example, as will a sample of the same size from a population of 1000 (i.e., a 10% sample). If no estimate of the variability is available, it may be necessary to conduct a preliminary survey for this purpose. In field surveys, an increase in sample size may, by adversely affecting the quality of the recording, increase the variability between sampling units, and thus yield less advantage than might be expected. It may be that the stated precision could be achieved only with a survey of much too large size. In this case, the investigator must either reconsider his specification of the desired precision or consider other methods of approaching the question that is to be investigated.

Systematic biases do not decrease with an increase in sample size, and they therefore limit the accuracy achievable even with complete enumeration. The importance of taking steps to remove systematic bias is discussed further in section 5.

In multi-stage sampling, the main source of variability is that between the first-stage units, and the most important decision to be taken about sample size is the choice of the number of first-stage units. The number of observations made within each first-stage unit is usually less important, because it is often possible to increase or decrease this number with little effect on the over-all precision and cost.

The accuracy required should depend on the purpose of the inquiry and the decisions that flow from its results. For some purposes it is quite unnecessary to have results of great accuracy.

In some situations, greater accuracy may be required for estimates of some population characteristics than for those of certain others. In such a situation, a useful device is the multi-phase survey, in which some information is collected for all the units selected in the main sample. Further information, for which a smaller sample size would suffice, may then be collected on a sub-sample of the main sample. Various features of the main sample may provide information on which stratification can be used for the selection of the second-phase sample.

## 5. NON-SAMPLING ERRORS

A considerable amount of statistical theory has been developed for the estimation of the likely magnitude of the sampling error, which arises from the fact that a sample is but a part of a population and not the whole of it. A principal consideration in the choice of sample design is to ensure that, for a given cost, the effect of sampling error is as small as possible, or that a required precision is achieved at minimum cost.

At least equally important are non-sampling errors, a term that covers all other ways in which discrepancies arise between the recorded characteristics of the sample and the corresponding characteristics of the population.

Non-sampling errors are not peculiar to sample surveys but occur also with complete enumerations, where errors of this kind may be even more serious. They are important in all forms of sample surveys and inquiries in the public health field, whether these are based on interviews, medical examinations or records. It is just as important to consider these errors and attempt to reduce them to an acceptable level as it is to consider ways of reducing sampling error.

Non-sampling errors can be either systematic or non-systematic. Systematic errors (often called biases) are those that affect the mean value of a variable (or the frequency of an attribute) and do not tend to disappear with an increasing sample size. For example, systematic error might occur in a country where the death rate is estimated by a field survey ; reluctance on the part of the population to report deaths might lead to systematic under-reporting.

Non-systematic errors are frequently associated with technical measurements—for example, in the determination of blood pressure by the usual sphygmomanometer, or the reading of induration sizes in tuberculin surveys. Non-systematic errors can be regarded as contributing to the basic variability of the observations made on different sampling units, and to this extent they merely inflate the sampling error and diminish as the sample size increases. Measurements of this type may also, however, suffer from systematic errors.

Major non-sampling errors can be placed in three broad classifications as shown below.

(*a*) *Coverage errors*, which are caused by failure to sample adequately the target population. This failure may arise through inadequacy of the sampling frame, or it may be due to unsatisfactory coverage of the units selected in the sample. Both imperfections may be caused by the omission or duplication of units to be investigated or by the inclusion of units that should not be investigated. The inability to make the required observations on all the sampling units, which in postal and interview surveys is termed non-response, may be due to non-cooperation, or there may be some other reason why the observation cannot be made (for example, an individual may have died or have left the specified address).

(*b*) *Observational errors*, which may be caused by the fault of the investigator (for example, by inaccurate recording), by imperfect measuring instruments, by the person or object observed (for example, an error of memory in an interview survey or an incorrect statement in a record), or by the interaction of two or more of these factors.

(*c*) *Processing errors* that may arise after the data have been collected. They may be theoretical errors in the method of statistical analysis, clerical errors in the copying of material or computational errors in the data processing.

The control of coverage errors is largely a matter of organization. It may be difficult to eliminate completely defects of the frame, but it may be possible to reduce them to negligible proportions. Non-response will usually be more troublesome. In sampling a human population, it may be useful to acquaint the population with the objectives of the survey, to seek the assistance of influential individuals, to provide guarantees of the confidentiality of information and, in some cases, to provide incentives such as free medical treatment. Even when such steps are taken, there will inevitably remain a group of non-respondents. Some additional measures that may then be taken are described below.

(*a*) Intensive effort may be made to obtain responses from the whole of this group by further approaches of various kinds.

(*b*) Such effort may be confined to a sub-sample of the non-respondents, the information collected from the sub-sample then being applied to the non-response group as a whole.

(*c*) If some information is available about characteristics of the non-respondents (for example, basic demographic features), adjustment may be made for these features in the analysis.

(*d*) Individuals of similar characteristics may be substituted for the non-respondents. This method is similar to the previous one and should be regarded mainly as a device to facilitate data processing.

It should be emphasized that methods (*c*) and (*d*) are never completely satisfactory, since the failure of an individual to provide an observation is likely to be related to the magnitude of that observation or to the presence or absence of an attribute, even when the associated information mentioned in (*c*) is known. These methods are likely to be useful only when the non-response rate is very low.

Losses are particularly liable to occur in longitudinal studies. Even though most of the intended observations are made at each point in time, it may be that few individuals provide a complete set of observations throughout the entire period of the survey. Furthermore, it is likely that the original sample will become progressively depleted with the passage of time. Under certain assumptions, special analytical methods are available for the estimation of trends in these situations, but there will be a serious selective effect if the persons who leave the survey are atypical of the population in respect of some relevant characteristic.

The report of a survey should give a complete description of whatever is known about the character of the non-respondents, such as age and sex distributions, likely reasons for the non-response, and the method of treating the non-response group in the analysis.

Several attempts have been made in recent investigations to evaluate the magnitude of observational errors. In surveys of the interview or postal inquiry type, the replies obtained from a sub-sample may be verified by reference to physicians, hospitals or insurance companies. In record-based surveys, the quality of diagnosis in medical records may be poor and other information may be defective. Here again, a check may be made on a sub-sample of records.

Most sample inquiries make use of more than one investigator (for example, many interviewers participate in an interview survey). Systematic differences between investigators in the general level of observations may be an important source of non-sampling error. This difficulty will not usually be eliminated, but it may be reduced by increasing the number of observers. Investigator variation should be reduced as much as possible by preliminary training (see section 3.4), but the remaining variation should be measured. A useful device is the use of interpenetrating samples, a technique whereby independent sub-samples are allotted to different investigators. In some surveys, it may be possible for certain observations to be made in replicate by more than one observer.

An important source of non-sampling error in health surveys is that attributable to subjective assessment of the severity of a disease. It may be advantageous to supplement such assessments by recording the more objective and reproducible observations (for example, presence of symptoms and signs) on which they are based.

Processing errors can be controlled mainly by suitable administrative action. Adequate clerical personnel, provision of mechanical punching and sorting devices, calculating machines and, in some cases, computers will help to eliminate these errors. Their importance should not, however, be overlooked, as expensive and time-consuming surveys have sometimes been spoilt by inadequate provision for controlling these errors.

The Committee wishes to stress the importance of non-sampling errors and urges that more empirical and theoretical studies should be devoted to this topic. The Committee recommends that WHO should collect and study all available information on the nature of non-sampling errors encountered in health and morbidity surveys and the ways in which they can be reduced. This information, like that on problems of planning and design referred to in section 3.7, should be made available to workers in these fields.

## 6. THE ANALYSIS, EVALUATION AND REPORTING OF RESULTS OF SAMPLE SURVEYS

### 6.1 Estimation

Estimation is the process of generalization from the sample to the entire population. In its simplest form, when one is estimating a total such as that of the number of individuals in a population who show certain symptoms, estimation may involve merely multiplying the survey results by the ratio of the population size to the sample size. This procedure is known as raising or *inflating*. Thus, if one unit in 100 was selected, the raising is performed by multiplying the total in the sample by 100.

In practice, there are two ways by which this procedure may be made more subtle and more efficient:

(*a*) The measure of " size " in terms of which the frame and the sample are compared need not be the number of sampling units. For example, if one area-unit in 100 is sampled, the number of people in the sample will not be exactly 1% of the number in the population. The survey results may be raised by the ratio of the population to the number in the sample, rather than by the factor of 100. In general, the proper use of supplementary information relating the sample to the population will improve the precision of the estimate. Various methods, such as ratio estimation, regression estimation, and difference estimation are described in textbooks on sampling.

(*b*) The raising factor may be different in different parts or strata of the population. Even where a uniform sampling fraction is used in all strata, the raising factor may take into account the actual *achieved* ratio of population to sample in each stratum. This procedure may be regarded as stratification after sampling or re-weighting the sample to match the population. Adjustment of this type can often be done at little added cost, particularly when a computer is used for the analysis (see section 7).

It sometimes happens that a sampling method is used that is intended to achieve a certain sampling fraction but fails to do so because of defective sampling-frame information. It is important to recognize this source of error and to evaluate it; if it is found to be substantial, the estimation procedure must take account of the true sampling fraction rather than the planned one.

For example, in some surveys carried out in developing countries, a first-stage area sample has been selected with probability proportional to census population, and at the second stage a fixed number of households or persons has been selected in each selected area unit. The assumption is that this procedure will provide a self-weighting sample (with a constant raising factor in all areas). In practice, this assumption is often vitiated by the large discrepancy (lack of proportionality) between the census population of each area and the number of households or persons found at the time of the survey.

### 6.2 Post-survey evaluation

Progress in survey design may be greatly enhanced by means of post-survey evaluation. This term is used here to denote a methodological appraisal of the survey design on the basis of data and experience collected in the course of the survey itself. Such a critical appraisal may provide valuable information that can be used in the planning and conduct of future surveys, and it should be allowed for in the initial planning stages.

There are two important ways in which a post-survey evaluation may provide information which will be useful in the planning of future surveys :

(*a*) In any sample design, it is possible to make efficient use of *a priori* estimates of the variances, costs and resources involved, such information being provided by evaluation of previous surveys. An example is the use of " optimum allocation " in stratified sampling. In the absence of *a priori* estimates of the variances and the costs of observing a unit in the various strata, it may become necessary to allocate the sample in proportion to the sizes of the strata. Such an allocation may be far from the optimum if either the variances or the costs in the various strata differ greatly.

In such a case, even rough estimates may make it possible to achieve considerable improvements in sampling efficiency. Analogous examples may be given with reference to all other sampling design procedures. Mention may be made of multi-stage procedures involving the use of clusters as sampling units ; the availability of even very rough estimates of some suitable measures of the homogeneity of the units forming a single cluster will make possible more efficient designs than are otherwise feasible.

When these estimates of variances, costs and resources have been obtained, it may be illuminating to consider the relative costs that would have been incurred and resources that would have been utilized by different sample designs that might have been used, and also the relative accuracy of estimates of population characteristics made by various methods.

(*b*) The appraisal should cover other aspects of the survey as actually carried out. A far as possible, the appraisal should aim to assign causes to any deficiencies observed, so that they can be removed in the execution of subsequent surveys. This type of evaluation may be applied, for example, to:

(i) The time-schedule of the survey; the appraisal should involve a comparison of the original time-schedule and that actually realized, together with an analysis of the reasons for any substantial discrepancies.

(ii) The non-response rate encountered in the survey; the appraisal should involve the level of non-response in various segments of the sample.

Post-survey evaluation can often be carried out at little extra cost and will normally be a very profitable exercise. Specific allowances for this evaluation should be made in the planning of the survey.

### 6.3 Preparation of the report of a sample survey

In addition to the description of the basic results of a sample survey, it is important that the report should contain information about its purpose, planning and organization.

Detailed recommendations on these aspects of a report on a sample survey are contained in a recent document of the Statistical Office of the United Nations. The 13 following recommendations are based largely upon this publication.[1]

(*a*) *Statement of purposes of the survey.* A general indication should be given of the purposes of the survey, the accuracy required and the ways in which it is expected that the results will be used.

(*b*) *Description of the intended coverage.* An exact description should be given of the target population.

(*c*) *Collection of information.* The nature of the information collected should be reported in considerable detail, including a statement of items of information collected but not reported upon. The inclusion of copies of the questionnaire or other schedules, and relevant parts of the instructions used in the survey (including special rules for coding and classifying) is often of value. Methods of checking the accuracy of data at the point of collection should be mentioned. It is of importance to describe the type and number of investigators, for example, whether full-time or part-time, permanent or temporary, with particulars of their training and qualifications.

(*d*) *Specification of the frame.* A detailed account of the specification of the frame should be given. This account should define the geographic areas and categories of material included and the date and source of the

---

[1] United Nations (1964) *Recommendations for the preparation of sample survey reports* (provisional issue), New York (Statistical Papers, Series C., No. 1, Rev. 2).

frame. If the frame has been amended or constructed *ab initio*, the method of amendment or construction should be described. Particulars should be given of any known or suspected deficiencies.

(*e*) *Design of the survey*. The sampling design should be carefully specified, including details such as types of sampling unit, sampling fractions, particulars of stratification and use of multiple stages. The procedure used in selecting sampling units should be described. If it is not random selection, the reporter should indicate the evidence on which he relies for adopting an alternative procedure.

(*f*) *Repetition*. It is important to state whether the survey is an isolated one or is one of a series of similar surveys. Where the survey is repetitive and some of the sampling units reappear in the successive stages, this should be stated.

(*g*) *Date and duration*. There are two periods of time that are important for any survey: first, the period to which data refer, or *reference period*, and second, the *period of collection*; that is, the period taken for the field work. The duration of both of these periods should be stated.

(*h*) *Numerical results*. A general indication should be given of the methods followed in the derivation of the numerical results. Particulars should be given of methods of weighting and of any use made of supplementary information. Any special methods of allowing for non-response should be described.

(*i*) *Accuracy*. A general indication of the accuracy attained should be given, and a distinction should be made between sampling errors and non-sampling errors.

(*j*) *Cost and resources*. An indication should be given of the number of personnel of various categories employed in the survey and the total time spent by each category. Where appropriate, the cost of the survey should be given, under such headings as preliminary work, field investigations and analysis. Resources used in the conduct of the survey but not included in the costs should be stated. An account should be given of the organization of the personnel employed in collecting, processing and tabulating the primary data, together with information regarding their previous training and experience. Arrangements for training, inspection and supervision of the staff should be explained. A brief mention of any equipment used is frequently of value to readers of the report.

(*k*) *Post-survey evaluation*. The extent to which the purposes of the survey were fulfilled should be assessed. In particular, any discrepancy between the actual population sample and the target population should be mentioned. Measures of cost and accuracy relevant to the planning of future surveys should be given.

(*l*) *Responsibility.* The names of the organizations sponsoring and conducting the survey should be stated.

(*m*) *References.* References should be given to any available reports or papers relating to the survey.

## 7. THE APPLICATION OF AUTOMATIC DATA PROCESSING SYSTEMS

The traditional method of processing large-scale sample-survey data is to transfer the information to punched cards and to use conventional sorting and tabulating equipment to obtain the required tabulations. In recent years, there has been a rapidly increasing use of high-speed computers for the automatic processing of data. This development provides the opportunity for cheaper, more rapid and more efficient analysis of data. However, the availability of automatic data processing in no way removes the need for careful planning of all stages of a survey, and in particular does not diminish the importance of collecting accurate and relevant observations.

### 7.1 General considerations

(*a*) The entire use of the computer should be planned as thoroughly as possible in advance of the survey. Inadequate preparation for the use of a computer is likely to be very costly.

(*b*) The potential economy in the use of a computer is likely to be vitiated if complex programs must be written separately for the processing of the results of each survey. For this reason, the use of general survey programs is of great importance. The Committee recommends that WHO should collect and disseminate information about such general programs and about other programs of particular interest for survey analysis.

(*c*) Sample survey data must be carefully edited before statistical tabulation and analysis. The identification of observations that fail to satisfy pre-determined requirements can be done efficiently by a computer, and editing processes of this sort should be regarded as an essential part of the use of automatic data processing for survey analysis.

(*d*) The output from the computer should be kept within reasonable bounds ; otherwise much time may be wasted in the examination of unnecessary tables.

(*e*) If a list of the population is stored in a computer, possibly together with auxiliary information for stratification, the process of selecting a sample may be done automatically by the computer.

### 7.2 Data input

At the present time, the most usual forms of input are punched cards or punched paper tape, which must be prepared from the original data sheets by a separate operation. Errors are likely to occur during this transfer stage and must be guarded against. In systems using paper tape, equipment may be used that provides a rapidly printed record of the information as it is put on to tape.

Increasing use is now being made of improved mark-sensing devices by which information is marked in a special way on the record form and then read directly into the computer. These methods have already proved valuable in many types of medical survey but the care required for the marking of the original record may make them unsuitable in some situations. Other methods of optical scanning, now in an experimental stage, may be widely used in the future.

### 7.3 Data analysis

Reference has already been made to the important function of the computer in editing data. The use of a computer makes possible efficient methods of statistical analysis (such as the estimation procedures referred to in section 6.1) which might otherwise be regarded as too complex. It also facilitates the analytical device of " imputation ", whereby values are automatically inserted for missing or apparently faulty observations, these values being estimated from associated data on the individual units of inquiry. When imputation is carried out, the method used and the proportion of data thus treated should be stated in the report. However, a large amount of computer imputation is quite unacceptable in any statistical investigations. Computers are very suitable for such operations as editing and correction, but the Committee wishes to emphasize that adequate original measurement is as important today as ever.

The use of a computer may permit types of statistical analysis that throw valuable light on the inter-relations of variables. In analytical surveys in which the effect of different factors on certain observed variables is of interest, a complex analysis may be required to disentangle the effects of several related factors. General computer programs are available for analyses of this sort. Similarly, when many variables (for example, several symptoms, signs and test results related to a particular disease) are observed in a survey, methods of multivariate analysis may provide a useful way of combining the information. Such methods of analysis are often prohibitively lengthy without the use of a computer.

## 8. A MANUAL ON SAMPLING METHODS IN MORBIDITY SURVEYS AND PUBLIC HEALTH INVESTIGATIONS

In view of the important role played by sampling in many types of public health investigation and the shortage of experts in the theory and practice of sampling, the Committee felt that health administrators, epidemiologists and other health workers should be provided with facilities for obtaining a basic knowledge of sampling principles and methods and some acquaintance with their potential applications in the medical field. Although a number of excellent textbooks are available on both the theoretical and practical aspects of sampling, few, if any, are directed especially to the health worker. A manual dealing with the general principles of sampling and describing in some detail the special problems and opportunities in the medical field would be a useful guide for many workers in public health, would assist the statistician or sampling expert with no previous experience of medical applications, and would also prove valuable for training courses.

The Committee felt that it would be appropriate for WHO to undertake the task of producing such a manual. It had the opportunity to discuss a preliminary draft produced for the secretariat by a consultant and made various detailed suggestions for changes that might profitably be made in a later draft.

# WORLD HEALTH ORGANIZATION

## TECHNICAL REPORT SERIES

*Recent reports :*

| No. | | Price s.d. | Price $ | Price Sw. fr. |
|---|---|---|---|---|
| 288 | (1964) **Enteric Infections**<br>Report of a WHO Expert Committee (35 pages) . . . . | 5/– | 1.00 | 3.— |
| 289 | (1964) **Joint FAO/WHO Expert Committee on Brucellosis**<br>Fourth report (65 pages) . . . . . . . . . . . . . . . | 6/8 | 1.25 | 4.— |
| 290 | (1964) **WHO Expert Committee on Tuberculosis**<br>Eighth report (24 pages) . . . . . . . . . . . . . . . | 3/6 | 0.60 | 2.— |
| 291 | (1964) **WHO Expert Committee on Malaria**<br>Eleventh report (46 pages) . . . . . . . . . . . . . . . | 5/– | 1.00 | 3.— |
| 292 | (1964) **Environmental Change and Resulting Impacts on Health**<br>Report of a WHO Expert Committee (23 pages) . . . . | 3/6 | 0.60 | 2.— |
| 293 | (1964) **WHO Expert Committee on Biological Standardization**<br>Seventeenth report (86 pages) . . . . . . . . . . . . . | 6/8 | 1.25 | 4.— |
| 294 | (1965) **Integration of Mass Campaigns against Specific Diseases into General Health Services**<br>Report of a WHO Study Group (25 pages) . . . . . . . | 3/6 | 0.60 | 2.— |
| 295 | (1965) **Viruses and Cancer**<br>Report of a WHO Scientific Group (60 pages) . . . . . | 5/– | 1.00 | 3.— |
| 296 | (1965) **Resistance of Malaria Parasites to Drugs**<br>Report of a WHO Scientific Group (65 pages) . . . . . | 6/8 | 1.25 | 4.— |
| 297 | (1965) **Environmental Health Aspects of Metropolitan Planning and Development**<br>Report of a WHO Expert Committee (66 pages) . . . . | 6/8 | 1.25 | 4.— |
| 298 | (1965) **Organization of Dental Public Health Services**<br>Report of a WHO Expert Committee on Dental Health (44 pages) . . . . . . . . . . . . . . . . . . . . . . | 5/– | 1.00 | 3.— |
| 299 | (1965) **WHO Expert Committee on Bilharziasis**<br>Third report (56 pages) . . . . . . . . . . . . . . . . | 5/– | 1.00 | 3.— |
| 300 | (1965) **The Effects of Labour on the Foetus and the Newborn**<br>Report of a WHO Scientific Group (32 pages) . . . . . | 3/6 | 0.60 | 2.— |
| 301 | (1965) **Protein Requirements**<br>Report of a Joint FAO/WHO Expert Group (71 pages) . | 6/8 | 1.25 | 4.— |
| 302 | (1965) **Nutrition in Pregnancy and Lactation**<br>Report of a WHO Expert Committee (54 pages) . . . . | 5/– | 1.00 | 3.— |
| 303 | (1965) **Mechanism of Action of Sex Hormones and Analogous Substances**<br>Report of a WHO Scientific Group (24 pages) . . . . . | 3/6 | 0.60 | 2.— |
| 304 | (1965) **Neuroendocrinology and Reproduction in the Human**<br>Report of a WHO Scientific Group (19 pages) . . . . . | 3/6 | 0.60 | 2.— |
| 305 | (1965) **Physiology of Lactation**<br>Report of a WHO Scientific Group (22 pages) . . . . . | 3/6 | 0.60 | 2.— |
| 306 | (1965) **Public Health and the Medical Use of Ionizing Radiation**<br>Fifth report of the WHO Expert Committee on Radiation (41 pages) . . . . . . . . . . . . . . . . . . . . . . | 5/– | 1.00 | 3.— |
| 307 | (1965) **WHO Expert Committee on Specifications for Pharmaceutical Preparations**<br>Twenty-first report (31 pages) . . . . . . . . . . . . . | 3/6 | 0.60 | 2.— |
| 308 | (1965) **Health Problems of Adolescence**<br>Report of a WHO Expert Committee (28 pages) . . . . | 3/6 | 0.60 | 2.— |

| No. | | Price s.d. | Price $ | Price Sw. fr. |
|---|---|---|---|---|
| 309 | (1965) **Specifications for the Identity and Purity of Food Additives and their Toxicological Evaluation: Food Colours and some Antimicrobials and Antioxidants** Eighth report of the Joint FAO/WHO Expert Committee on Food Additives (25 pages) | 3/6 | 0.60 | 2.— |
| 310 | (1965) **Diabetes Mellitus** Report of a WHO Expert Committee (44 pages) | 5/– | 1.00 | 3.— |
| 311 | (1965) **Special Courses for National Staff with Higher Administrative Responsibilities in the Health Services** Report of a WHO Study Group (31 pages) | 3/6 | 0.60 | 2.— |
| 312 | (1965) **WHO Expert Committee on Dependence-Producing Drugs** Fourteenth report (16 pages) | 3/6 | 0.60 | 2.— |
| 313 | (1965) **The Biochemistry and Microbiology of the Female and Male Genital Tracts** Report of a WHO Scientific Group (15 pages) | 3/6 | 0.60 | 2.— |
| 314 | (1965) **Nutrition and Infection** Report of a WHO Expert Committee (30 pages) | 3/6 | 0.60 | 2.— |
| 315 | (1965) **Immunology and Parasitic Diseases** Report of a WHO Expert Committee (64 pages) | 5/– | 1.00 | 3.— |
| 316 | (1966) **The Technical Basis for Legislation on Irradiated Food** Report of a Joint FAO/IAEA/WHO Expert Committee | 5/– | 1.00 | 3.— |
| 317 | (1966) **Chemotherapy of Bilharziasis** Report of a WHO Scientific Group (71 pages) | 6/8 | 1.25 | 4.— |
| 318 | (1966) **Water Pollution Control** Report of a WHO Expert Committee (32 pages) | 3/6 | 0.60 | 2.— |
| 319 | (1966) **WHO Expert Committee on Leprosy** Third report (31 pages) | 3/6 | 0.60 | 2.— |
| 320 | (1966) **University Health Services** Fourteenth report of the WHO Expert Committee on Professional and Technical Education of Medical and Auxiliary Personnel (21 pages) | 3/6 | 0.60 | 2.— |
| 321 | (1966) **WHO Expert Committee on Rabies** Fifth report (38 pages) | 5/– | 1.00 | 3.— |
| 322 | (1966) **Cancer Treatment** Report of a WHO Expert Committee (55 pages) | 5/– | 1.00 | 3.— |
| 323 | (1966) **Requirements for Biological Substances – Revised 1965** Report of a WHO Expert Group (71 pages) | 6/8 | 1.25 | 4.— |
| 324 | (1966) **WHO Expert Committee on Malaria** Twelfth report (48 pages) | 5/– | 1.00 | 3.— |
| 325 | (1966) **Human Viral and Rickettsial Vaccines** Report of a WHO Scientific Group (79 pages) | 6/8 | 1.25 | 4.— |
| 326 | (1966) **Clinical Aspects of Oral Gestogens** Report of a WHO Scientific Group (24 pages) | 3/6 | 0.60 | 2.— |
| 327 | (1966) **The Use of Human Immunoglobulin** Report of a WHO Expert Committee (29 pages) | 3/6 | 0.60 | 2.— |
| 328 | (1966) **Planning of Radiotherapy Facilities** Report of a Joint IAEA/WHO Meeting (44 pages) | 5/– | 1.00 | 3.— |
| 329 | (1966) **WHO Expert Committee on Biological Standardization** Eighteenth report (130 pages) | 10/– | 2.00 | 6.— |

# WHO PUBLICATIONS

## SUBSCRIPTIONS AND PRICES 1966

### Global Subscription

The global subscription covers all WHO publications, i.e., the combined subscription " C " and, in addition, the *Weekly Epidemiological Record, Monograph Series* and any other occasional publications.

£32 $120.00 Sw. fr. 375.—

### Combined Subscriptions

Special prices are offered for combined subscriptions to certain publications, as follows :

| *Subscription* | | | | |
|---|---|---|---|---|
| A | *Bulletin, Chronicle, Technical Report Series,* and *Public Health Papers* | £10 | $38.00 | Sw. fr. 120.— |
| B | *Epidemiological and Vital Statistics Report* and *World Health Statistics Annual* | £8 10*s*. | $30.00 | Sw. fr. 100.— |
| C | *Bulletin, Chronicle, Technical Report Series, Public Health Papers, Official Records, International Digest of Health Legislation, Epidemiological and Vital Statistics Report, World Health Statistics Annual,* and *World Health* | £23 | $85.00 | Sw. fr. 270.— |

WHO Distribution and Sales Unit will be pleased to make a quotation for any other type of combined subscription desired.

### Individual Subscriptions

| | | | |
|---|---|---|---|
| *Bulletin*, vol. 34 and 35 (12 numbers) | £6 5*s*. | $23.00 | Sw fr. 75.— |
| *Chronicle*, vol. 20 (12 numbers) | 17/6 | $ 3.00 | Sw. fr. 10.— |
| *International Digest of Health Legislation*, vol. 17 (4 numbers) | £2 10*s*. | $ 9.00 | Sw. fr. 30.— |
| *Technical Report Series* | £2 10*s*. | $ 9.00 | Sw. fr. 30.— |
| *Official Records* | £3 15*s*. | $14.00 | Sw. fr. 45.— |
| *Epidemiological and Vital Statistics Report*, vol. 19 (12 numbers) | £4 | $15.00 | Sw. fr. 48.— |
| *Weekly Epidemiological Record*, 41st year (52 numbers) | £3 15*s*. | $14.00 | Sw. fr. 45.— |
| *Vaccination Certificate Requirements for International Travel* | 13/4 | $ 2.75 | Sw. fr. 8.— |
| *World Health*, vol. 19 | £1 | $ 5.00 | Sw. fr. 12.— |

Subscriptions can be obtained from WHO sales agents for the calendar year only *(January to December)*.

* * *

*Specimen numbers of periodicals and a catalogue will be sent free of charge on request.*