NATIONAL HOUSEHOLD SURVEY CAPABILITY PROGRAMME

Non-sampling Errors in Household Surveys: Sources, Assessment and Control

UNITED NATIONS

DP/UN/INT-81-041/2

NATIONAL HOUSEHOLD SURVEY CAPABILITY PROGRAMME

Non-sampling Errors in Household Surveys: Sources, Assessment and Control

Preliminary version

UNITED NATIONS
DEPARTMENT OF TECHNICAL
CO-OPERATION FOR DEVELOPMENT
and
STATISTICAL OFFICE

New York, 1982

PREFACE

This study on assessment and control of non-sampling errors is one of a series of studies designed to assist countries in planning and implementing household surveys in the context of the National Household Survey Capability Programme. The United Nations revised Handbook of Household Surveys is the basic document in the series. It provides technical information and guidance at a relatively general level to national statistical organizations charged with conducting household survey programmes. In addition to the Handbook, a number of studies are being undertaken to provide reviews of issues and procedures in specific areas of household survey methodology and operations. The major emphasis of this series is that of continuing programmes of household surveys.

The content, design and arrangements of household survey programmes will differ from country to country, reflecting varying needs, circumstances, experiences and availability of resources. In studying the options available to them, countries will choose those which make the best use of their resources to meet their specific needs. The objective of these studies is to examine the factors involved in these choices, and to promote good practices in the design and implementation of household survey operations.

Errors in survey results may be divided into two broad types: sampling errors and non-sampling errors. In many situations, the contribution of non-sampling error to the total error exceeds the contribution of the sampling error. The present study provides a broad review, with numerous illustrations, of major sources and types of non-sampling errors in household surveys so as to enhance the awareness among survey practitioners of factors which determine the quality of data they produce. The study also aims to provide a practical guide to the measurement and control of non-sampling errors, and to indicate strategies for building up the quality of statistical data in the long run. The form of the study is non-technical as far as possible, and the emphasis is on illustrations from a variety of experiences.

Basic standards and guidelines for the planning, design, organization and conduct of household surveys are contained in Part One of the revised Handbook, and much of the material (especially Chapter VIII) is relevant to the control of non-sampling errors. Issues relating to quality and operational control of survey data processing are further discussed in the NHSCP study Survey of Data Processing: A Review of Issues and Procedures. Issues of survey design and data quality in different subject matter areas have been discussed in Part Two of the Handbook and in various other publications of the United Nations and its specialized agencies.

In the preparation of this document, the United Nations was assisted by Mr. S.S. Zarkovich and Mr. T.B. Jabine serving as consultants. The document was reviewed at a Technical Meeting on the NHSCP held in April 1981 at New York, and subsequently revised in light of discussions at the meeting. It is being issued in draft form to obtain comments and feedback, from as many readers as possible, prior to its publication in final form.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

A. Objectives, audience and scope

1. Objectives and context

This document is one of a series designed to provide technical support to countries participating in the United Nations National Household Survey Capability Programme (NHSCP). It addresses issues relating to assessment and control of errors in the context of new opportunities and approaches offered by continuity and integration in household survey operations. Its first objective is to provide a broad review of major sources and types of non-sampling errors in household surveys, so as to enhance the awareness among survey professionals of the factors which determine the quality of data they produce. Its second objective is to provide a practical guide to the measurement and control of these errors, and to indicate strategies for building up the quality of statistical data in the long run. The form of the study is non-technical as far as possible. The emphasis is on illustrations from a variety of experiences.

The context of this study is defined by the objectives and scope of the NHSCP. The Programme is designed "to help interested developing countries obtain, through household surveys and in conjunction with data from censuses and other sources, a continuing flow of integrated statistics for their development plans, policies and programmes, and in line with their own priorities. For this purpose, the NHSCP aims to assist the interested countries to develop enduring national instruments and skills for survey-taking" (United Nations, 1980b). Being a country-oriented programme, the NHSCP does not propagate any fixed model of surveys. The complexity and extent of the data collection programme will differ from country to country, depending upon specific data needs and potentialities.

However, an essential feature of all NHSCP country programmes is continuity and integration of household survey activities. Such continuity

and integration provides new opportunities for the improvement and evaluation of the data quality. The establishment of permanent teams of professional, semi-professional and other staff engaged in survey work makes it possible for the organization to undertake a systematic collection of experience on all aspects of survey design and implementation. This would include empirical information on the suitability and clarity of concepts, definitions, classifications, wording of questions, on fieldwork conditions, etc. At the same time, in a continuing programme, quantitative information relevant for more efficient and cost-effective design of surveys can be accumulated and fed from one survey into the design of the next. The development and use of common survey-taking facilities permits the spreading of costs over a number of surveys, making it possible to use improved procedures which, for cost as well as technical reasons, are not available when individual surveys are undertaken in isolation and on an ad hoc basis. For example, the preparation and maintenance of accurate sampling frames is often one of the most expensive elements in sample survey operations, specially in developing countries. It is rarely feasible to devote the substantial resources required for this purpose if the frame is to be used only for a single or a few ad hoc surveys. However, in the context of a continuing and planned series of household surveys, it becomes cost-justifiable to devote considerable resources to the development of adequate sampling frames and to the measurement and control of coverage errors. The presence of regular and permanent enumerators in sample areas permits more thorough and cost-effective follow-up or call-back strategies, and hence greater control of non-response. Similarly, the cost of training of field and office staff can be shared between surveys; furthermore, continuity of operations facilitates building into the primary data collection activities special studies for evaluation and assessment of errors at substantially reduced costs.

At the same time, it bears emphasizing that to reap the benefits of a continuing programme the survey organization must avoid repeating past methods and procedures in routine fashion. It must be alert to

changes that may be called for as a result of changing circumstances and engage in a continuing process of evaluation and assessment of methods, procedures and data quality.

The gradual building up of the quality of the data is a long-term task which has to be seen as an integral part of the day-to-day operations. Adequate resources must be allocated and organizational arrangements set up for continuing work on the evaluation and improvement of data quality, as a part of the programme designed to meet long-term as well as immediate data needs.

2. Audience

This document is intended for a broad spectrum of survey practitioners: designers and managers of household surveys and other professional staff working on survey programmes in national statistical organizations in developing countries. A precondition to the building up of the quality of survey data is the development of an understanding of the nature and sources of errors and an assessment of their magnitude and effect on survey results. This understanding is needed not only by the technicians involved in the survey work, but also by the managers and organizers responsible for resource allocation. It is in this context that the present document has been prepared.

Beyond this primary audience, it is expected that the document will also be of interest to household survey managers and technicians in developed countries and to users of data from household surveys in all countries.

3. Scope

Household surveys are sample surveys conducted to obtain data on various subjects for households and for the families and individuals of which they are comprised. This document will be concerned mostly with household surveys in which the data are obtained by face-to-face interviews with one or more members of each sample household. The other two principal methods of data collection -- mail and telephone interview -- are, for obvious reasons, not yet used very much in developing countries.

In a broad sense, this document is about the quality of survey data. Quality of survey results is determined by three elements -- relevance, timeliness and accuracy. For statistics to be useful and actually used, none of these three elements can be neglected. The scope, timing, interrelationship and content of surveys in a programme have to be determined such that the relevance to users as well as timeliness and accuracy of the resulting data are maximized for given cost, or the cost of achieving the requisite quality are minimized. This process, sometimes referred to as "total survey design", is concerned with the broader issues of survey development - assessment of user needs, determination of scope and content of surveys, their arrangement in the survey programme, operational planning etc. -, as well as with more specific issues of optimization of survey design in terms of cost, accuracy and balance between different sources of error.

Within the framework of overall data quality, this document focusses on certain aspects of accuracy of household survey data. The accuracy of survey estimates depends on the sampling and non-sampling errors of the estimates that result from the survey. Sampling errors in survey estimates occur because data are collected only for a sample of the target population; therefore, the estimates may differ from the values that would have been obtained from a complete census, conducted under the same general conditions. The magnitude of sampling errors is a function of the sample design and estimation procedures used.

Non-sampling errors, with which this document is concerned, may be defined initially as a residual category, that is, as all errors of estimation that are not the result of sampling. "Residual" does not necessarily mean "small" or "unimportant". For many estimates in many surveys, the contribution of non-sampling error to total error exceeds the contribution of sampling error. Unlike sampling errors, non-sampling errors have many different sources, can occur at any stage of a survey, and are often exceedingly hard to detect and control. A good understanding of the sources of non-sampling errors and of the methods available for measuring and controlling them is, therefore, essential for the successful design and execution of household surveys.

B. The significance of non-sampling errors in survey design

1. Optimal allocation of resources

Designing a household survey calls for numerous decisions as to the procedures, instruments and personnel that will be used to collect and process the data. Many of these decisions are critical to the control of non-sampling error. Does the sampling frame need updating? How long a reference period should be used? What respondent rules are acceptable? How many follow-up visits should be made when no one is at home on the first visit to a sample dwelling unit? What kinds of manual and computer edit checks should be applied to the responses obtained? Decisions on these and other similar matters affect the allocation of resources. In the face of such numerous decisions, as well as many practical constraints and the absence of sufficient information, optimization of design and procedures can at best be only an approximate one. Ideally, it requires detailed quantitative estimates of the costs and errors associated with alternate designs. Such information is seldom available with the accuracy and detail needed to make an analysis of all feasible design alternatives.

There are a number of reasons why no overall models have been developed for the solution of the most general survey design problem, namely, the problem of balance between different sources of errors:

(i) There are many practical constraints. The survey designer does not have all courses of action open to him. For example, the availability and quality of sampling materials, field staff, data processing facilities, etc. has a determining effect on the choice of the survey design. The process of survey design begins with a process of identifying, in the light of practical constraints, a set of feasible approaches. The time constraints of the survey may inject another set of considerations very much related to the balance between different sources of error.

(ii) Major alternatives of survey design do not necessarily present themselves as feasible alternatives within a fixed budget. Different approaches may involve costs and require facilities which differ in order of magnitude or at least differ greatly. The initial cost of setting-up a new operation or procedure may be quite different from its recurrent cost as an ongoing process. Furthermore, a single technique may affect several components of the error, in complex ways often impossible to see in advance.

(iii) Almost all household surveys are "multi-purpose", that is they are designed to estimate numerous separate statistics on a variety of topics. The best allocation of resources may differ substantially for different estimates and classes of estimates, so that some compromise is necessary in the overall design.

(iv) Just as a major survey is seldom designed to measure a single variable, similarly it is seldom designed to measure variables at a single level of aggregation. The compromise which is necessary between the often conflicting requirements of national and sub-national requirements may have important repercussions on the survey design. Similarly, there are conflicts between the need to accumulate data over a period to get the best estimates of levels, and to measure differences and time trends.

In spite of these and other limitations, the underlying principle of total survey design, more specifically optimization in terms of cost, accuracy and balance between different sources of error, is of great value in making many difficult decisions that face the survey designer. Often the problem of improving survey design presents itself in the form of the need to concentrate additional resources where they would do the most good. As Fellegi and Sunter (1974) note, "being able 'to do the most good', however, presupposes that we know where the most good

may be done, i.e. which sources of error are potentially the most important under different alternative designs and how to reduce them". It is necessary to identify the components of error which are potentially the most important in the sense that they are amenable to the largest reductions. It is in this sense that the development and implementation of continuing household survey programmes offers new and cost-effective opportunities for improving the total survey design.

2. Evaluation in the context of continuing survey operations

Many additional factors are involved when the design of a number of surveys in a programme is considered simultaneously. Indeed, while a given strategy may not be cost-justifiable for a single survey, it may become so in the context of a survey programme. For example, while the cost of establishing an efficient sampling frame of well-defined, small and uniformly sized area units may be prohibitive in the context of a single household survey, such a frame may become feasible when considered in the context of a series of surveys conducted as part of a national survey programme.

Similarly, the costs of monitoring and evaluation of survey operations and results appear in a different light when considered in the context of a survey programme. Monitoring or quality control (i.e. activities designed to control errors *during* the operation) is essential for any survey -- whether *ad hoc* or part of a programme. However, in a continuing programme, the cost of monitoring can be substantially lower due to the establishment of regular procedures and facilities; and at the same time, its usefulness is enhanced because it provides information for better management of future surveys.

The contrast is even more marked in relation to the cost of *post hoc* evaluation. Apart from evaluation of survey results for substantive reasons, the expenditure on evaluation is in a sense "wasted" so far as the results of a current survey are considered in isolation. By contrast, evaluation becomes an extremely important aspect of continuing survey operations where the objective is to gradually improve survey

design and data quality and to continuously strive to achieve "more with less". Such evaluation will be more effective if the recording or capture of relevant information has been built into the survey procedures; and if the evaluation is carried out soon after completion of each survey, while the key participants are still available and the experience is fresh in their minds. Application of the principle of total survey design to household surveys in a continuing programme will be greatly aided if, as noted earlier, the survey organization adopts a policy of accumulating and cataloguing detailed information relating to errors of all types and to the costs of each phase of survey operations. In addition to its utility for improving the design of new surveys, such information can be used by the survey organization to make a realistic appraisal of the quality of results already produced and to caution users as to their limitations.

3. Some references on total survey design

There is beginning to be some literature on total survey design which the reader may usefully consult. Fellegi and Sunter (1974), referred to above, give a lucid discussion of both the advantages and limitations of its application, followed by several illustrations based on experiences in Canadian censuses and surveys. Other applications have been described by Andersen, et al (1979), and Kalsbeek and Lessler (1978) for surveys of health and medical care expenditures; by Jabine and Tepping (1973) for the collection of data on occupation and industry; and by Huang, et al (1980) for surveys of registration and voting. Some specific models for total survey design have been discussed by Lessler (1974), Horvitz and Wolter (1977), and Wolter and Pyne (1978). Rao and Sastri (1974) provide a general description specifically within the context of comprehensive system of household surveys.

C. Organization of this Document

Chapter II introduces the basic concepts associated with non-sampling error. As explained in that chapter, survey errors are of

numerous types and arise from a variety of sources. They may be classified in a number of ways. One way, particularly convenient for their description, is to categorize errors according to the main stage of the survey operation in which they arise. This scheme is followed in Chapters III to V to describe three major categories of errors, namely coverage errors, error due to non-response and response errors, which arise at the data collection stage. The material presented for each category of error includes basic concepts and definitions, principal sources, examples, effects, and estimation and control of error. Where relevant, related issues of survey design, content and conceptual errors are also discussed under these headings. Some discussion of the control of errors arising at the data processing stage is contained in the NHSCP study on data processing (United Nations, 1982).

Chapters VI and VII deal with the control of non-sampling errors. Chapter VI identifies steps that can be taken at each stage of a survey to control and minimize non-sampling errors, starting with the survey design and planning stage, and following on through questionnaire development, data collection and data processing. Emphasis is on low-cost procedures which experience has shown to be effective.

Chapter VII places this discussion more explicitly within the context of continuing, integrated programme of household surveys. It discusses what can be done to assess, control and reduce non-sampling errors in this context. The importance of accumulating information pertinent to data quality is emphasized, and long-run strategies for building up data quality are presented.

D. Additional sources of information

This document is an introduction to the subject of non-sampling error. Many survey professionals will want to know more about theory and practice in the measurement and control of non-sampling error. The methods used must be carefully adapted to the survey environment in each country; nevertheless, it is helpful to review the experience of others and to learn what has worked well and what has not, and why.

Numerous illustrations of different types of non-sampling error in household surveys have been included in this document, and references are provided to the books, articles and reports from which these were taken. These are listed in Part 2 of Annex A. Because it may be difficult for any individual or survey organization to obtain all or most of the publications listed, a short annotated list of basic references is provided in Part 1 of Annex A.

Obtaining access to and keeping abreast of the relevant literature is difficult and expensive. In addition, a major problem for survey professionals in developing countries is that most of the literature on applications comes from developed countries where survey conditions are quite different from those of developing countries. A special effort has been made to locate and include examples and references from developing countries in this document, but for some topics this was not possible. This gap highlights the need for all countries to document and catalogue their experiences in conducting household surveys. Doing this will help individual country organizations to improve the quality of household survey results over time, and will be especially beneficial if the information can be shared with counterparts in countries where the survey environment is similar.

A catalogue of problems and survey methodology should cover useful experiences from any survey as to the adequacy or otherwise of the conceptual framework, definitions and classifications, units of sampling and enumeration, questionnaire design, question wording and translation, pretesting of questionnaires and procedures, and sample implementation, and should give special attention to difficulties encountered in field work and data processing. For this, it is necessary to ensure that all field and office staff prepare structured reports on their individual experiences, and that these reports be systematically compiled and summarized.

In this connexion, one may note the well-known efforts of the United Nations Statistical Office and the Food and Agriculture Organization in relation to population and agriculture censuses, as well as

of the work of Regional Commissions and other specialized agencies of the United Nations in the area of censuses and surveys.

Another illustration of organized collection can be found in the World Fertility Survey programme. The WFS organized a systematic effort to collect experiences from the countries which undertook fertility surveys under the programme. Much of this information has been published, including guidelines for the preparation of survey reports (WFS, 1977), preparation of reports on problems encountered in some of the countries (Sahib et al, 1975; Ramírez et al, 1976); a study of language problems (Ware, 1977); investigations of sampling error (Kish et al, 1976; Verma et al, 1980) and of response variability (O'Muircheartaigh et al, 1979, 1980, MacDonald et al, 1978) in a number of countries; a large number of studies on data evaluation; and finally, organized collection of design and operational features of country surveys including detailed reports on "technical monitoring" (Scott and Singh, 1980).

Another example is the Survey Methodology Information System (SMIS) maintained by the United States Bureau of Census till recently. The system consists of a collection of articles and reports covering most aspects of survey design and methodology, with special emphasis on measurement and control of errors that occur in the data collection phase. Most of the articles and reports cover surveys in the United States.

A separate Section on Survey Research Methods of the American Statistical Association was established in 1978, and since then separate Proceedings of the papers presented at annual meetings under the auspices of the Section have been published each year. These Proceedings, which are relatively inexpensive, are perhaps the best single source of information on recent developments in survey methodology, at least with respect to government-sponsored surveys. Although a majority of the articles are based on survey research in the United States, many are of broad general interest; and articles based on survey work in developing countries are frequently included.

CHAPTER II

BASIC CONCEPTS

A. The quality of survey data

Before starting a detailed discussion of non-sampling errors in household surveys, it will be useful to consider a broader concept, namely, the quality of survey data. The quality of survey data depends on three characteristics of the data: their relevance to the needs of users, their timeliness, and their accuracy. None of these three characteristics can be ignored in planning and conducting surveys: data are useful only if they are sufficiently relevant, timely and accurate.

1. Relevance

The determination of what data the users need and in what detail is an essential first step in survey planning. This defines the objectives of the survey as far as its content and scope are concerned. In practice, prospective users usually express their needs in fairly general terms and the subject-matter specialists of the statistical office must work out, in co-operation with the users and the survey organization, the specification of the items to be collected and their concepts and definitions. A continuing programme renders this process more effective as it is conducive to the opening up of regular channels of communication and personal contacts with user government departments and subject-matter specialists.

The survey organization will need to be on its guard against promising more than can reasonably be expected from a household survey, given resource and other constraints. From the sampling point of view, the resources for household surveys are seldom adequate to produce data for a large number of small areas - say below the province or state level. Even to produce data for all provinces or states may not always be feasible.

The ultimate objective of any statistical survey is to produce estimates for specified characteristics, applicable to a specified population at a given time. The concepts or characteristics have to be operationalized into sets of questions and response categories. If the operationalization is not done carefully, the survey may not measure what it is intended to measure. Consider, for example, a survey with the objective of measuring annual income and expenditures in a country where the population is predominantly engaged in subsistence agriculture and animal husbandry. If the survey questionnaire leaves out some major sources of income or expenditures; if the duration of the survey is not sufficient to cover significant seasonal variations; or if important sections of the population are excluded, then the survey will not measure what it is intended to measure.

In this document, it is generally assumed that fundamental errors due to serious incompatibility between survey objectives and survey design and procedures are avoided, as are purely accidental catastrophes of the kind that sometimes hamper the implementation and destroy the validity of particular surveys. In other words, it is assumed that the surveys are designed to be basically in accordance with clearly defined survey objectives, and are implemented in more or less normal circumstances. Designing to survey objectives is not, however, a simple task, and readers are urged to consult the revised Handbook of Household Surveys (United Nations, 1981) for guidance.

2. Timeliness

It is possible to imagine situations in which the results of a survey would be of little interest to most users if not available by a specified date. Political polls relating to specific elections fall

in this category. Another example would be a survey of income and expenditures of wage and salary workers to provide data needed in connexion with a scheduled revision of the minimum wage.

More often, the utility of survey results falls off gradually with the passage of time following the data collection stage of the survey. The rate at which utility declines over time depends on the content and objectives of the survey. In some developed countries, monthly labour force surveys provide short-term economic indicators, and the results are needed very quickly. In developing countries, the results of labour force surveys are more likely to be used for intermediate and long-range planning; the value of their results for such purposes lasts longer.

Users of survey data often press for timeliness at the expense of accuracy. Delay in the availability of results is very obvious, whereas lack of accuracy, especially when it arises from non-sampling error, is much harder for users to detect. The survey organization must produce timely data to facilitate their actual use, but it also has a responsibility to produce accurate data, and to strike a reasonable balance between these objectives.

3. Accuracy

The objective of a sample survey is to make estimates of certain values for a population, using observations obtained from or for a limited number of units (a sample) of the population. The term population is used in the general sense: it may refer to people, or it may represent all the households, all the farms, or all the shops for which estimates or inferences are required.

The accuracy of a survey estimate is generally taken to mean the closeness of the estimate to an exact or true value. The true value, which is nearly always unknown, is the value that would be obtained if data could be collected and processed, without error, for all of the units of observation in the population. The error of a particular survey estimate is the difference between that

estimate and the true value of the quantity being estimated.

Errors arise from several sources, for example: the sample selected may not be representative, in some sense, of the study population; the survey results may be distorted because of defects in implementation, including failures to obtain information from some of the units selected; the responses obtained from individual units in the survey may depart from the truth; and errors may be committed in the processing and analysis of the data. All these errors are called non-sampling errors. In addition, there are sampling errors which arise from the fact that the observations are confined to a sample of the population rather than the whole population.

The statistical concept of "error", as defined above, differs somewhat from the everyday meaning of the term. According to Webster's dictionary, the term error refers to an "ignorant or imprudent deviation from a code of behaviour ... an intentional deviation from truth or accuracy ... something produced by a mistake ... a deviation from the correct". Error is taken as a synonym of mistake, blunder, slip, lapse, faux pas, etc. In line with these concepts, "sampling error", for example, would imply that something wrong was done in the process of sampling. However, the statistical concept is more specific and does not necessarily involve any subjective value judgement. It refers to the deviation of an estimate from the true value that it is supposed to estimate. Furthermore, most statistical measures of error are indicative of the expected margin of uncertainty rather than of the actual or exact departure of a particular estimate from the true value. Exact departure is seldom known, since in most statistical enquiries the true value is not known.

B. Methods of classifying non-sampling errors

The discussion in the previous paragraphs led to a definition of non-sampling error, which is the subject of this document. A detailed treatment of the sources, measurement and control of non-sampling errors requires that they be further broken down and

categorized in ways that facilitate understanding of their nature. Several schemes for classifying non-sampling errors are possible; none is perfect, but each serves a purpose.

One approach is to classify non-sampling errors by the stage of the survey in which they occur. The three major stages of a survey are:

- Survey design and preparation
- Data collection
- Data processing and analysis

Each of these stages can be subdivided. For example, data processing and analysis could include:

- Manual coding and editing (data preparation).
- Data entry
- Computer editing
- Tabulation
- Dissemination (publication and other forms of release)
- Analysis

This classification is especially useful in discussing the control of non-sampling errors. It is used in Chapter VI, which describes procedures that can be used to avoid or control non-sampling errors at each stage in the planning and execution of a household survey.

A second method of classifying non-sampling errors is on the basis of the source or type of the error. For example, the three major categories of data collection errors, as described in chapters III to V, are:

- Coverage errors
- Non-response errors
- Response errors

Though these three types of error can be thought of as occurring during the data collection phase of the survey, frequently they are the result of poor decisions or choices in the survey design and preparation phase.

Finally, another important and useful approach is to divide non-sampling errors into variable errors and bias. Each source or

type of error can be divided into these two categories; however, the classification has its greatest value in connexion with the discussion of response errors. Similarly, in the data processing phase, this classification is useful in the study of errors in the manual coding and editing and the data entry. The underlying concepts and their implications for the measurement and control of non-sampling errors are discussed in the next section of this chapter.

C. Variable error and bias

Errors in survey estimates depend on the conditions under which the survey is taken. Some of these conditions may be considered as essential to the survey design and operation. Examples of essential survey conditions include the general socio-economic situation in the country, the subject matter and procedures of the survey, the particular estimation procedures used, and the type of interviewers employed. Errors in surveys also result from transient or chance factors such as the particular units selected into the sample, the particular interviewers and coders used, and the conditions under which a particular interview is conducted. One can conceive of the survey being repeated under the same essential conditions; if this were done, each repetition would give different results due to transient or chance factors. Variable error measures the variability between different estimates of the same quantity from the hypothetical repetitions of the survey. The average of all these estimates is their expected value (under the given essential survey conditions). The difference between this expected or average value and the true value for the study population is the bias of the estimate.

1. Application of the concept of variable error and bias to sampling error

The distinction between variable error and bias can be applied to both sampling and non-sampling error. In the case of sampling error, the variable error component is called the sampling variance.

It is a function of the sample design used, including the overall sample size, the type of stratification, the allocation of the sample to strata, the extent of clustering, and the statistical estimation procedure. The sampling variance of an estimate is a measure of how the estimate would be expected to vary over repeated sample selections; fortunately, it can be estimated from the results of a single sample. Generally, sampling variance decreases proportionately with increasing sample size.

Sampling bias results from the statistical estimation procedure used. Frequently it so happens that a slightly biased estimation procedure is preferable to a strictly unbiased procedure because of the substantially smaller sampling variance of the former. For example, ratio estimators which are commonly used in household surveys are actually "biased"; however, in a well-designed sample with good control over cluster sizes, the bias in many situations is negligible.

It may sometimes happen that a biased estimator is used unintentionally, for example, one that does not properly reflect the sample design and the actual probabilities used in selecting the sample. In such cases, the biases may be substantial, and the mean square error, which reflects the combined effect of sampling variance and bias, can be considerably larger than the variance of an unbiased estimate.

Errors in the execution of the sample design, such as unintentional omission of some areas from the sampling frame, incorrect application of a sampling interval, or applying the wrong weights to a particular set of observations, are considered to be non-sampling errors, even though they are related to the sampling process.

2. Application of the concept of variable error and bias to non-sampling error

Sampling variance is the outcome of a stochastic process, namely, the application of a probability selection scheme to select the units to be included in the sample. To establish an analogous concept for

non-sampling variation, it is necessary to assume that the response obtained for a fixed unit of observation in the population is a variable subject to some probability distribution and thus has an expected value and a variance. In other words, as stated earlier, it is necessary to conceive of the possibility of independent repetitions of the data collection and processing operations for a fixed sample, and to recognize that the resulting values for individual units of observation would vary from one repetition to the next.

Using this survey model, which was introduced by Hansen, Hurwitz and Bershad (1961), non-sampling error can be partitioned into non-sampling variance and non-sampling bias. Non-sampling variance measures the variation of observed values for fixed samples in hypothetical repetitions of the survey process carried out under the same essential survey conditions. Non-sampling bias is a measure of the difference between the expected values of these repeated observations and the corresponding true values. In subsequent chapters, the term bias will be understood to mean "non-sampling bias", which, in any case, is the predominant component of the bias in most situations.

As will be described further in chapter V, the same survey model permits decomposition of non-sampling variance and bias to reflect the separate contributions of survey respondents, interviewers, supervisors, coders, etc. It was noted earlier that one way of reducing sampling variance is to increase sample size. The same is true for non-sampling variance; however, the meaning of "sample size" in this context requires careful attention. While the contribution of respondents to non-sampling variance (the simple response variance) can be reduced by increasing the number of units in the sample, this would have no effect on the contributions of interviewers, supervisors and coders unless their numbers were also increased.

3. Some implications for survey design

Variable errors and biases differ in their sources, effects and methods of measurement and control. Sampling errors are largely

variable errors, while biases are an important and often the main component of non-sampling errors.

Variable errors from different sources are, in the aggregate, additive. Biases from different sources can, on the other hand, act in opposite directions. In a retrospective survey, for example, failure to recall some events may be partly offset by incorrect placement into the survey reference period of events which actually are outside it. Biases from the same source may offset each other when one is comparing sub-classes from the same sample, or values of the same variable from different rounds of a survey.

Variable errors and biases differ in their method of measurement. Variable errors can usually be estimated directly from the survey results, although the extent to which all components of non-sampling variance are included will depend on the sample design and the particular method used to estimate the variance. By contrast, the measurement of bias depends almost entirely on comparisons with data external to the survey proper.

The most important distinction between variable error and bias has already been suggested, namely, that biases can generally be reduced only by "doing something more" (e.g., designing a better questionnaire or improving supervision of field work), while variable errors are reduced by "doing more of something" (e.g., increasing sample size).

D. Sampling vs. non-sampling errors

It is sometimes said that non-sampling errors contribute much more to the total error than do sampling errors. Such a sentiment is probably a reaction to the fact that, in designing surveys, sampling statisticians have frequently paid more, or even exclusive, attention to sampling error, and have disregarded non-sampling error. This may be primarily because sampling errors are more easily quantified. However, this sentiment is tantamount to saying that most surveys in practice use samples which are too large in relation to the requirements

for maximizing overall precision. Though, obviously, the appropriate sample size has to be determined in relation to the survey objectives in each particular case, the above statement regarding the relative magnitude of non-sampling and sampling errors is meaningless unless applied to specific survey estimates.

It is true that in a typical household survey with a sample of several thousand units, the sampling error for estimates for the total sample or for other highly aggregated statistics may be very small in relation to the non-sampling error present. However, one should bear in mind that sampling variance increases in proportion to a decrease in sample size, while the same is not true of all components of non-sampling error. Therefore, the contribution of the sampling error at disaggregated levels (for example, for regions or provinces of a country, or for other sub-groups or sub-classes of the population) may be substantial. This consideration can be even more significant when the objective is to compare and contrast different sub-classes.

Nevertheless, it remains true that hitherto many survey designers have not paid adequate attention to the control of non-sampling errors, and it is an objective of this document to attempt to redress the balance. What is at issue is not simply the choice of the appropriate sample size, but the whole question of sampling design and field procedures. Sampling design is one of the central elements of whole survey design, and it is necessary that "practical sampling ... be rooted in the nature and inter-relationship of field operations involved, not only in the implementation of a particular sample design but also in the data collection stage of the survey as a whole. The sampler must constantly remain aware of what can or cannot be expected of the field and office workers actually involved in implementation of that design". (Verma, 1977)

The concept of mean square error, which was introduced earlier in this chapter in connexion with sampling error, can also be applied to the combined effects of sampling and non-sampling errors. Quantitatively, the mean square error of a sample estimate is expressed as

the sum of all variable errors plus the square of the bias. These concepts, as developed in the Hansen, Hurwitz and Bershad survey model, have important implications for total survey design, insofar as it relates to the accuracy of survey results. These implications were described by Jabine and Tepping (1973) as follows:

> "Conceptually, the magnitudes of the various components of the mean square error are functions of the design of the survey. Thus, we may modify the sampling variance by introducing stratification, changing the size or nature of the sampling units at the various stages of sampling, or by other well-known devices. Similarly, we may modify the response variance components and the coding variance components by changing the form or content of the questionnaire, the methods of recruiting, training or supervising the interviewers, the procedures for editing or coding responses, or other aspects of the survey. Each change alters the mean square error of the results and also the cost of the surveys. We have the ultimate goal of seeking those changes which minimize the mean square error attainable with given resources or, alternatively, which minimize the resources required to attain a specified mean square error."

E. Further reading on basic concepts

The discussion in this chapter of some basic concepts relating to non-sampling error and total survey design has deliberately omitted the use of mathematical symbols and formulas. Readers who have had a course in statistics and especially those who have some knowledge of elementary sampling theory will derive a fuller understanding of these concepts and how to apply them by studying some of the references listed below:

1. Hansen, Morris A; Hurwitz, William N. and Max A. Bershad. Measurement errors in censuses and surveys. Bulletin of the International Statistical Institute 38(2): 359-374, 1961. This is the "classic" paper in which the so-called "Census error model" was first presented.
2. Jabine, Thomas B. and Bejamin J. Tepping. Controlling the quality of occupation and industry data. Bulletin of the International Statistical Institute 45(3), 360-89,

1973. The section on "The Model" (pp. 362-64) gives a concise summary of the Census error model and its implications for survey design.

3. Food and Agriculture Organization of the United Nations Quality of Statistical Data, by S. S. Zarkovich. Rome, 1966. Chapter 1, "Some Basic Concepts", gives a more detailed treatment of most of the ideas presented in this chapter.
4. Kish, Leslie. Survey Sampling. New York, Wiley, 1965. The first part of Chapter 13, "Biases and Non-sampling Errors", examines the distinction between bias and variable error and its implications for survey design and interpretation of results.
5. Verma, Vijay. Assessment of errors in household surveys. Bulletin of the International Statistical Institute, vol. 49, 1981, provides a summary of various types of errors and reviews some of the basic issues in the assessment of errors in household surveys in developing countries.

CHAPTER III

COVERAGE ERRORS

A. Basic concepts

The objective of a sample survey is to make inferences about a desired or target population from the observation of units confined to a sample. An objective basis for such inference is provided by using probability sampling, that is, a selection procedure which gives each unit of observation in the population a known, non-zero relative probability of being selected. The selection is done by applying an appropriate randomized procedure to a sampling frame in which all units of the target population are supposed to be represented.

A frame provides, explicitly or implicitly, a list of sampling units as well as other auxiliary information required and specific sets of rules and procedures for sample selection. In household surveys the sampling operation typically involves a number of "stages" - moving say from larger area units to smaller segments and ultimately to dwellings. At each stage, it is necessary to link the units involved by rules of association, for example rules which specify which dwellings are included in any selected segment. Furthermore, often the ultimate sampling units used are not identical to the units of observation (for which the information is actually collected), and it is necessary to link them by using rules of association which specify which units of observation should be included in the sample when a particular sampling unit is selected. To give a simple illustration, if the units of observation are households (groups of people) and the sampling units are dwelling units, an association rule is needed to link dwelling units and households. If members of a single household live in more than one dwelling unit, a more complex association rule may be needed (for further discussion; see Hansen, et al, 1963).

Non-coverage is failure to include some units of observation,

either directly or implicitly (by rules of association) in the operational sampling frame. Since such units have a zero probability of selection into the sample, they are in fact excluded from the survey results.

The effect of non-coverage is obvious in the case of censuses, the main aim of which is to determine the total number of units, or units possessing certain attributes (for example the total population, or the members of a particular ethnic group). Failure to cover all units will result in an undercount of the total population; further, the degree of non-coverage can vary from one group to another (or from one area to another) and can result in serious distortions in the distribution of the groups in the population.

On the other hand, it does not follow that non-coverage of a significant proportion of the units of observation will necessarily result in an equally significant underestimation of certain population totals. For example, the proportion of cultivated area missed in an agricultural census is likely to be smaller than the proportion of farms missed, insofar as the missed farms tend to be the smaller ones.

The above considerations also apply when a sample survey is used to estimate population totals. However, the objective of sample surveys is usually to estimate proportions, means, distributions, etc. rather than totals. (Where totals are estimated from sample surveys, it is usually done indirectly from the estimated means or proportions in combination with some external or auxiliary information.) In such cases, non-coverage will lead to error in the sample results if the missed units differ in characteristics from the units covered. This frequently occurs, as in the example in the previous paragraph. Similarly, the survey results can be distorted if the extent of non-coverage differs between different parts of the population, such as regions, sex and age groups, ethnic groups or socio-economic classes.

Errors of non-coverage should be distinguished from deliberate and explicit exclusion of sections of the population from the defined target population. Survey objectives and practical difficulties

determine such deliberate exclusions. For example, in a national survey, certain areas of the country may be excluded due to extreme remoteness or inaccessibility, political or security problems, or a judgment that the value of the additional information would be too small compared to the costs involved, or simply because the survey objectives are confined to certain segments of the total population (to urban areas, or agricultural households, etc.). Such explicit exclusions from the study population are not errors of non-coverage; however, care should be taken to emphasize in the survey report that the results do not apply to any parts so excluded.

Non-coverage refers to the negative error of failure to include some units in the frame used for sample selection. This is a special, though common, case of errors resulting from defects in the sampling frame. Just as some units may not be represented in the frame (non-coverage), other units may appear in the frame more than once, giving them a larger than intended chance of selection into the sample. This kind of error can be referred to as <u>over-coverage</u> or <u>duplication</u>. Even if the frame is complete and unduplicated, non-coverage or over-coverage can result from use of defective association rules, or from incorrect application of technically correct association rules.

A practical sampling frame not only provides a list from which units can be selected into the sample, but also must provide sufficient information on the basis of which the selected units can be uniquely identified in the field. Failure to do so can also result in distortions in probabilities of selection and in the sample structure generally, with consequences similar to coverage errors.

Coverage errors differ from another type of error of non-observation -- namely non-response (discussed in Chapter IV). Non-response results from failure to obtain observations on some units selected and designated for the sample, due to refusals, failure to locate addresses or to find respondents at home, loss of questionnaires, etc. The extent of non-response can be measured from the sample results by comparing the selected sample with that achieved; by contrast,

the extent of non-coverage can be estimated only by some kind of check external to the survey operation. It is this fact which makes the measurement of coverage error difficult.

The distinction between errors of coverage and other types of non-sampling error is not always clear. If an interview is conducted with a sample household, and the interviewer fails to list and obtain information for all household members, should this be considered coverage error, non-response error, or response error? The answer would depend on which units of observation are affected by the error -- households or individuals -- and the reasons for the error. The category used, however, is of secondary importance, as long as the nature of each type of error is understood and appropriate measures are taken to control it.

B. Sources of coverage error: frames

1. The use of frames in multi-stage household surveys

To appreciate more fully the nature of coverage errors and how they arise, it may be instructive to consider typical sampling frames and selection procedures employed in household surveys in developing countries.

Generally, household surveys employ multi-stage, clustered samples. Clustering is introduced to save costs of travel and supervision and, particularly in developing countries, to save costs of sample frame construction and sample selection. A sampling frame is more than a list of units from which a sample may be selected; besides actual lists the frame includes procedures that can account for all the sampling units without the physical effort to actually list everyone. In a multi-stage area sample, for example, the frame may consist of:

(i) Maps showing the first-stage or primary sampling units (PSUs), and auxiliary information about each PSU, such as its estimated population size;

(ii) Maps showing the second-stage area units within the

selected PSUs (no such maps being required for the PSUs which are not in the sample). Maps will also be required for area units resulting from further stages of selection, if any, till a sample of the lowest-stage or ultimate area units (UAUs) is obtained;

(iii) Appropriate lists for units within the sample UAUs. Here it is necessary to distinguish between sampling units, units for listing and the units of observation, the last being the units for which substantive data are to be collected in the survey. Several options are possible. One may, for example, use "compact cluster" sampling, in which all units of observation associated with each selected UAU are taken into the sample; this approach may not require a separate operation for listing of units. When listing and subsampling are involved, the units listed may or may not be of the same type as the units of observation. For example, one may list and sample dwelling units within each sample UAU, and interview all households associated with each selected dwelling unit. The operations of listing and sampling within area units involve several complex issues, some of which are addressed in later paragraphs.

It is necessary that units at each stage of the sampling frame be (i) exhaustive, (ii) non-overlapping, and (iii) uniquely identifiable in the field. Coverage errors arise because one or more of these conditions are violated. Exhaustive means that all elements in the target population are accounted for. Non-overlapping means that each lower stage unit belongs to one and only one unit at the next higher stage. Unique identification refers to an unambiguous description of each unit, and correspondence between the frame and the actual situation in the field. These points are illustrated by some examples below.

2. Errors associated with area frames

In some countries there exist no maps on the basis of which a frame of PSUs can be constructed. A common practice in such situations is to select localities (say villages) from administratively available lists. Even when such lists are reasonably complete, localities may not be area units and hence may not cover the entire country exhaustively. The more scattered population not clearly associated with any of the listed localities will not be covered by the frame, and the survey results can be biased if the characteristics of this missed population differ from the rest. Conversely, due to population growth or settlement, two localities may merge, leading to possible over-coverage. Similar problems and coverage errors may arise when boundaries of the localities cannot be uniquely identified, or when no account is taken of new localities formed since the original lists were made, or when the lists are incomplete, or have duplications.

When pre-existing frames of appropriate area sampling units are not available, a special operation to map and segment some larger area units is required before the sample can be selected. Coverage error can result if the segmentation is not done properly. For example, problems have been reported in surveys where the field workers responsible for enumeration were also responsible for segmentation: it was found that they tended to bias the selection in favour of the smaller segments, or to make the selected segments smaller after selection. Similar biases have been reported in cases where precise demarcation of the segment boundaries was undertaken only for the selected segments, rather than for all the segments in the frame prior to sample selection.

The main problem of coverage in relation to area units concerns their boundaries. For example, in crop-cutting surveys when small grids are used to mark plots for selecting stalks of small grains, over-coverage typically occurs. Many stalks appear on the borderline, and the field workers, when in doubt, tend to include them. The smaller the plot, the larger the perimeter in relation to its area, and proportionately higher the risk of over-coverage of stalks at the boundary.

Because of this, experts have advocated using larger plots than would be used if minimizing the sampling variance were the only concern (Sukhatme 1947; Zarkovich 1966). This advantage of larger segments is also pertinent to social surveys. Boundary problems are proportionately less serious in larger segments, and better boundaries can be found more often for larger segments than for smaller ones.

3. Errors associated with list frames

More commonly, one thinks of coverage error in terms of imperfections in listings of units such as households or dwellings. In a population census, where the objective is to enumerate all households and all persons in each enumeration district (ED), errors can occur both in the listing of households and of persons within the listed households. Similar problems also exist in sample surveys. In the course of listing, some units may be omitted from the list (under-coverage), some may be listed more than once (over-coverage); units may also be erroneously included in or excluded from the area, especially in the case of households or persons temporarily away from their usual place of residence.

Listing operations require careful planning and supervision, particularly if the area involved is large, as for example in the case of census enumeration districts with several hundred households in each district. The amount of error will, in general, also depend on the physical size and layout of dwelling units in the area. If listers become aware that their work is not being checked, they are less likely to pay adequate attention to the quality of their work. Similarly, the absence of good quality maps showing area boundaries, physical features and other details can result in unsystematic work, and consequently in serious coverage errors. Cartographic work is relatively expensive, and the cost of production of good maps must be spread over a number of surveys.

Coverage errors can be particularly serious if pre-existing lists are used, especially if they are of households rather than dwelling

units. Out-of-date lists fail to account for changes in the units, or for units formed or constructed since the lists were originally made. In most countries, a list of households from a census would not be suitable as a sampling frame for more than a few months following the census. A list of dwelling units or addresses might be useful for a longer period, provided special procedures are developed to deal with newly-constructed units and other changes. The coverage of administrative lists, e.g., lists maintained in connexion with public health programmes, elections or taxes, should be carefully checked before deciding to use them as a sampling frame.

In some household surveys the listing and data collection operations are integrated into a single field operation. In other words, whenever a population unit is identified the interviewer proceeds immediately to the data collection operation for that unit. Such an arrangement is generally possible only when "compact cluster" sampling is used. With such an arrangements, coverage errors due to incorrect identification of area or area boundaries, erroneous inclusion or exclusion of units, etc., can occur as with other arrangements. However, the frame is completely up-to-date and errors due to changes occurring between the listing and interviewing operations are eliminated. On the other hand, no independent lists are available to check the coverage of the sample by the interviewers; this would indicate the need for closer supervision of field work.

The listing operation is an application of a set of rules, concepts, and definitions. The field staff will find it difficult to list accurately if the units to be listed are only vaguely defined. Difficulties with the definition of households as units are well known. A study of this problem notes that "a household by standard definition is taken to mean a group of people normally living together, pooling together their financial resources, and eating from a common cooking pot. This definition is by no means precise. Questions have arisen as to how to consider a person who lived a couple of months in a year with one child and a few months with another and could not claim either

of the households as being his normal household. 'Gang Quarters', where some only eat together and others only stay together, presents another problem. Then there are servants who stay the day and eat in the master's house, but go back to their own home to sleep; and then there are chauffeurs who eat outside but sleep in the master's home". (Palan, 1978)

4. Additional references

Lessler (1980) gives a systematic presentation of errors associated with frames. She identifies 6 types of errors associated with frames:

"(1) Population elements missing from the frame;
(2) Non-population elements included in the frame;
(3) Population elements associated with the frame more than once;
(4) Failure to recognize that the frame units are clusters of elements;
(5) Incorrect auxiliary information; and
(6) Information insufficient to locate target elements".

Hansen, et al (1963) describe problems met in using "imperfect" lists as frames and methods adopted to cope with these problems in surveys conducted by the U.S. Census Bureau.

C. Other sources of coverage error

The previous section was about coverage errors directly associated with the quality of area and list sampling frames used in household sample surveys. This section will discuss coverage errors arising from other sources, especially those that occur in the processes of selecting samples and in identifying and obtaining survey information for the units of observation associated with the selected sampling units.

1. Use of inappropriate sampling frames

The use of fixed area sampling frames assumed in much of the above discussion is not generally applicable for nomadic, semi-nomadic or other highly mobile populations. For such populations, special methods such as sampling through tribal structures or enumeration at water points have to be tried. Whatever the method used, serious under-coverage of such populations can be expected in many circumstances (for a detailed discussion of techniques for enumerating nomads in African censuses and surveys, see United Nations, 1977).

2. Incorrect application of sampling procedures

Strange though it may seem at first, errors in applying selection procedures are classified as non-sampling errors. A review of the definitions given in Chapter I (A.3) will show that this is logical. Because such errors lead to inclusion of the wrong units in the sample, it seems reasonable to treat them as coverage errors.

When sample selection is done in a central location, it is relatively easy to check the results and to eliminate arithmetic errors and errors resulting from failure to follow instructions. Errors are more likely to occur and are more difficult to control when interviewers are asked to select the sample themselves as they go (say to interview at every tenth household they find). It is not unlikely that some of them will tend to favour smaller households to keep their workload small; others, with good intentions, may substitute larger neighbouring households for small households. Similarly, some interviewers may favour more accessible, centrally located units to those at the boundaries of the sample area. Clearly, asking interviewers to select the sample is a bad practice and should be avoided as far as possible.

Selection errors can also result from incorrect specifications by the survey designer, from field workers' ignorance of the correct procedures, from failure to identify the actual units selected, or simply from mistakes, such as enumerating the wrong unit.

Occasionally dishonest interviewers complete questionnaires for

imaginary or "made-up" households and submit these in place of questionnaires based on actual interviews with the sample households assigned to them. In the United States census, this practice has been given the colorful name of "curbstoning". The practice is relatively easy to detect, either by occasional revisits to sample households or by careful review of the questionnaires that are turned in.

3. Rules of association: households and persons

A potential source of coverage error is failure to develop clear rules of association between sampling units and units of observation, and to apply them correctly. In determining which individuals to include as members of sample households, a choice has to be made between de jure and de facto coverage definitions, the first defined in terms of usual residence and the second in terms of presence of the individual in the household at a given point in time. De jure coverage is more suitable for the purpose of linking the survey data with data from other sources at the level of the individual. However, since residents away during the entire interview period cannot be contacted, de facto coverage can be expected to result in lower non-contact. However, this consideration is more relevant to non-response (discussed in the next chapter) than to coverage. A de facto rule may be less successful in covering the transient population. In many descriptions of surveys, a great deal of attention has been paid to relative merits and demerits of the two coverage definitions. However, the problem may not be a significant one in all circumstances, and its relation to coverage error in a particular situation is an empirical question (Verma, 1980 provides a number of illustrations).

Regardless of which of these coverage definitions is chosen, there are likely to be borderline cases where it is difficult to determine which households to include in the sample and which persons to include in each household. Rules are needed to cover persons temporarily away at school or in institutions, and persons with more than one usual place of residence. In some cases, it may not be obvious how many

distinct households exist within a sample dwelling unit.

The survey designer needs to display a sense of proportion in dealing with these issues. It is important that field workers be given a clear set of instructions and adequate training. On the other hand, the instruction and training materials should emphasize the more common situations, rather than being overburdened with discussion of those that occur only rarely.

4. Rules of association: farms and other enterprises

If information on farms and/or non-farm family enterprises is to be collected in connexion with a household survey, rules of association are needed to link these enterprises with area or list sampling units used in the survey. Farms, which may be separated from the residence of the operator and may not be located entirely within the selected ultimate area units (UAU's), pose particularly difficult problems.

If the survey is designed only to produce estimates of total area (farmland, cropland, pastureland, acres harvested of particular crops, etc.), one procedure is to collect information only about land located inside the boundaries of the sample UAU's regardless of who is responsible for the farming activities. This is called the closed segment approach.

However, if data on farm characteristics are needed, then a rule is needed which uniquely associates each farm with a single UAU, even though it may be physically located in more than one UAU. The simplest rule is to associate each farm with the UAU in which the farm operator resides. This method, called the open segment approach, requires clear definitions of "farm" and "farm operator". What procedures should be followed if, for example, farms are operated under partnership or communal arrangements? Typical farming arrangements differ substantially between countries, and the definition and association rules adopted should be designed to fit the local situation.

Rules of association for non-farm family enterprises are generally more straightforward than those adopted for farms. These enterprises are usually associated with a single household and are

physically located in or very near to the dwelling unit in which the household members live. Perhaps the greatest potential for coverage error lies in failing to define non-farm family enterprises in such a way that they can be readily distinguished from other (commercial) non-farm enterprises.

5. Identifying individuals with special characteristics

In some household surveys, coverage is restricted to specified classes of persons, such as women of child-bearing age, disabled persons, or persons participating in specified government programmes. The usual procedure in such surveys is to interview a sample of households and to apply a screening procedure to identify household members who belong to the target population. The amount of coverage error for the target population in interviewed households will depend to a considerable extent on the effectiveness of the particular screening procedure used.

As in all household surveys, coverage error can result from incorrect application of the survey rules that associate individuals with households. In this type of survey, errors can also result from use of imprecise definitions of the attributes (age, disability, etc.) that identify members of the target population, or from incorrect responses to questions about these screening characteristics. Non-coverage of some members of the target population is particularly likely if the screening characteristics are regarded by some respondents as sensitive, for example income and some types of disability.

D. Some examples

1. Examples from censuses

In most surveys, the extent of non-coverage can be estimated only by checks external to the survey procedure -- either from an auxiliary investigation or by comparing survey estimates with data obtained from an independent source. Especially in the first case, the task of checking

is arduous. Reliable estimates of non-coverage are scarce. However, it would be wrong to conclude from the scarcity of information on non-coverage that it is not an important source of non-sampling error in household surveys.

For the reasons noted, much of the available information on coverage error is from statistically advanced countries, and relates to population censuses rather than to sample surveys. Even though the focus of the present study is on household surveys, some illustrations from censuses are relevant because censuses often provide the frames on which sample surveys are based, and also because many of the reasons for coverage error are the same in censuses and surveys.

An extreme case of coverage error was reported for the 1946 census of industrial and business establishments in France (Chevry, 1949). The errors were found to be so serious that it was decided not to publish the census data in order to avoid misleading the users of the data.

Interesting results from a number of recent censuses of agriculture in the United States are given by Chapman and Rogers (1978).

Table 3.1

Estimated per cent net undercoverage in U.S. censuses of agriculture

Item	Census Year				
	1954	1959	1964	1969	1974
Number of farms	8.1	8.4	11.3	15.0	11.5
Land in farms	5.4	6.0	6.1	9.1	7.4
Value of product	na	na	2.9	3.3	2.9

The overall level of (net) under-coverage is relatively high. The higher under-coverage of farms in 1969 was associated with a switch from a field canvass to a mail data collection procedure. The authors state that the lower under-coverage of farms in the 1974 Census was largely the result of a change in the farm definition whereby some of the smallest units were eliminated from the target population. The

table shows that under-coverage was smaller for land area than for number of farms, and considerably smaller for the value of farm products. This is because the farms missed were more often the smaller ones and those less intensively cultivated.

Several studies have been conducted on coverage errors in U.S. population censuses (U.S. Bureau of the Census, 1978). It was estimated from a reinterview study that in the 1950 census, 2.3% of the population was omitted, and 0.9% erroneously included, giving a net under-coverage of 1.4%. The estimated undercount varied by age, race and other characteristics; for example, net under-coverage for whites was 1.2% and for others 3.3%. For the 1960 census the following estimates of coverage error, also based on a reinterview study, have been reported:

Table 3.2

Estimates of under-coverage in the 1960 Census of Population: United States

A.	Under-coverage in EDs		3.2%
	- in missed housing units	1.8%	
	- in partially enumerated housing units	1.4%	
B.	Over-enumeration in EDs		1.3%
	- in over-enumerated housing units	0.1%	
	- within properly included housing units	1.2%	
Net undercount (A - B)			1.9%
Gross Coverage Error (A + B)			4.4%

Considerable variation by race and sex was also found. For non-white males, a net under-coverage of 4.2% was reported, as compared with 1.6% for white males. From a record-check study for the 1970 census, it was estimated that "3.1 per cent of the population was missed in 1970 (1.6% in missed units, 0.9% in enumerated units, and 0.5% in

occupied units enumerated as vacant). In general, these rates were higher for Blacks than for non-Blacks, higher in the South and Northeast than in the West and North Central areas, higher in rural areas than urban (but high in central city portions of SMSA's), and highest for low income and educational attainment levels".

The results cited above for the 1950, 1960 and 1970 U.S. censuses were based on reinterviews or record checks. Estimates of undercoverage for the same censuses based on demographic analyses are generally higher. Table 3.3 compares estimates of net under-coverage in the 1960 Census derived by two different methods.

Table 3.3

Estimates of the per cent of net undercount by race and sex, by source of estimate: U.S. 1960 Census of Population

Source of estimate	White		Non-White	
	Male	Female	Male	Female
Reinterview study	1.6	1.7	4.2	3.4
Demographic analysis	2.4	1.6	9.7	6.3

The differences in Table 3.3 illustrate the difficulty of estimating the size of census under-coverage. Each method used has certain strengths and weaknesses, and the results obtained by different methods are unlikely to be in full agreement. These issues are discussed further in section F.2 of this chapter.

The 1966 Sample Census of Population in England and Wales was the first census in the country conducted using the sampling method, and many steps were taken to check the quality of the operation, including a check on coverage. For the coverage check, two samples were used. First, a sample of plots of about 15-20 households each was selected with the objective to estimate the rate of omission of households. The study indicated that around 1% of the households were omitted (Gray and Gee, 1972).

Secondly, a sub-sample was selected from the sample of households in which persons were re-enumerated. The study indicated a 1.4% net undercount of households, but a 0.6% net overcount of persons.

Studies of census coverage have also been made in many other countries. A sample check following the 1953 Census of Population in Yugoslavia found no undercount; rather, a slight overcount was reported due to double counting of persons (Zarkovich, 1966). The 1956 Demographic Yearbook of the United Nations reports, for example, a net undercount of 1.1% in the 1951 census of population in India, and 0.7% net undercount in the 1953 census in Ceylon (now Sri Lanka). A post-enumeration survey of the 1974 population census of Liberia, using the so-called "dual systems approach", estimated under-coverage to be 11.0 per cent, with the highest under-coverage occurring for males under 25 and females below child-bearing age (Marks and Rumford, 1976). A convenient summary of estimates of census undercount and the methods of estimation used for 24 countries is given by Zitter and McArthur (1980).

2. Examples from surveys

A useful study of non-coverage of dwellings in sample surveys is provided by Kish and Hess (1958). The authors note that of the two possible methods to study non-coverage, a post-enumeration quality check is necessarily expensive, both because the check has to be on a large enough scale to measure errors which are generally rare and likely to be concentrated in a few areas, and because the data-gathering procedure is likely to have a high unit cost. The alternative procedure, used in the study, calls for a reliable estimate from an outside source. One difficult condition that must be met in such a comparison is that the outside estimates must be based on the same units not only by formal definition, but also operationally. The differences between the two definitions must at least be small enough to allow for meaningful comparisons between the two sources. Kish and Hess compare the occupied dwellings in their sample, inflated by the sampling fraction, with the occupied dwellings reported in the census. Occupied dwellings rather

than individual persons were chosen for the comparison, because the coverage of persons in surveys and censuses is not compatible -- since the latter but not the former include persons not living in private households.

The study reports a substantial improvement in coverage with improved procedures: from 90% or lower even after years of effort with one set of procedures, to 97% with improved procedures. The authors attribute this change to improved procedures rather than merely to improved performance of interviewers. One of the major sources of improvement is stated to be the use of a sample of segments averaging about 4 dwelling units each, rather than a sample of individual dwelling units.

Available information about coverage error in the United States monthly labour force survey (the Current Population Survey) has recently been presented and analyzed by Brooks and Bailar (1978). The authors estimate that "because of missed structures less than 3 per cent of the target population is not included in the frame". Persons missed within interviewed households are thought to account for a larger percentage of the total under-coverage. Conservative estimates of total under-coverage are obtained by comparing sample estimates for population sub-groups with intercensal estimates for the same subgroup derived independently of the survey. In 1975, for example, the estimated under-coverage on this basis was 4.9 per cent and 2.3 per cent for white males and females, respectively, and 15.5 per cent and 7.5 per cent for males and females of black and other races. Actual under-coverage is greater, because the independent estimates do not reflect under-coverage in the population census.

Turner et al (1979) give an example of under-coverage which occurred in a 1976 national survey of farm production in the Dominican Republic. A sample of farms for the survey was selected from a listing of all farms for which the residence of the "farm producer" was inside the boundaries of an existing sample of area segments used in a continuing quarterly survey of agricultural area and production. An analytical study of the results led to the conclusion that the survey

estimate of the number of farms was low by from 15 to 22 per cent. This under-coverage was believed to result from the fact that the survey interviewers "... were accustomed to listing land parcels in their regular (quarterly survey) work but not to accounting for farm operations or operators per se".

Sometimes it is proposed to use a population census as a sampling frame for a follow-on survey of some sub-group of the population, for example, persons with certain kinds of disabilities. If the sample for the follow-on survey includes only persons or households reporting disabilities in the census, response error in the census can lead to coverage error in the follow-on survey.

Table 3.4, taken from McNeil (1981), shows some relevant results from a 1980 pretest for a disability survey planned as a follow-on to the United States 1980 Census of Population (the planned survey was not taken). A pilot census had been conducted in the Richmond, Virginia metropolitan area in 1978. The disability survey pretest used a sample of households enumerated in the pilot census two years earlier, stratified by whether or not any disabled persons had been reported at that time. In the disability survey pretest, a detailed series of questions was used to identify and determine the time of onset of various kinds of disabilities. Individual responses to these questions were matched against responses for the same persons to the shorter disability inquiry used in the pilot census.

Table 3.4

Disability status of matched persons aged 18 to 66 as reported in 1978 pilot census and 1980 survey pretest: Richmond, United States (weighted results)

Status in 1978 pilot census	Total	With a work disability that began in 1978 or before	
		Number	Per cent
With a work disability	15,574	9,345	60.0
With no work disability	175,504	10,894	6.2

The important result shown in Table 3.4 is that more than half of the persons identified in the 1980 pretest as having work disabilities starting in 1978 or before were not identified in the 1978 pilot census as having work disabilities. By extension, it can be concluded that if the sample for a follow-on survey included only those persons identified as disabled from a simple screening question in a census, serious under-coverage could occur. Hence, to ensure proper coverage of disabled persons in the follow-on survey it is necessary to cover both (a) persons identified as disabled in the census and (b) persons not identified as disabled. However, the likelihood of finding a disabled person during the follow-on is much higher in group (a) than in group (b) (60 per cent versus 6 per cent), so that it will be efficient to substantially over-sample group (a) for the follow-on survey.

An example of coverage error resulting from incorrect application of the sampling procedure is provided by the quality check following a 1951 census of livestock in Yugoslavia. Supervisors responsible for the quality check were to revisit a random sample of holdings. On the day of the check the weather conditions happened to be very bad: low temperature, rain, mud and wind making movement in the field very difficult. When the check data were matched with the corresponding census data, considerable discrepancies were found and doubt arose as to the quality of the procedure followed in the check. Analysis of the results indicated that field workers in the check did not go out to visit the more distant units; instead they replaced these units by others that were nearer to village centres. However, the substituted units were significantly different from the original ones, resulting in substantial over-estimation of numbers of cattle and poultry in the first enumeration check. This sort of distortion is a common consequence of allowing field workers to freely select the sample "as they go". Further illustrations are provided in Zarkovich (1966).

E. The effects of coverage errors

Under-coverage in a census can have very important effects for a number of reasons. Data from a census of population are often used to determine the number of representatives of various political subdivisions to the national governing body, and sometimes to assign weights to votes of individual representatives in that body. In such cases, the presence of coverage errors in census data may become the basis of very serious legal and political issues. If various minorities have rights that depend upon their numbers as found in the census, unequal incidence of under-coverage in different sub-populations could result in inequitable distribution of resources. Difficulties might also arise if various geographic units make contributions to the government or receive assistance based on their population sizes and per capita incomes. Generally recognized undercounts of some population groups, and differences in the completeness of coverage of various regions, cities and other administrative units have in many cases generated heated public discussion of equity and fairness in the use of government resources and representation of various units in the legislative and decision-making bodies (for an example relating to the 1980 U.S. Census, see Keyfitz, 1979).

The percentage error in counting or estimating the <u>total</u> number of households or other units in the population is usually about equal to the percentage net under-coverage of persons. This is true both for complete censuses and for sample surveys. However some totals, such as the total value of agricultural product in one of the examples given above, may be affected to a different degree depending upon the characteristics of the units missed.

Except possibly for very large-scale sample surveys, the interest in survey results is primarily in estimating proportions, distributions, means, rates and ratios, etc., rather than population totals as such. For example, demographic surveys are usually conducted to estimate birth and death rates, rather than the total number of births and deaths

in the population. The effect of under-coverage on such estimates depends mostly on the extent of differences in the characteristics of the units covered and the units not covered, as for example was illustrated in Table 3.1.

An example of the effects of coverage error on estimates of totals and per cents from the U.S. Current Population Survey is given by Hirschberg, et al (1977). The effect of adjustments for under-coverage on labour force estimates is shown in Table 3.5.

Table 3.5

Labour Force Estimates for March 1973 before and after Coverage Adjustment
(Numbers in millions)

Labour Force Category (16 years and over)	Initial (1)	Published (2)	Adjusted for census under-coverage (3)
Total civilian labour force	84.7	87.3	90.3
Employed	80.3	82.8	85.3
Unemployed	4.5	4.5	5.0
Unemployment rate (per cent)	5.3	5.2	5.6

Estimates in column (1) are obtained by inflating sample results by the reciprocal of selection probabilities. Estimates in column (2), which are the ones actually published, are the results of adjusting initial survey estimates of population by age, sex and race to agree with control totals obtained by updating 1970 Census counts for the same groups, using data on births, deaths and migration. The column (3) estimates result from a further adjustment to take account of estimated under-coverage in the 1970 Census. After analyzing the effects of similar adjustments on several income and labour force items, the authors conclude

that the adjustments "... have their major effect on aggregates with the effects on overall percentages and rates being much less pronounced".

Another common objective of sample surveys is to produce estimates of differences between various sub-groups of the population. For the non-coverage bias to be important in the comparison of sub-classes, the effect of the bias on one of the classes must differ significantly from its effect for the other sub-class. If both sub-classes are affected in the same way and to the same extent, the bias in their difference will cancel out. However, more often than not, the assumption of such an automatic cancellation is quite inappropriate.

The importance of bias resulting from non-coverage should be judged in relation to other errors in the survey. For example, if coverage errors are known to be appreciably smaller than errors due to non-response, it would be appropriate to devote additional resources to reducing non-response rather than to eliminating the already small non-coverage. In the reverse situation (which probably is more common in developing country surveys) of relatively small non-response but large under-coverage, resources should be devoted to reducing the latter. Similarly, in deciding on the sample design and size, possible coverage errors need to be compared with sampling errors, not only for the estimation of population totals, but also of other statistics such as population means, means and differences of sub-classes etc. For example, even in samples of reasonable overall size, the sampling error in estimates of differences between small sub-classes can be quite large; consequently, the relative importance of bias due to coverage errors may be much less for these estimates.

F. Estimation of coverage errors

1. Introduction

As mentioned previously, the estimation of coverage error is difficult and expensive, but nevertheless generally an important task. One of the major advantages of regular programmes of integrated surveys,

as opposed to one-time ad hoc surveys, is the opportunity they offer for the measurement and control of errors, particularly coverage errors. Within the context of integrated survey programmes, it is possible, and even necessary, to develop sampling frames or master samples so that the cost of their construction and maintenance can be distributed over a number of individual surveys. The repeated use of a master sample makes it all the more important that adequate resources be allocated to the measurement and control of coverage errors.

2. Evaluation of coverage errors in censuses

Several methods have been developed to estimate under-coverage in censuses, and some of them are also relevant to household surveys. A description of the nature and relative advantages and disadvantages of these methods follows (for a more detailed account, see Siegel, 1974):

(i) Re-interview studies. These studies can take many forms. One straightforward version is to conduct an independent re-enumeration, following the census, in an area probability sample. The resulting listings of dwelling units, households and persons are matched against those from the census for the same areas. In some studies, there may be additional field work to reconcile and explain differences found in matching. A major advantage of this method is that it identifies separate components of coverage error, for example, duplication, missed dwelling units, and persons missed in enumerated dwelling units. A major disadvantage is that, as shown by experience, persons missed in enumerated dwelling units in the census also stand a good chance of being missed in the reinterview survey in spite of all precautions. Another disadvantage is that matching operations are prone to errors which may significantly affect the re-interview study estimates of coverage error.

(ii) Record checks. In general terms, a record check study

consists of taking a sample list of persons from some source other than the current census, determining the residence of each person at the time of the census, and checking against census records to see whether these persons were enumerated in the census. Source lists may be from a prior census, a household survey, various types of administrative records, or some combination of these.

Omissions from the record check source lists are less likely to be correlated with omissions from the census than are omissions from re-interviews; consequently, underestimation of under-coverage is less likely. However, up-to-date lists covering the entire census target population are difficult to locate, and matching problems may be an even greater source of error than in re-interview studies.

(iii) Analytical techniques. In this approach, data from various sources, such as prior censuses, vital records and records of external migration, are used to develop values for the total population and sub-groups classified by age, sex and race, etc. as of the census date. These expected values are compared with the corresponding census counts.

If reasonably good data are available to derive the expected values, this method may provide the best estimates of under-coverage. There are no matching problems involved. However, the method provides only estimates of net coverage error at the aggregate level and does little to identify components of and reasons for coverage error.

(iv) Comparisons with aggregates from administrative records. For some segments of the population, indications of

coverage error may be obtained by comparisons with aggregated data from various kinds of administrative records: voting lists, population registers, birth records, social security records, etc. Such comparisons do not provide information on components of coverage error, and when differences are observed it may be hard to determine which of the totals is in error.

The above review clearly shows that there is no single best method of estimating coverage error. To obtain best results it is desirable to use more than one method and compare results. In many developing countries, however, the methods available are quite limited. The analytical techniques require the existence of a reasonably complete system of vital registration. This prerequisite is not met in most developing countries. The same is generally true of record checks and uses of aggregate data from administrative sources. Hence most developing countries have to depend upon sample checks involving re-enumeration for evaluating census coverage, even though recognizing the disadvantages of that approach.

3. Evaluation of coverage errors in household surveys

In theory, all of the methods described above are available for estimating coverage errors in household surveys. In practice, the methods which require additional field work -- re-interview and record check studies -- are seldom used for this purpose; survey organizations rely almost entirely on analytical techniques for after-the-fact evaluation of coverage.

An exception is the re-interview programme of the U.S. Current Popualtion Survey. A sub-sample of units is re-interviewed each month and a sample of area segments relisted. Emphasis, however, is on the measurement of response error, and the re-interview programme accounts for only a small proportion of missed persons, as estimated by analytical techniques. An experiment was conducted in October 1966 and June 1967 in which the entire re-interview was devoted to coverage. This

experiment showed that the re-interview programme could provide better estimates of coverage if more attention were given to that aspect; nevertheless, estimates of under-coverage were still well below those obtained by analytical methods (Brooks and Bailar, 1978; U.S. Bureau of the Census, 1968b).

Re-interview and record check studies are done somewhat more often in conjunction with household surveys to estimate response variance and bias, probably reflecting an implicit judgment that these components of non-sampling error affect survey results more than coverage error does. Procedures for re-interview and record check studies are discussed in Chapter VII.B.

Comparison of the data from a household sample survey with other sources such as a population census can usually provide at least a broad indication of the coverage or representativeness of the sample. One may, for example, compare the distribution of the population by region, type of place (urban-rural), and by characteristics such as age, sex and ethnic group from the survey with corresponding data from a recent census. The census counts should be adjusted to make them as nearly comparable as possible with the survey estimates. The principal adjustments normally are those needed to account for differences in reference dates and for differences in the target populations. For example, institutional population may have been covered in the census but not in the survey. Even though effects of all sorts of sampling and non-sampling errors are confounded in such a comparison, the major factor involved is probably the quality of coverage of the sample. It should be a routine practice to make such comparisons whenever alternative sources of data are available.

In many developing countries, probably the only external source available for a more specific evaluation of coverage errors in a sample survey is the population census. If the area units used in the sample are compatible with areas used in census (for example, enumeration areas, or other area units such as individual localities), and if the census data have been tabulated by those areas, then a comparison between the

sample and the census at that level may be possible. Similarly, if the census is reasonably recent and sufficient information exists to identify individual dwellings, households or persons as enumerated in the census, then case by case comparison with the sample survey may be possible. Whether the comparison is at the area level or at the level of individual units, one necessary condition is that the census and survey results that are compared must be based on the same units not only as formally defined, but also operationally. Also, such a comparison will only be meaningful, so far as coverage errors are concerned, if the sampling frame is at least partly independent of the census.

G. Control of coverage errors

No study of coverage errors can compensate for poor work in the original survey. Post-survey evaluation studies can be instrumental in avoiding misinterpretation and misuse of the data, in identifying areas where improvements in methodology are called for, and in formulating measures to improve coverage in future surveys. However, the errors in the original survey remain. Hence, the emphasis in all survey work must be on the prevention of errors.

To control coverage errors, sufficient resources must be devoted to the construction and updating of sampling frames. Given the normal resource constraints, the best way to ensure this is to distribute the costs involved over several surveys. Depending upon circumstances, this can be done in various ways. In many developing countries, the first priority would be to develop a master sample system with suitable area units, sufficient in number to meet the precision requirements of any household sample survey which the country might wish to conduct. Also, the master sample should be designed to permit repeated and inexpensive sub-sampling for individual surveys. With such an integrated system it may be easier to devote more resources and efforts to map the sample areas accurately, obtain auxiliary

information as needed and institute procedures for keeping the area frame up-to-date between censuses.

Many multi-stage designs for household surveys involve preparation of listings of dwelling units or households for selected ultimate area units and, in some cases, selection of samples from these listings. Because individual dwellings or households are likely to be less stable than area units, the time interval between the establishment of lists and their use for survey purposes should be held to a minimum. The preparation of fresh lists for each survey is not always feasible because of the high cost of such an operation. The strategy should be (i) to distribute the cost of listing over a number of surveys, i.e. to plan individual surveys so as to ensure sharing of listings as far as possible, (ii) to avoid making new lists whenever the purpose can be served by a cheaper operation to update existing lists, (iii) to choose a sampling design which avoids unduly large listing work and (iv) to plan the listing operation and procedure so as to minimize the risk of coverage errors and to devote sufficient resources to ensure good quality of listing.

The development of integrated survey programmes, and specifically the maintenance of a master sample of area units would go a long way towards ensuring objectives (i) and (ii). The implication of (iii) is that, generally, the sample design should be determined such that the ratio of the number of units to be listed to the number in the sample for which substantive data are actually collected is not large. This may, for example, call for greater clustering of the sample in situations where the available area units for sampling are large in size, and/or segmentation of such area units before dwellings or households are listed. Where appropriate, compact cluster sampling may be used to avoid a separate listing operation altogether. Further, the survey programme should be designed such that lists prepared for a master sampling frame are more or less exhaustively used by the surveys in the programme. The point is simply to pay due attention, in determining the survey design, to the relatively high costs of

good listing and to potential trade-offs between sampling errors, which can be reduced by the use of listing and sub-sampling, and non-sampling (in this case, coverage) errors, which may be increased as a result of listing errors and changes occurring between the listing and interviewing operations. A good discussion of some of these issues is given in World Fertility Survey (1975a) and Verma (1977, 1980).

Once sufficient care is taken to keep the size of the listing operation within manageable limits, it may be more feasible to devote adequate resources and care to its quality. The choice of listing units is important. Small "chunks" of a few dwelling units each, or even individual structures or dwellings, may be more suitable units for durable lists than individual households. Listers should be trained to proceed with listing in a systematic and specified way, use maps to indicate the itinerary followed during listing, and develop sketch maps for areas where adequate addresses do not exist. Supervisors should do spot checks of the listers' work, especially in the early stages of the operation and for less experienced listers.

Listing and the subsequent use of lists by interviewers, especially in areas lacking well-defined street patterns, is facilitated if identification numbers appear on the outsides of all structures. In some areas, all or most structures will already have such numbers, and they should be recorded by the listers. Where there are no such numbers, it may be possible to arrange for listers to mark or affix identification numbers specifically for the purpose of the survey programme. However, this should only be done after obtaining appropriate clearance with local authorities and giving due consideration to possible adverse effects of this practice on the co-operation of residents in future surveys.

Another determinant of the amount of coverage error is the quality of work done by interviewers. This is especially important with respect to the coverage of persons who are members of the target population within occupied sample listing units. The forms and procedures used by interviewers to list household members should be tested prior

to the survey. As was recommended for listers, supervisors should do spot checks of interviewers' work, with due attention to the completeness of household listings.

CHAPTER IV

ERRORS DUE TO NON-RESPONSE

A. Basic concepts

1. Definition of non-response

Non-response arises when households or other units of observation which have been selected for inclusion in the survey fail to yield all or some of the data that were to be collected. This failure to obtain complete results from all the units selected can arise from several different sources, depending upon the survey situation. For example, in a household survey, the interviewers may fail to locate a selected household, or the respondent may not be at home when the interviewer calls, or the respondent may be unwilling or unable to respond, or the completed questionnaires may be lost.

Non-response may be total or partial. Total non-response refers to the failure to collect any survey data from a sample unit. Partial or item non-response occurs when a co-operating unit fails or refuses to provide some specific items of information. For example, in an income-expenditure survey, some otherwise co-operating respondents may refuse to give information on some sources of income, or on some items of consumption. Sometimes items are omitted because the interviewer fails to ask the question or to record the answer.

Total non-response is relatively easy to define in principle and very noticeable in practice: unlike non-coverage discussed in the previous chapter, non-response can generally be measured by comparing the lists of units selected for the sample with the units for which information is returned from the field. In practice, however, the situation can be more complex. For example, the selected sample may not take the form of lists which can be directly compared with returns from the field; or the units used for listing and sample selection (for example, dwellings) may not be directly comparable to

the units of observation (for example, households or individuals). Even when there are no such difficulties, it may be difficult to differentiate between "blanks" in sample lists (such as vacant dwelling units) and probable cases of non-response (such as units which cannot be located due to inadequate description in the list). The latter case is characterized as "probable" non-response, since there is a possibility that the unit could be vacant.

Partial non-response is also easy to define in principle, but will only be recognizable in practice if the questionnaire is suitably designed and the instructions for completing it are followed by the interviewers. The most common problem occurs in the recording of amounts, such as the amount of income from each of several sources. It is desirable (but often not accomplished in practice) that the recording procedure should make it possible to distinguish between those cases where the respondent had no income in a particular category and those cases where no response was obtained for that category. The latter case represents item non-response; the former does not.

2. Measures of non-response

In most survey reports, the extent of non-response is indicated by reporting its complement, usually called either a response rate or a completion rate. Bailar and Lanphier (1978) define the response rate as "the number of eligible sampled units responding divided by the total number of eligible sampled units". This is the generally accepted definition, although various other measures are used from time to time. The above definition actually gives a ratio or proportion; it may be multiplied by 100 to express the response rate as a per cent.

If ultimate sampling units have been selected with varying probabilities, it may be useful for some purposes to calculate a weighted response rate, in which the appropriate weights are applied to the individual eligible sampled units in the numerator and denominator.

In a household sample survey, response rates are normally computed

for households. To apply the above definition, there must be agreement on what constitutes an "eligible" household and what is meant by a "responding" household. For example, if responses to a few core items for an absent household have been obtained from neighbours, but many items are missing, should this household be classified as responding?

In most cases response rates can be measured well if accurate accounts are kept of the disposition of all eligible elements that fall into the sample. Such records are necessary for understanding the sources of non-response, for its control and reduction, for predicting it in future surveys, and for estimating its possible effects on survey results. Furthermore, reporting the extent of non-response to users has become an accepted practice in household surveys.

The aims noted above can be better served if the many possible sources of non-response are sorted into a few meaningful classes. A good classification of non-response depends upon the survey situation, and the terminology used in the description of non-response depends upon the type of survey. For example, in interview surveys one talks of "not at homes" and "refusals", while in mail surveys these terms may be meaningless. In this chapter, terminology appropriate to personal interview surveys will be used.

Item response rates are generally computed only for households or persons counted as responding to the survey and for whom a response to the specific item should have been obtained. Thus, if a specific item applied only to women aged 14 and over, other persons responding to the survey would not be included in the denominator of the response rate for that item. Item response rates are an indication of the potential effects of non-response bias on the items and point to specific problems that may call for revision of the questionnaire and/or the instructions to interviewers.

When a survey variable, such as household income, is based on responses to several questions, the interpretation of item response rates is less straightforward. Assuming that each household not

responding on one or more of the relevant questions is counted as not responding on income, then the item response rate will depend, in part, on the number of separate questions used and on whether or not the questions are asked separately for each household member. However, a relatively low response rate on income resulting from the use of a large number of separate questions is not necessarily an indication of significant bias resulting from non-response. In fact, the combined effects of non-response and response bias are likely to be smaller if more questions are used.

3. Major issues

In evaluating the significance of non-response for any particular survey, the important questions to be considered are:

- what are the rates of response (expected and actual)?
- do the responding units differ from the non-responding ones, and if so, to what extent?
- what are the sources of non-response; how can it be reduced?
- what can be done to eliminate, or at least to reduce, the effect of non-response on the survey results?
- what part of the survey budget should be allocated specifically to reducing non-response?

With respect to the last of three questions, it should be recognized that the same sources may give rise to various errors in survey results, including non-response. For example, inadequate training and supervision of field staff may result in poor coverage, high non-response, serious response errors, etc. Hence, the "cost" of reducing non-response cannot in general be looked at in isolation.

4. Organization of chapter

Various sources of non-response are discussed in Section B. The relative importance of these factors and the measures needed to control them will vary. In any given situation, the objective of identifying

the major sources of non-response is to devise measures to control it and also to estimate its effects on survey results. Section C provides some illustrations of the level of non-response from surveys in both developed and developing countries. In general, refusals to participate in the survey are an important source of non-response in developed countries, but a relatively minor component of non-response in developing country household surveys. Section D examines the effects of non-response on survey results, and gives some illustrations.

Section E provides general guidelines for improving response rates in household surveys in developing countries. Section F covers procedures used in household surveys to deal with non-response in both the data collection and data processing stages, with special attnetion to imputation of missing data. Finally, Section G identifies some additional sources of information on the subject of non-response in household surveys.

B. Sources of non-response

1. Introduction

Many factors can lead to non-response. The relative importance of different factors and the measures needed to control them vary from country to country, from one culture to another, and from one survey to another. In any situation the objective of identifying the major sources of non-response, and the characteristics of the field staff and respondents associated with non-response, is to devise measures to control non-response, to adjust for it, and to estimate its effects on survey results.

Two major categories on non-response may be identified: non-contact and refusal. By and large, the major problem in most developing countries, specially in rural areas, appears to be non-contact due to difficulties in access to and identification of units, resulting from defects in the sampling frame and other adverse circumstances. In the

more developed countries, the most significant problem may be an increasing incidence of potential respondents' refusal to co-operate with the survey (Verma, 1980). Failure to find an eligible respondent "at home" after repeated attempts is also a significant factor.

In a continuing programme of surveys, it is usually possible and cost-justifiable to devote more attention and resources to improving the sampling frame, and consequently to reduce the incidence of non-contact -- which, as noted above, is often the most significant component of non-response in developing countries. The deployment of permanent enumerators, often stationed within or near sample areas, also helps in the same direction. Enumerators can become familiar with local conditions as well as the respondents, and are able to make the necessary call-backs more economically. On the other hand, greater attention needs to be paid to respondent burden, specially if the same respondents have to participate in a number of different surveys or in a number of rounds of a longitudinal enquiry.

2. Failure to gain access to sample units

A possible source of non-contact, especially in developing country surveys, is the inaccessibility of some of the sample areas due to conditions such as civil disturbances, political or security problems, or floods and other natural calamities. Sometimes, due to unfavourable publicity, most of the respondents in a particular sample area may refuse to co-operate. In such cases, most of the non-response in the survey may occur in a few clearly identified areas. An attempt should be made to anticipate such difficulties during the planning phase, and special arrangements should be made to ensure the coverage of the problem areas as far as possible. For example, rescheduling of the survey, or the use of locally recruited interviewers may reduce the problem. If a problem of some magnitude remains, it may be necessary to redefine the study population by explicitly excluding the inaccessible domains.

In many survey situations, it is necessary to employ locally recruited interviewers who are familiar with and able to work in local conditions and language. It is not always possible to replace field

workers unable to continue their duties by other interviewers or by new recruits in time to complete scheduled interviews. Such situations are similar to the case described in the previous paragraph in the sense that loss of field staff may result in some sample areas becoming "inaccessible" within the survey period and resources.

Defects or deficiencies in the sampling frame are another common source of non-contact, especially in developing country surveys. For example, an inadequate description of the area sample units may result in failures to locate or correctly identify the assigned units. In such situations non-response may be difficult to distinguish from non-coverage. The incidence of such problems depends upon factors such as the sampling procedures used, the quality of sample lists and how up-to-date they are, the mobility of the population, and of course the qualifications, training and supervision of the field staff. It has been a common experience that the living arrangements and mobility of the population in urban areas results in such problems being more common in those areas than in rural areas.

3. Failure to contact respondents

Once the interviewer has located, identified and assessed the sample unit to be interviewed, he or she may still fail to contact a suitable respondent. Success in contacting an appropriate respondent depends upon several factors. First, the type of respondent considered acceptable for the survey is important. Non-contact is likely to be less common in household surveys where any adult member can provide the required information; it is likely to be more common if in each household one or more specified members need to be contacted individually. The nature of the respondent also makes a difference: for example, it is easier to find people who work at home than those who work away from home.

The degree of difficulty in contacting an acceptable respondent depends strongly on whether or not proxy responses are permitted. Proxy interviewing refers to the collection of data for selected persons from some other person, such as another member of the household. With the survey resources and sample fixed, one can expect lower non-response

(but greater response error) if proxy interviews are allowed. This is illustrated by the following data from labour force surveys in Canada (Platek, 1977; see also Fellegi, 1973). Non-response rates are compared for a "Methods Test Panel" (MTP) in which the incidence of proxy interviewing permitted was lower, and the regular Labour Force Survey (LFS) with more proxy interviewing permitted.

Table 4.1

Comparison of non-response rates from the Canadian Labour Force Survey and a Methods Test Panel

Item	Domain A		Domain B		Total	
	MTP	LFS	MTP	LFS	MTP	LFS
Total non-response	11.6	6.9	9.3	3.9	11.0	5.9
-temporarily absent or not at home	6.6	4.4	4.7	3.0	6.1	4.0
-refusals	4.4	2.0	4.5	0.7	4.4	1.5
-others	0.6	0.5	0.2	0.2	0.5	0.4
% of proxy interviews	23.4	52.3	7.1	44.8	19.3	49.7

Domain A corresponds to "self-representing units", which are normally the larger urban areas. In this domain, proxy interviewing was more common and at the same time overall non-response was higher, reflecting the effect of type of place. The data show that proxy interviewing lowered not only the non-response due to temporary absence, but also that due to refusals.

The survey rule concerning the use of proxy-interviewing is an important consideration affecting the cost of the survey. In addition to response rates, it will also affect the accuracy of the data. Further treatment of this question may be found in Haase and Wilson (1972); Singh and Tessier (1975); Kovar and Wilson (1976); and U.S. National Center for Health Statistics (1977).

Another factor influencing success in contacting respondents is

the timing of the survey. For example, during the holiday season, large numbers of people may be temporarily away from their usual place of residence. Agricultural workers may be away during the harvest season. In fact, the timing of field work can be a major design decision affecting not only success or failure in contacting respondents, but also the whole quality and feasibility of the survey operation.

In this connexion, it is useful to distinguish between two categories of non-response: (i) temporarily away, and (ii) not at home. The first refers to respondents who are known to be away during the entire operational survey period, and is influenced by the timing of the survey operation. The second refers to respondents who are not at home when the interviewer calls, excluding cases in (i). The incidence of not-at-home respondents depends upon factors such as the type of respondent involved (for example, daytime during the week is particularly unsuitable for finding employed members of the household), the care with which the interviewer chooses the time of each visit, the availability of information on the respondents' movements and especially the number of call-backs made.

Finally, reference can be made to the problems of contacting respondents in successive rounds of longitudinal surveys, in which information is to be obtained for the same persons or households over several survey rounds. The problem to be faced is that persons and households move between rounds; sometimes within the same sample area, sometimes to another sample area, and sometimes to an area not in the sample. Special procedures must be established to ensure that the proportion of units lost to observation in later rounds of the survey is held to an acceptable level.

4. Failure to gain co-operation

Following a successful contact with respondent, the next step is to obtain the respondent's co-operation to provide the required information. This effort may be completely, partially, or not at all successful, with the second and third possibilities leading to partial and total non-response, respectively.

As indicated earlier, refusals are an increasingly serious problem in social surveys in developed countries, and may be becoming so

in some developing countries, particularly in urban areas. In large-scale and compulsory operations such as a census of population, a publicity campaign through mass-media can be used to make the operation and its timing and obligatory nature known to the public. Such efforts can result in higher response rates. However, most sample surveys are neither compulsory nor have budgets that permit a broad publicity campaign. Nevertheless, publicity campaigns may be useful, if they are focused on the areas selected into the sample. In many areas, preliminary discussions with local authorities are essential as a first step in gaining co-operation from the residents who have been selected for the survey.

Many factors can affect the incidence of refusals, including the auspices of the survey, the reputation of the organization conducting it, the nature of the questions to be asked, the length of the interview, the technique and skill of the interviewers, and the nature of the respondents themselves. On the other hand, people's reaction towards any particular survey may be greatly influenced by more general attitudes, either favourable or unfavourable, which they have developed towards all survey activity on the basis of past experience and information received from other sources. It is, for this reasons, important for any survey organization to conduct its work in such a way that respondents are always treated courteously and with respect, and that no harm comes to them as a result of their having provided information for statistical purposes.

It is hard to generalize on the subject of how refusal rates are affected by the nature of the questions asked in the survey. There is little evidence to suggest that people are any less willing to answer questions on many sensitive topics than they are to answer simple "factual" questions. On the contrary: surveys on fertility and contraception for example have experienced little difficulty in many countries. Of course, refusal rates can be higher if the questions are perceived to be threatening in some way, for example in surveys concerned with what may be regarded as undesirable behaviour (e.g. the use of alcohol

or drugs) or with income and wealth (which may generate fear of taxation or confiscation). Even if such questions do not result in explicit refusals, the quality of the information provided may be adversely affected. Some "refusals" to answer particular questions may simply reflect the respondent's inability to understand the question or to provide the information requested. In such cases it is often difficult to distinguish between an actual refusal and a "don't know" response. The important thing in a particular situation is to anticipate such problems and to pretest the feasibility of the questionnaire before launching a full-scale operation.

Some developed countries have undertaken studies to find out why people refuse to co-operate; this is generally done by conducting limited interviews with non-respondents to specific surveys. One such study was conducted by the Swedish Central Bureau of Statistics (Bergman et al, 1978). Generally speaking, it was found in this survey that people are very much "concerned about privacy and confidentiality, and about being computer registered". Some of them did not believe that the information supplied by them would be treated as confidential. Some interviewees were "suspicious people negative to contacts with strangers". Others were critical of authorities and the community as a whole; therefore, they did not want any contacts. Some older people said that they had "neither the strength nor the energy to receive visitors". Some asked the question "why should it be just me who has to participate? You can surely take someone else". Some respondents said they were negative to the survey because of lack of time to co-operate. Among poor people, cases were found of the attitude that statistics is the tool of the bosses who should not be helped through co-operation with surveys. One person said the survey was "damned unpleasant and absolutely forced on me, nothing less than an intrusion". Some people referred to the survey sponsors as "they": for example "They take surveys in order to conceal unemployment". With this background, it was concluded by the researchers that some level of non-response had to be accepted as being beyond the statisticians' control.

5. Response burden and non-response

Non-response is often associated with the burden that the survey puts on the respondent. The burden may refer to the length of the questionnaire; the strain it places on the respondent's memory; the need for the respondent to search for information or documents, or to keep a diary; and, of course, the subjection of the respondent to repeated interviewing. Respondent burden may affect not only the response rate, but also the quality of the information obtained. This consideration can be a very significant one in the context of integrated programmes of surveys. It may appear attractive, from the point of view of statistical efficiency, to retain the same sample of households or persons in a multi-round longitudinal study, or to subject it to a number of surveys within a short period. However, this may adversely affect both the response rates and the data quality. The effect of response burden on response rates was studied in labour force surveys in Canada (Platek, 1977). In some of these surveys the burden was increased by asking the same households to participate at the same time in other surveys. This led to a higher non-response rate. In the Indian National Sample Survey, a number of subjects may be canvassed simultaneously in any given annual round. However, the different questionnaires are administered to different samples of households located in a common sample of areas. This is done primarily to reduce respondent burden (Rao and Sastry, 1975).

6. Interviewers and response rates

Another issue which has been the subject of several studies is the relationship between interviewers' characteristics, such as age, sex, level of education, experience and race, and response rates. These studies describe specific situations, and broad generalizations are not possible. Reflecting the experience in a number of countries, one report notes that "the (surveys) have successfully used housewives, graduates and nurses as interviewers; similarly, in all categories there have been failures" (Verma, 1979). Lyberg and Rapaport (1979) report that in Sweden no differences attributable to sex of interviewers

were found in the response rates, but younger and older interviewers did better than those in the middle-age ranges, and interviewers with modest education did better than their colleagues with higher education. A study on the Canadian experience emphasized the role of the interviewer, noting that "respondents are persuaded to co-operate in a survey as much by the sincerity and capability of the interviewer as they are by the validity of the arguments he or she presents. In the recruitment of interviewers, it is thus necessary to pay full attention to their abilities and willingness to get along with people as it is to check their intelligence, general competence and health. It is also necessary to keep building their ability to motivate the respondents (to co-operate)". The following table (Gower, 1979) shows how a relatively small proportion of the interviewers can be responsible for a substantial proportion of the non-response. It is important to identify such interviewers, and to retrain or replace them as necessary.

Table 4.2

Distribution of Canadian Labour Force Survey Interviewers by non-response rate

Non-response rate (%)	0	0.1-5.0	5.1-10.0	10.1-15.0	15.1-20.0	over 20	Total
Per cent of interviewers	15.0	40.0	31.3	9.2	2.7	1.2	100.0

An illustration of an experiment designed to study the effects of different interviewer characteristics on response rates is given in Gales and Kendall (1957).

7. Characteristics of non-respondents

It is important to try to identify relevant characteristics of non-respondents. This is necessary to design measures to reduce and adjust for non-response and to estimate its effects on survey results.

In Canadian Labour Force surveys it was found that non-responding households differed from the responding ones in household size and labour force status (Gower, 1979). In a study by Barnes and Birch (n.d.) respondents and non-respondents were compared on a number of characteristics. Substantial differences were found by tenure: owners of houses had a refusal rate of 14.1%, while for local council tenants the rate was 5.3%. In a study by Demaio (1980), older people and people with higher incomes refused more frequently, though no difference was found by race or sex. There have been numerous articles and reports on this subject, including Bennett and Hill (1964); Cobb, King and Chen (1957); Dunn and Hawkes (1966); Lubin, Levit and Zuckerman (1962); Lundberg and Larsen (1949); Newman (1962); Ognibene (1970); Pan (1951); Reuss (1943); Skelton (1963). Although no generally valid conclusions are possible, many useful results have been obtained regarding particular types of surveys and methods of data collection.

C. Some examples

1. Introduction

The overall incidence of non-response can vary greatly from one country to another, and from one survey to another. In some voluntary surveys in developed countries, quite high proportions of the sample may fail to respond in spite of strenuous efforts on the part of the survey organization. For example, non-response rates of 16% to 48% have been reported in major continuous government-sponsored survey in the United Kingdom (UK Office of Population Censuses and Surveys, 1978). In many developing country surveys, much lower-non-response rates have been reported. Within a given country, non-response is frequently higher in urban than in rural areas. Also, the composition or source of non-response is different in different situations. While refusals or failure to contact respondents may be more common in developed countries, failure to gain access to or locate sample units may be the main source in some developing country surveys.

Examples of non-response rates in surveys from several countries are presented in this section, first for developed countries and then for developing countries. These examples document the substantial variation between countries and between surveys in the levels and types of non-response. Some of the examples from developed countries also illustrate trends in non-response in recent years.

2. Illustrations from developed countries

For three major continuous surveys in the United Kingdom, non-response rates in 1971 were as follows (United Kingdom, Office of Population Censuses and Surveys, 1978):

General Household Survey	15%
Family Expenditure Survey	30%
National Food Survey	46%

These surveys differ widely in the burden which they impose both on the survey organization and on responding households. This may account in part for the difference in response rates. The General Household Survey comprises a household schedule plus separate interviews with all adult members of the household; however, proxy interviews are taken if one or more household members are unavailable. The Family Expenditure Survey requires completion of a household schedule covering major and regular expenditure items, an income schedule for each adult in the household, and a comprehensive record of purchases made by each spender over a two-week period. If any individual fails to respond, the response of the household as a whole is invalidated. The National Food Survey is a very detailed enquiry involving multiple visits and record-keeping.

Some additional details are available for the General Household surveys. Barnes and Birch (n.d.) report data collected in the course of several years. The table below includes separate data on partial non-response, i.e., the percentage of cases for which less than complete information was obtained; such information is not usually provided in studies on non-response.

Table 4.3

Data on non-response from the UK General Household Survey: 1971-74

Outcome	1971	1972	1973	1974
Successful response	70.8	69.2	69.7	74.3
Partial response	14.6	14.7	14.0	11.6
Non-response total	14.6	16.1	16.3	14.1
Non-contact	2.7	2.6	2.8	2.4
Refusal	11.9	13.5	13.5	11.7

Note that the non-response rates are stable from year to year, and that refusals generally account for over three quarters of the total non-response. On the other hand, it reflects well on the survey organization that the rates of non-contact are uniformly low.

The presence of high non-response rates in most government surveys in Sweden has created considerable concern about the problem, and has resulted in much research to shed more light on the sources of non-response. Non-response rates for a number of surveys over several years are presented by Lyberg and Rapaport (1979):

Table 4.4

Non-response and refusal rates for government surveys in Sweden: 1970-78

Survey Year	(1) LFS* T**	R**	(2) PPS* T	R	(3) HIS* T	R	(4) SCBE* T	R	(5) SLC* T	R	(6) SHEC* T	R
1970	1.7	1.2										
1971	3.0	2.3										
1972	3.4	2.7										
1973	3.6	2.6	11.7	6.6	10.9	8.9						
1974	4.2	2.7	14.5	7.5	11.8	9.4	17.6	6.9	17.6	16.5		
1975	7.1	3.8	14.7	7.9	12.8	8.2	15.3	10.8	19.1	18.0		
1976	7.4	4.1	14.8	8.9	12.9	9.1	16.1	10.8	21.3	19.5		
1977	7.0	3.9	12.9	7.1	11.9	7.9	15.4	10.3	20.0	18.2		
1978	6.5	3.6	12.4	6.6	12.3	8.1	14.4	9.8	15.2	13.9	17.7	12.0

*The results are shown above for the following six surveys: (1) Labour Force Survey, (2) Party Preference Survey, (3) Household Income Survey, (4) Survey of Consumer Buying Expectations, (5) Survey of Living Conditions, and (6) Survey of Household Energy Consumption.

**T = total non-response rate, **R = refusal rate.

The Labour Force Survey (LFS) shows the lowest non-response rates. During the first half of the 1970s, the rate increased considerably, but stabilized at around 7% after that. Refusal rates have also increased, but account for a smaller proportion of the total in the latter part of the period. In the other surveys, non-response rates are substantially higher and no overall time trend is indicated. There are differences between surveys in the level of refusals as a proportion of the total non-response. In the party preference survey (PPS), for example, refusals account for about one-half of the non-responding cases; by contrast, non-response in the survey of living conditions (SLC) is largely due to refusals.

A recent study on non-response in a variety of surveys in Norway (Thomsen and Siring, n.d.) concludes that:

"(1) Completion rates show a serious decline during the last ten years in most interview surveys done by the Central Bureau of Statistics of Norway.

(ii) Refusals account for approximately fifty per cent non-response, independent of completion rates.

(iii) If no extra efforts had been made to increase completion rates, the decline in response rates would have been larger than that actually observed."

Table 4.5 shows some examples of the upward trend in non-response for selected types of surveys, summarized from this study:

Table 4.5

Trends in non-response for selected surveys in Norway

Political opinion surveys	Year	1969	1971	1972	1973	1977
	Non-response(%)	9.9	12.6	19.2	19.4	21.6
Family expenditure surveys	Year	1967	1973	1974	1975	1976
	Non-response(%)	21.8	28.6	32.6	32.3	31.0
Surveys of housing conditions	Year	1967	1973	) The two surveys were		
	Non-response(%)	9.9	22.9	)-almost identical in) procedures.		

The study by Thomsen and Siring also illustrates the effectiveness of call-backs in reducing non-response. Table 4.6 shows the percentage of non-response after termination of "normal data collection" (which includes some call-backs), and the final level of non-response after extra call-backs were made by telephone and by specially-trained interviewers.

The results are for a labour force survey in Norway, conducted every quarter since 1972. For illustration, only data for the second quarter of each year are shown.

Table 4.6

Effect of extra call-backs on non-response rates in Norwegian Labour Force Survey: 1972-78 (2nd quarter)

Year (data are for 2nd quarter	% Non-response rate		Per cent distribution by reason		
	Normal Procedure	After extra call-backs	Refusals	Not at home	Other
1972	8.4	7.3	53	39	8
1973	10.5	9.5	NA	NA	NA
1974	10.6	8.8	41	42	17
1975	11.1	7.4	45	43	12
1976	14.5	9.9	45	38	17
1977	14.1	10.4	47	37	16
1978	16.4	11.9	45	42	13

Even though this survey allowed for proxy responses by other members of the household, as much as 40% of the non-response resulted from failure to find a suitable respondent at home.

Data contained in a recent evaluation of labour force statistics in the United States (National Commission on Employment and Unemployment Statistics, 1979) show non-response rates by reason over a period of 24 years.

Table 4.7

Non-response rates, total and by reason, for the United States Current Population Survey: 1954-77

Period	Total non-response rate*	Non-response rate by reason			
		No one at home	Temporarily absent	Refusal	Other
1954-55	4.2	1.6	1.6	0.6	0.4
1962	5.0	1.9	1.8	1.0	0.4
1967	4.9	1.4	1.2	1.8	0.5
1972	4.0	1.0	1.0	1.8	0.3
1977	4.1	0.7	0.7	2.5	0.2

*Per cent of eligible households for which an interview was not obtained.

These data show that the total non-response rate has been maintained at about the same level, but only by a successful effort to reduce non-response for reasons other than refusals in order to compensate for the increase in refusal rates over the period.

The illustrations given above provide a broad enough ground for concern about non-response in interview surveys in developed countries. The experience in these countries is, of course, not uniformly discouraging. For example, a rather optimistic picture of the effectiveness of successive call-backs is given by the data presented by Cole (1956) for a household income and expenditure survey in Great Britain. The results in Table 4.8 are unusual in that the call-backs resulted in complete coverage of the sample (note, however, that the results refer to a period more than 25 years ago).

Table 4.8

Interviews completed, by number of calls, for a household expenditure survey in Great Britain

No. of calls	1	2	3	4	5	6 and more	Total
Additional % households contacted	62.0	22.3	9.0	4.1	1.6	1.0	100.0

3. Illustrations from developing countries

There is a measure of agreement among survey statisticians that the co-operation of respondents with household surveys in developing countries is generally much better than that in developed countries. Once the interviewer is able to locate and approach the respondent, the respondent is very unlikely to refuse to co-operate. For example, in a large number of national fertility surveys in various parts of the world, the refusal rate rarely exceeded 0.5% (Scott and Singh, 1980).

This does not mean that the problem of non-response in developing countries can be assumed to be non-significant. Rather, and more typically, the problem arises from the failure to implement the selected sample, the failure of the interviewer to locate and contact the respondent, and in many cases, from the inability of the respondent to provide the desired information.

As an illustration, data are presented in Table 4.9 from several surveys conducted in developing countries under the World Fertility Survey programme (Verma, 1980). The table shows response rates for a two-phase operation: first, households were interviewed (the left hand panel), on the basis of which women in the child-bearing ages were identified for a more detailed individual interview (right hand panel). The data show that once a household or individual was contacted, refusal to co-operate was extremely rare. Failure to find the

respondent at home was somewhat more common, though it was still under 1 per cent for the household interviews in most Asian surveys and under 3 per cent in most surveys in Latin America. However, the largest category responsible for non-response was the inability of the interviewer to locate a sample address or to find an eligible household or respondent at the address.

Table 4.9

Sample outcome and response rates for fertility surveys in selected developing countries

	HOUSEHOLD INTERVIEWS Per cent of selected dwellings or hh:					INDIVIDUAL INTERVIEWS Per cent of identified women:			
	Inter-viewed	Re-fusals	Not at home	Other non-response	Vacant, not found, etc.	Inter-viewed	Re-fusals	Not at home	Other non-respons
Ghana	83.8	0.8	4.9	0.7	9.8	99.8	0.1	0.1	0.0
Kenya	82.6	0.3	1.4	4.7	11.0	95.8	-	-	-
Tunisia	85.2	--------4.8--------			10.0	92.2	-	-	-
Turkey	79.2	1.0	8.5	1.6	9.7	-	-	-	-
Bangladesh	94.0	0.2	0.4	1.1	4.3	98.0	0.5	1.0	0.5
Fiji	91.0	0.2	0.6	4.7	3.5	97.5	0.2	1.4	0.9
Indonesia	96.7	-	-	-	-	97.7	0.0	1.7	0.6
Malaysia	95.7	0.1	0.1	0.0	4.1	99.2	0.1	0.1	0.6
Pakistan	93.6	0.1	0.4	1.3	4.6	99.0	0.0	0.5	0.5
Philippines	86.4	0.4	1.0	2.7	10.5	96.5	0.3	2.4	0.8
Sri Lanka	89.4	-------0.3--------			10.3	99.4	-	-	-
Thailand	95.2	-------1.3--------			3.5	94.4	-	-	-
Guyana	95.0	0.7	0.9	0.8	2.6	95.6	0.4	1.4	2.6
Haiti	80.9	-	-	-	-	94.8	-	-	-
Jamaica	82.7	3.2	3.0	0.2	10.9	93.6	2.8	2.2	1.3
Costa Rica	89.9	0.5	1.0	0.0	8.6	96.7	1.1	2.0	0.2
Mexico	89.3	0.6	2.9	0.2	7.0	95.3	0.5	3.5	0.7
Panama	91.4	0.1	0.6	0.0	7.9	97.5	0.0	2.5	0.0
Peru	89.2	0.5	3.0	0.0	7.3	93.0	0.4	6.2	0.4

('-' = not available)

The figures in Table 4.9 illustrate a basic difficulty in accurately reporting the incidence of non-response in such surveys. The figures for non-contact depend critically on distinguishing dwellings which are unoccupied from those whose inhabitants are temporarily away. Unoccupied dwellings merely amount to "blanks" in the sample lists and do not represent non-response, while dwellings whose inhabitants are temporarily away do represent cases of non-response. Field interviewers are not always able to make this distinction clearly. Similar confusion can arise between sample addresses which ceased to exist since the lists were made (and do not represent cases of non-response), and addresses which the interviewer is unable to locate due to inadequate description in the list. Most of those in the latter category are likely to be cases of non-response. Inadequate documentation and the use of out-of-date lists for sampling increase such problems and at the same time they also increase the incidence of non-contact.

Some surveys in developing countries have reported substantially larger non-response rates than those presented above. In addition, there can be substantial regional and urban-rural differentials in response rates. For example, in a recent survey of married couples in Thailand, only 73% of the eligible males were successfully interviewed in the Bangkok Metropolis while the overall response rate in the country as a whole was over 92% (Thailand, 1977). In a national survey of married women in Turkey, the response rate in urban areas of one of the regions of the country was only 65%, while the overall response rate for the survey was 79% (Turkey, 1980). The latter study noted that "the timing of the field work contributed substantially to non-response, particularly in urban areas. Field work had to be completed prior to the onset of winter, which unfortunately resulted in the interviewing period coinciding with vacation months. Moreover, in some regions such as (the region with the highest non-response rate mentioned above), dwellings were rather scattered making call-backs and

sometimes even location of the house difficult for the interviewer. Exactly the same problems, in fact, have been faced in previous Turkish surveys, though their magnitude could have been reduced with more careful updating of the frame".

D. Effects of non-response on survey results

1. General considerations

Non-response bias results when the non-responding units differ in relevant respects from those which respond to the survey or to particular items. In this sense, the effect is similar to that of non-coverage as discussed in the previous chapter. However, while non-coverage is difficult to measure, the incidence of non-response is or can be available in most cases from the results of the survey itself. In diagnosing the causes of known deficiencies in the responding or achieved sample, it is often hard to distinguish the effect of non-response from the effects of shortcomings in the construction of the sampling frame or the selection of the sample (i.e. from coverage and implementation errors). Thus, under-representation of particular categories of households within the achieved sample may occur because they tend for some reason to be omitted from the sampling frame or to be missed by the field enumeration procedure, rather than because they are less likely to respond. Correct diagnosis of the source of the error is nevertheless vital, since the measures necessary to reduce the bias will be quite different according to whether the cause is sample deficiencies or non-response.

The effect of non-response on survey results has to be judged in relation to the magnitude of other possible errors. In some surveys, non-response may not be a very significant source of error. If the total non-response rate is 1 per cent, then the bias due to non-response, at least for estimates of attributes, cannot be large, even if non-responding units differ radically from those for which data are obtained. As described in Section F of this chapter, the

use of appropriate imputation and estimation techniques can further reduce biases due to non-coverage and non-response.

However, there is no room for complacency about non-response. Most of the characteristics of non-responding units are not known and may be assumed to be different from those of responding units. Some illustrations of these are given below. Non-response can be kept at a low-level only through strict attention to the development of survey questionnaires and data collection procedures and to their use in the field.

2. Direct indicators of potential non-response bias

There are several methods of obtaining quantitative information which give, if not a direct measure, at least a direct indication of the potential non-response bias in a household survey. One indication of the bias can be obtained from a comparison of values reported for units who respond after the first call with values for units who responded only after subsequent calls. Mail surveys, though not very relevant so far as household surveys in developing countries are concerned, provide a convenient basis for such an analysis of the effect of non-response. In many mail surveys, the first attempt results in much lower response rates compared to most interview surveys, while substantial improvements have been reported after subsequent reminders; also repeated attempts are more feasible because in mail surveys explicit refusals are not common. As an illustration, in a mail survey in the State of North Carolina in the United States, successive mailings resulted in the outcome shown in Table 4.10 (Hendricks, 1956).

Table 4.10

Results from initial and follow-up mailings for an agricultural survey in North Carolina

Mailing	No. of Responses	Average number of trees per farm
1	300	456
2	543	382
3	434	340

It is obvious from this table that an estimate of the average number of trees per farm would be very biased if the farms not responding to the first mailing were omitted. Clearly, smaller farms are less likely to respond to the enquiry. Finkner (1950) reported similar experiences from surveys of peach growers in North Carolina.

Data from the United Kingdom (Gray and Corlett, 1950) which indicate the same effect in an interview survey on household income and expenditures are shown in Table 4.11. The table shows (a) the average expenditure on meals in catering establishments per adult per week (shillings), and (b) the average number of visits to the cinema per adult per month, tabulated against the number of call-backs required to obtain an interview with the household.

Table 4.11

Selected results by number of calls required from a UK income and expenditure survey

No. of calls	(a) Average Expenditure	(b) No. of Visits
1	1.83	2.02
2	3.10	2.44
3	3.88	2.47
4 or more	3.61	2.00
All persons interviewed	2,69	2.26

The data show that people who go out for meals or to the cinema more frequently were more difficult to contact. Their exclusion from the survey would have resulted in an underestimation of these items of consumption.

Another method of studying non-response bias is to obtain some of the survey data for non-respondents from another source, such as a population census. This method is illustrated by a study of non-response in the General Household Survey in Great Britain which used the data from the 1971 Census for the non-respondents in the survey (Barnes and Birch, n.d.). The household survey had many questions which were also included in the 1971 Census of Population and Housing. The procedure of the study was to list the non-responding units in the household survey and locate their respective questionnaires from the more publicized and obligatory census. The availability of the census information for the units not responding in the survey made it possible to obtain data to compare respondents and non-respondents in various sub-groups. It was possible both to examine differential response between various sub-groups and to calculate, for a number of substantive variables, the correction factors required to compensate for non-response. Some significant differences in response rates by sub-group were revealed, but most of the correction factors were inside the range 0.95 - 1.05. This suggested that the effect of non-response bias upon the characteristics for which comparisons were possible was only moderate. The overall non-response rate in the survey was 15%, though it varied from one sub-group to another.

In the previous example, the basis of evaluation was the presence of an appreciable number of substantive items common to the survey and a comparable external source (census). Frequently, however, the common information may be confined to a few characteristics by which the respondents in the two sources may be classified. In such cases it may be possible to estimate the effect of non-response through a sample reweighting procedure. This was done for the British Family Expenditure and National Food Surveys (United Kingdom,

Office of Population Censuses and Surveys, 1978). The main substantive measures produced by these surveys were not included in the 1971 census, but the surveys contained information for classification of respondents by certain characteristics which were also available from the census for the units in the sample. Comparison of the distribution of the sample units by these characteristics between the census and the surveys yielded direct evidence of differential non-response. Indirect estimates of bias affecting substantive measures (e.g. income, expenditure and food consumption) could be made by reweighting the sample to correspond with the census distribution on the common characteristics. (These estimates are themselves unbiased only if the mean values for respondents and non-respondents are the same for each of the classes of household reweighted separately.) The study report notes that "the most consistent overall findings to emerge from the checks were that elderly households, single person households and households without children were significantly under-represented in the responding survey samples. As expected, between-group differentials in response and hence the potential for substantive bias tended to increase as the overall level of response decreased. A check, using a reweighting method, on bias in FES (Family Expenditure Survey)-derived estimates of the incidence of taxation and benefits suggested that such bias was only rarely serious enough to affect general conclusions drawn from the analyses". The overall response rate in the survey was 70%.

The most direct method (but also a difficult one to use) of estimating the effects of response bias is to mount a special intensive follow-up study to obtain data for the originally non-responding cases, or for a sub-sample of them. Table 4.12 presents the differences in demographic and labour force characteristics of respondents and non-respondents in a labour force survey in Canada (Gower, 1979), estimated by using this method.

Table 4.12

Characteristics of respondents and non-respondents in a Canadian Labour Force Survey

Survey item	Responding household	Total	Refusals	No one at home	Temporarily absent
Mean size of household	2.24	1.80	2.19	1.64	1.75
Unemployment rate	8.3	8.9	9.5	9.7	8.5
Participation rate	62.5	63.6	63.7	74.2	55.8

Anderson (1979) reported on the effects of non-response on fertility data from surveys in Sao Paulo State, Brazil and in El Salvador. Certain fertility items were collected for all women 15 to 49 in sample households during a listing operation, and the same information (plus other more detailed data) was obtained in interviews from those listed in each sample household. Table 4.13 shows selected estimates based on the two sources of data.

Table 4.13

Comparison of fertility measures estimated from two sources in recent contraceptive prevalence surveys

Survey and item	Source of estimate (1) Interviews with women in sample	(2) Listing Operation	Ratio (1)/(2)
Sao Paulo State, 1978			
Mean number of children born alive	2.2	2.0	1.10
General fertility rate	104	96	1.08
Total fertility rate	3.1	2.8	1.09
Crude birth rate	26	24	1.08
Non-response rate (%)	11.5	NA	NA
El Salvador, 1978			
Mean number of children born alive	3.4	3.2	1.05
General fertility rate	219	208	1.05
Total fertility rate	6.6	6.3	1.05
Crude birth rate	45	43	1.05
Non-response rate (%)	21	NA	NA

The non-response rate shown for each survey in the first column is the percentage of women who were selected from the household listing as respondents for detailed individual interviews but for whom no interviews could be obtained. The author points out that "since whether a woman has children or not has an effect on whether she stays at home, non-response in a household survey can be related to fertility". He attributes the higher values of fertility measures estimated

from the individual interviews primarily to the differences in fertility between respondents and non-respondents, although recognizing that other factors, such as the use of proxy respondents for some women in the listing operation, could have accounted for part of the differences observed. It may be noted that at best, table 4.13 provides an indirect estimation of the effect of non-response.

3. Indirect indicators of potential non-response bias

Indirect indications of non-response bias may be had by comparing estimates derived from the sample with externally available information such as population totals by age, sex and political subdivision. Such evidence is indirect in the sense that it does not normally permit non-response bias to be distinguished from bias arising from other sources. Reference has been made to such comparisons in section F.3 of the previous chapter.

The compounding of non-response bias with other biases can be a serious problem in estimating the magnitude of the bias involved. For example, in a survey of alochol consumption, a downward bias in the level of consumption may result from (i) lower-than-average response rate on the part of households or individuals with higher-than-average level of alcohol consumption, (ii) deficiencies in the sampling frame leading to omission of some of the heaviest drinkers, for example due to many of them not having a permanent place of residence, (iii) under-reporting of alcohol consumption, especially by the heavier drinkers, due to the social stigma often attached to high consumption, (iv) alterations in behaviour due to the very fact of being interviewed, especially if the survey is an elaborate one with multiple interviews and keeping of detailed diaries or records of consumption. All these biases might act in the direction of under-estimation of consumption, and the under-estimate may be serious insofar as a relatively small proportion (for whom the results may be more seriously biased) account for a large proportion of the total alcohol consumption of the population. An approximate idea of the combined effect of these biases might be had in some countries by

comparing the survey estimate of total consumption with data from other sources on the total production or sales of alcoholic beverages, but it would not be possible to single out the separate contribution from each source of bias.

E. Improving response rates

1. General considerations

Whatever the actual magnitude of bias resulting from non-response in a particular survey, a survey organization in practice has little choice but to take steps to maximize response in household surveys, within the resources available to it. This is because non-response is frequently the most visible index of the quality of whole survey operation. The response rates achieved reflect the reception of the survey among the population, the quality and completeness of the sampling frame, the appropriateness of the content of the survey and the burden it puts on the respondents, the timing of the field work, the personal qualities, training, experience and supervision of the interviewers, and the organization's control over execution of the operation.

Fundamentally, improving response rates is a matter of careful planning and design of the survey operation. This is so at least in countries other than the few where general antipathy towards statistical surveys may have become rather widespread. Particularly in a continuing programme of surveys, much can be achieved by instituting long-term measures to identify sources of non-response and to control its incidence. Further, the co-operation of the population with statistical surveys is likely to depend on a long-term educational process, as has been demonstrated by the success of population censuses in many countries.

2. Contacting respondents

In many developing countries, the problem, as noted earlier, is

largely that of locating sample units and contacting suitable respondents. Here the initial focus has to be on the improvement of the sampling frame and, as far as possible, of means of transport. Most household surveys in developing countries use listing and sub-sampling of the ultimate area sampling units. If the household is the unit for listing and sub-sampling, it is essential that the lapse of time between listing and interviewing be kept short, preferably less than one month. Whether the listing units are households or housing units, enough information should be recorded on the listing form so that they can be readily located and distinguished from other nearby units by the survey interviewers.

For various economic, administrative and even statistical reasons, constraints are usually placed on the length of time interviewers are allowed for completing their assignments. However, from a response viewpoint, it is crucial that these constraints do not make it unreasonably difficult for the interviewers to cover all households assigned to them and to make the necessary call-backs to obtain responses.

Interviewers should be able to make calls at sample units at times when prospective informants are likely to be at home. Persons recruited as interviewers must be able and willing to do this. This is particularly important if the survey uses part-time interviewers. Interviewers must be given thorough training in how to plan and update calling schedules so as to make best use of their time and journeys in securing response. In this context, the advantage of having interviewers stationed in or near sample areas is obvious.

3. Gaining co-operation

Particularly in countries and types of areas where refusals are more common, it is useful to establish methods for interviewers to win the confidence of respondents. In this connexion, several factors may be important: the way in which the interviewers introduce themselves, the identification they carry, what they say about the

survey, how they dress and behave, and the courtesy they show to respondents. Statistics Canada has issued a booklet called "Doorstep Diplomacy" where all these problems are discussed, as a part of the instructions for the enumerators. It appears that the interviewers' diplomacy has an important bearing on the respondents' reaction and their decisions to refuse or co-operate. It can also be an important factor in gaining the permission of local officials to proceed with survey work in their areas, and in obtaining their help in locating and gaining the co-operation of sample units.

Provision of small incentives to respondents, as a token of appreciation for their co-operation, can be effective. However, such methods may be considered unduly expensive for surveys in a poor country. In some developed countries such as Canada and Sweden, information booklets about survey procedures or survey results have been distributed to the respondents to enhance their interest in ongoing and future surveys. Incentives are most commonly used in surveys that place the most burden on respondents, such as income, expenditure and food consumption surveys, and longitudinal surveys.

In recurring surveys, it is possible to prepare respondents for subsequent visits. Once a personal contact is established, and the respondent is aware of the time of the next visit and the content of the questionnaire, contact and co-operation may be more easily secured. Respondent fatigue in repeat surveys is thus only one side of the equation. It may also be useful in recurring surveys to record the names of relatives or friends who are likely to know the location of respondents who have moved or are temporarily absent in subsequent rounds.

Finally, it is important to recognize that when strong pressures are used to obtain interviews from reluctant respondents, the information obtained may be less accurate than that obtained from more willing respondents. Thus, for the "hard core" of non-respondents, it may actually be a superior strategy to abandon further attempts to

obtain interviews and to rely on other methods to minimize the resulting bias. To some extent, the differences in non-response rates for different countries which were presented earlier in the chapter may have resulted from differing philosophies concerning the amount of pressures or persuasion that should be applied to obtain interviews.

4. Interviewer selection, training and supervision

In personal-interview surveys, the interviewer is bound to play a crucial role in strategies designed to maximize response. Often, the interviewer is the only link between the survey organization and the informant. It is, therefore, absolutely essential for good response, as well as for satisfactory data, that the interviewers used in household surveys be carefully selected and well-trained and motivated. The development of a regular cadre of interviewers would permit such thorough training to be cost-effective. Studies of interviewer variance (discussed at length in Chapter V, Section D) have shown significant differences between interviewers in rates of item non-response, suggesting that some interviewers are much more thorough and persistent than others in completing their work.

Close supervision of interviewers' work, and particularly an up-to-date check on their achieved response rates is essential. Often a small proportion of interviewers is responsible for a disproportionate share of the non-response. However, it would be a mistake to automatically blame an interviewer with below-average response rates without first investigating the external factors which may be responsible for it. The important thing is to design the interviewing procedures in such a way that the outcome of each sample case assigned is recorded, including information on reasons for failures to obtain interviews. These data should be summarized separately for each interviewer and steps should be taken to improve sub-standard performance.

F. Dealing with non-response

1. Introduction

Several methods can be used to try to compensate for the effect of non-response on the survey results. Some of these methods are part of the data collection procedure, for example, intensive follow-up of a sub-sample of non-respondents, or the collection of limited data from neighbours for households that are away during the data collection period. The substitution of other units for those units which cannot be interviewed is a controversial practice and will be discussed in this section.

Other procedures, generally less costly, can be used during data processing. These come under the general headings of imputation of missing data and estimation procedures which attempt to compensate for missing data. There is not always a clear distinction between the two classes of procedures. In general, the procedures used in data processing (as well as substitution in the data collection) rely on assumptions of similarity between responding and non-responding units, either in the whole population or, preferably, within more homogeneous sub-groups of the population. While such assumptions are unlikely to be completely valid, the procedures are often simple to use, and may result in significant improvement of the survey estimates at little cost.

Non-response in household surveys cannot be effectively dealt with unless it is properly identified during data collection. It should be standard practice for interviewers to account for the outcome of every sample unit assigned to them. This means recording whether or not an interview was obtained and, if not, explaining the circumstances in sufficient detail so that each unit not interviewed can be classified as eligible for interview, not eligible, or eligibility not determined. This information should be transmitted to the data processing unit for use in connexion with whatever imputation and estimation procedures are to be applied. Similarly, it is

essential that questionnaires be so designed that the processing staff (and eventually the computer) can distinguish between items for which no response was obtained and those for which the response was "none", "zero", or not applicable.

2. Methods used in the data collection stage

Various methods have been proposed for dealing with non-response at the data collection stage. Some of them are completely unbiased; others assume some degree of similarity between responding and non-responding units, at least within defined sub-groups of the target population. Except for substitution, these methods have not been used very much in personal interview household surveys, and none of them is recommended here. However, they will be discussed briefly in order to explain why they are not generally considered suitable or applicable for household surveys in developing countries.

The recommended way of dealing with non-response during the data collection stage of the survey is to make a vigorous and thorough effort to obtain responses for all, or nearly all of the eligible units in the assigned sample. If no acceptable respondent is available when a unit is first visited, call-backs should be planned. If possible, the interviewer should ask neighbours when the occupants are likely to be at home. The optimum number of call-backs in a particular survey depends on several factors. Primary considerations are whether or not call-backs will require separate visits to the area units (blocks, villages, etc.) where they are to be made and, if so, the time needed and distance to be travelled to reach those area units. Deming (1953) examined the cost-effectiveness of call-backs and showed that in some situations a large number of successive calls may be justified.

When repeated call-backs fail, it may be possible to partially complete a questionnaire for the assigned unit by observation (e.g. for many housing characteristics) and by asking neighbours for information. Interviewers should be instructed which of the survey items they may properly ask of neighbours. These items should include only

non-sensitive questions for which neighbours might be expected to give reasonably accurate answers. Information about the basic characteristics of the members of the sample household will be useful in the data processing stage, at which point the missing items can be imputed, using data obtained for persons and households with similar characteristics.

Methods of dealing with non-response that have been suggested but are not recommended are:

(a) Intensive follow-up of a sub-sample of non-respondents. It has been proposed, especially in connexion with mail surveys, that, following a reasonable effort to obtain response by standard survey procedures, a sub-sample of the remaining non-respondents should be selected and more intensive procedures used to obtain the survey information for these units. This method is generally attributed to Hansen and Hurwitz, who gave a detailed treatment in a 1946 article; however, Dalenius (1957) points out that the idea was first suggested by Cornfield (1942).

If interviews are actually obtained from all of the sub-sample of non-respondents, and the selection probabilities are properly reflected in the sample estimates, this approach will completely eliminate bias due to non-response. However, this is done at the cost of an increase in the sampling variance. Using the method in a personal interview survey would substantially complicate the sample design and estimation procedures, and would be likely to delay completion of the data collection. These and other disadvantages (see Fellegi and Sunter, 1974, pp. 132-3) explain why the method is seldom used, and is not recommended for household surveys in developing countries.

(b) The Politz-Simmons method of eliminating call-backs. An ingenious but in the final analysis impractical and biased method of eliminating call-backs was proposed by Politz and Simmons(1949). It involved visiting each sample unit only once, at a randomly selected time of day. Respondents contacted were asked how often during the previous 5 days they had been home at that same time of day, and the sample responses were weighted accordingly, e.g., those at home on all 6 days received only half the weight given to those only at home on 3 of the 6 days. The potential biases and difficulties of execution of this method are fairly obvious, and it has not been used in any government-sponsored surveys.

(c) Adding a sample of non-responding units from previous surveys to the initial sample. This method, which has been used by the Survey Research Center of the University of Michigan, was described in an article by Kish and Hess (1959). The initial sample for a survey is augmented with addresses for which no response was obtained in recent similar surveys. A specified number of calls is made to obtain responses for each of the two sets of addresses. Subsequently, the responses obtained in the current survey for units not responding in prior surveys are treated as "replacements" for the non-response units in the new sample for the current survey. The authors argue that "... roughly speaking, the ... calls made on the replacement addresses can be considered as additional to the original calls at the same addresses".

Under certain reasonable assumptions this technique can reduce the bias due to non-response. However, it is not likely to be effective if a large proportion of the non-response comes from refusals, and its use is only feasible

where there have been recent prior surveys with quite similar content and collection procedures. Furthermore, it introduces some complications into the survey design, and requires very careful controls for proper execution.

(d) Substitution. Unlike the three methods just discussed, substitution of other units for the non-responding units has been used in several household surveys, both in developed and developing countries. The rationale for its use is usually to insure that completed interviews will be obtained for the exact number of sample households specified in the initial design.

Many different methods of substitution have been used. A common procedure is to divide the sample of households in each ultimate cluster randomly (or systematically) into two groups: the assigned sample and the reserve sample. Under specified conditions, which usually include making a certain number of call-backs, if it has not proved possible to complete an interview, the interviewer is permitted to substitute the first available unit in the reserve sample.

Substitution does not eliminate non-response bias. This may be understood if we view the survey universe as being divided into two groups: those households for which it is possible, by following specified survey procedures, to obtain an interview, and those for which, using the same procedures, it is not possible to obtain an interview. Substitution increases the sample size for the first group, but does not provide any representation of the households in the second group. Characteristics of the two groups are certain to differ, and the substitution process has done nothing to reduce the bias resulting from these differences.

Substitution does control sampling error by achieving the desired sample size, but this can be done almost as well by making the initial sample large enough to allow for the expected proportion of households which cannot be interviewed.

The main argument against substitution is that in actual practice its use is likely to introduce additional biases. This happens because:

(i) Frequently the rules established for substitution are biased, e.g., when an occupied housing unit is substituted for a vacant one or when a reserve household is used because a new household is occupying the housing unit that was previously occuped by the specified sample household.

(ii) It is extremely difficult to prevent interviewers from making unauthorized substitutions. It will be a great temptation to make a substitution after one unsuccessful visit to a household, rather than making the specified number of call-backs.

(iii) The use of substitution diverts attention from the problem of non-response bias without doing anything to alleviate it. It essentially forecloses any effort to reduce the effects of non-response bias through the use of appropriate estimation and imputation procedures.

3. Methods used in the data processing stage

(a) Estimation-based methods. Typically, the weights assigned

to sample data in order to produce estimates for the survey population have three components:

(i) Factors needed to adjust for non-response in the survey.

(ii) Factors reflecting the selection probabilities of the individual survey units.

(iii) Factors needed to make estimated totals from the survey agree with comparable totals available from other sources.

In practice, the weights applied in the tabulations to produce estimates usually represent the product of all three factors. However, to understand their derivation, it is necessary to look at each component. In this section, we will examine only the first component, which is intended to reduce non-response bias in the survey estimates.

Probably the most common approach is to reweight the sample clusters by a factor inversely proportional to the response rate in each cluster. This procedure does not depend upon any external information, and is based on the assumption that the most important factor determining the effect of non-response is resulting distortion in the geographic distribution of the sample. A number of examples are available in the surveys conducted under the World Fertility Survey Programme (for a description of the weighting procedure, see World Fertility Survey, 1977). These studies do not report estimates of the effect of non-response; rather they describe a weighting procedure which is used to "compensate" for differential non-response by sample cluster.

If certain information is available for both respondents

and non-respondents, this information can be used in adjusting for non-response. For example, the number of persons in each household may have been recorded in a separate listing operation. If so, it is possible to make separate non-response adjustments for households in each size category. For technical and procedural reasons, it is usually preferable to make this kind of adjustment for strata, or groups of ultimate area units, rather than separately for each such unit. This technique is sometimes called post-stratification, because it establishes strata that were not used in the process of selecting the sample. Post-stratification can be based on whatever information is available for all sample units, including those not responding. The objective should be to establish strata likely to have different response rates and different values of key variables. From this point of view, size of household is a good variable to use, because it is usually harder to complete interviews for small households than for large ones.

Non-response bias remaining after application of the weighting factors just described may sometimes be further reduced by the introduction of the third weighting component mentioned earlier, namely, the factors used to adjust the survey estimates to externally available counts or projections of universal totals. However, this specific effect is virtually impossible to measure, because the same adjustment may be compensating for under-coverage in the survey, or for the fact that the particular sample selected over-estimated or under-estimated the population or some sub-groups.

(b) Imputation. In a broad sense, imputation means replacing missing or unusable information with usable data from other sources. These sources can include the same questionnaire (if partial response was obtained), another questionnaire from the same survey, or external sources, such as another survey or an administrative record. The use of imputation techniques is expanding rapidly in scope and sophistication as a result of advances in computer technology. A review of techniques is provided by Chapman (1976).

If computers are used for imputation, two approaches are available: the cold-deck procedure and the hot-deck procedure (Nordbotten, 1963). A distinction will be made between imputation of partial non-response and imputation of the whole set of data for non-responding units. The former case arises if, say, a household responds to the whole questionnaire except for information about income. The latter case refers to the failure to obtain any information from the household.

If, for example, income is to be imputed for a household, the cold-deck procedure may start with an earlier survey of the same type and in the same population. The households in the earlier survey may be cross-classified by a few variables, such as the size of household and number of persons employed. The classification will result in a number of cells that should be as homogenous as possible. As it is not known in advance where the missing information will arise in the next survey, it is necessary to have at least one household from the earlier survey in each cell. In order to speed up the data processing of the current sample, it is important to carry out these classifications in advance. A household with missing

information on income in the new survey is classified in the same way as before and the appropriate cell for it is ascertained. The missing income is then taken from the same cell in the previous survey and all the subsequent data processing is carried out as if the imputed values were collected in the field.

Various methods have been used to select missing response from the cell concerned. One of them is to select one of the available responses in the cell at random. Another could be to take the average of responses in the cell and use it as the imputed value. In the latter procedure extreme values are avoided, but the distribution within the cell can be distorted.

The cold-deck procedure depends upon the availability of data from a previous survey of the same type. As this condition is frequently not met, an alternative is to use the hot-deck procedure, which is carried out as aobve but makes reference only to the data collected in the current survey. The survey itself provides the classification variables and the establishment of cells. Whenver a response is missing, the imputed value is taken from the last preceding unit that happened to fall in the same cell. The use of the nearest unit in the cell rather than a random selection is considered preferable because it suits the computer operations. In addition, it might be expected that the response of a nearby unit will be closer to the true value of the missing information than the response of a unit from a different part of the country.

When the survey design is such that households remain in the sample for more than one round, it might be possible to impute missing data by using information

obtained from the same household in an earlier round. Again using income as an example, instead of taking the actual value reported in the previous round, it might be preferable to apply an adjustment based on average change in income between the two survey reference periods. This technique, however, involving linkage of individual household data from two surveys, is only feasible if the data processing unit has a fairly sophisticated data-base management system.

In the case of total non-response, the situation is more complicated. As there are few or no auxiliary variables to classify the non-responding cases, the choice for imputing the whole questionnaire has to be made from a large group of responding units. Such imputation is equivalent to duplicating questionnaires. Of course, the procedure may be modified to impute within sample clusters or strata; also field staff may be instructed to collect, if possible, some auxiliary information from neighbours or friends, on the basis of which non-responding units could be classified into some broad categories for the purpose of imputation.

It is at this level that one sees the similarity between weighting and imputation. Duplicating a questionnaire to adjust for a missing unit is equivalent to giving that questionnaire an extra weighting factor of 2. From the viewpoint of sampling error, adjusting for non-response by an estimation procedure applied at the cluster or stratum level is preferable to duplication of individual questionnaires. The only reason for using the latter procedure would

be to maintain a sample design that is self-weighting at the stratum level. This is accomplished by actually duplicating the computer record for the selected questionnaire, rather than giving it an extra weight of 2.

Operational simplicity and flexibility of procedures for automatic imputation may make them very attractive. However, careless use of such procedures can result in serious distortions in survey results, or at least in changes the effects of which are not fully understood. The procedures may also smooth over shortcomings of the basic data and hence convey a misleading impression of data quality to the user.

Survey reports should include as much information as is available on the levels of total non-response and of partial non-response by item, as well as on the estimation and imputation procedures used to adjust for non-response. If public-use micro-data files based on the survey are to be distributed, it may be desirable to "flag" all imputed records and items so that users can distinguish these from items actually reported in the survey.

G. Additional sources of information

The literature on survey methodology includes many articles and reports on various aspects of non-response. Several of these have been referred to in this chapter and are identified in full in the bibliography. Additional items, for those who are interested, may be found in the sources described in Chapter I, Section D.

A major new work on non-response is expected to be published in 1982. In 1977, the Committee on National Statistics, which operates under the United States National Academy of Sciences,

appointed a Panel on Incomplete Data, composed of distinguished statisticians from government and universities. The mandate of this group was to review the sources of incomplete data in surveys and the current "state of the art" in dealing with incomplete data, and to make recommendations for coping with this problem which is common to survey research and many other scientific endeavours.

The final report of the Panel on Incomplete Data is expected to be published, in 3 volumes, by Academic Press in 1982. The report will have 5 sections:

i. The report proper, with recommendations.

ii. A section on examples of surveys emphasizing the treatment of incomplete data.

iii. A section on theory.

iv. The final proceedings of a symposium on incomplete data, held in Washington, D.C. in 1979. (A preliminary version of these proceedings was published by the U.S. Social Security Administration in 1979.)

v. Two bibliographies, one of which is annotated.

Sections iii and v are expected to be included in the same volume.

This publication promises to be an authoritative reference work, for many years to come, on the sources and treatment of incomplete data in surveys.

CHAPTER V

RESPONSE ERRORS

A. Introduction

Chapters III and IV were about coverage errors and errors of non-response which occur when no information is obtained for some items and/or for some units of observation. A different kind of error occurs when information is obtained but is incorrect. These errors, which are the subject of this chapter, are called response errors.

The boundary between response errors and errors of non-response was discussed in section A.1 of Chapter IV. If an incorrect response of zero is obtained for, say, a person's wage and salary income, this will be considered a response error. If no response is obtained for the item, this is an error of non-response.

Response errors, as defined here, occur in the data-collection phase of a survey, and are distinguished from errors which occur in the data-processing phase. Sometimes it may be difficult to say exactly where data collection leaves off and data processing begins. It is asserted here, somewhat arbitrarily, that the data collection phase is complete when the survey questionnaire leaves the hands of the interviewer for the last time. The phrase "for the last time" takes into account that, in some surveys, a field or office reviewer may decide that some questionnaires require further respondent contacts in order to correct apparent errors.

The use of the term "response error" must not be taken to mean that all of these errors are the fault of respondents. The outcome of the data collection process depends on all persons who take part in it, starting with the designers of the data collection forms, manuals and procedures, and including interviewers, trainers and field supervisors as well as respondents.

One common source of response errors is obviously the inability of respondents to provide the desired information. Respondents' co-operation with statistical surveys cannot go beyond the limits imposed by their knowledge. For some respondents, information such as their age in completed years, the size of their holdings in hectares, or their branch of industry

may have no precise meaning. Respondents' social backgrounds and levels of education are clearly very important factors in this respect.

Even when respondents are or have been aware of certain facts, their ability to recall or report those facts at the time of the interview will depend upon the salience of those facts in their minds. This can be so, especially if the facts concerned relate to the distant past in their lives. For example, it is well-known that some older mothers are not able to report accurately even the number of children they have had. Even major items of expenditures incurred some time ago may not be recalled during an interview. Matters which have lesser significance for most people, such as small expenditures, minor ailments, or hours worked on a particular day cannot be recalled accurately by most people more than a few days or weeks after the event.

Respondents sometimes purposely report certain information incorrectly to protect their dignity, prestige, or simply to conform to what they think is appropriate. A study in England noted that "It is a common place for expenditure surveys to produce under-estimates of consumers' purchases of alcoholic drinks and tobacco. From our own survey it would appear that Cambridgeshire households' expenditure on these two items is for alcoholic drinks more than 50 per cent, and for tobacco 30 per cent below what we estimate to be the national average. We can be fairly sure that a good deal of this under-estimate results from people's unwillingness to record just how much they spend on these two items" (Cole and Utting, 1956). Similarly, an extensive study on the quality of statistical data notes that "It is well-known that women do not like to disclose their age and often declare themseves younger, while young people often declare themselves older. Illiterate people report that they are able to read and write. Some people raise the level of their education, others the grade of their occupation. This is how medical assistants become medical practitioners, operators become foremen, bank employees become directors, etc. Similarly, people often exaggerate the salaries they receive and the rents they pay, the price of their food or clothing, the amount of money spent on books, concerts, theaters, and other items showing their cultural interests..." (Zarkovich, 1966). In his report on an experiment, McCord (1951) showed how far some respondents go in

trying to appear well-informed and to build up their prestige in the eyes of the interviewer. In the experiment, respondents were asked if they had ever heard about some particular magazines, writers and pieces of legislation, etc., which in fact never existed. A very large proportion of the respondents answered "Yes".

Some of the above are, of course, extreme examples, and most respondents in most circumstances probably try to provide the most accurate information they can. Even though many surveys involve a fairly taxing and time-consuming task for the respondents, the degree of respondent cooperation in interview surveys, particularly in most developing countries, has generally been very encouraging.

The interview is an interactive process between two (or sometimes more) individuals. A skilled interviewer can help the respondents to provide accurate responses. At the same time, interviewers can be a source of error by failing to put the question clearly, by influencing respondents to answer incorrectly, or by misrecording correct responses.

Above all, the incidence of response error depends upon the design of the survey operation, and the nature and complexity of its content. The system of concepts and definitions, the instructions to the field staff, the content, design and layout of the questionnaire, the wording of questions, and the procedures for training and supervision of the field staff are all major determinants of the level of response error.

Section B of this chapter gives a definition of response error, and explains how it may be considered as consisting of two components: response bias and response variance. The decomposition of response error into these two components is of considerable practical significance. It allows for a more specific identification of the sources of errors, and for development of methods for their measurement and control.

Section C provides several illustrations of response bias in survey results. The identification and measurement of response bias, generally speaking, requires comparison of survey data with more reliable data from some external source. However, in some situations, particularly in the demographic field, past experience and theoretical knowledge about the expected patterns make it possible to identify gross biases or distortions

in the data from a direct examination of the survey results.

Section D is about response variance. Response variance can be further decomposed into two components: uncorrelated (or simple) response variance and correlated response variance. Procedures for estimating each component of response variance are described and examples are given.

B. Basic concepts

Response error will now be defined more formally in terms of the deviation of the value of an item actually obtained for a particular unit of observation from the "true" or "correct" value of that item. For aggregate items such as income or expenditures, the response error for a particular response can be zero, positive or negative. A positive value would indicate a response in excess of the true value and vice-versa. For attributes, response error in the aggregate might be represented by a 2x2 table. For example, responses on labour force status might be tabulated as follows:

	Correct Status	
Status in survey	In labour force	Not in labour force
In labour force	a	b
Not in labour force	c	d

In the table, the net effect of response error for the category "in labour force" would be

$$(a + b) - (a + c) = b - c$$

and for the category "not in labour force" it would be

$$(c + d) - (b + d) = c - b$$

A gross measure of response error for the labour force status item would be b + c, or in relative terms:

$$\frac{b + c}{a + b + c + d}$$

For a categorical variable with more than 2 possible values, such as age by 5-year intervals, an analogous table and measures of response error can readily be constructed.

The above definition of response error assumes the existence of a "true" or "correct" value of the item for each unit of observation. In practice, especially for complex variables such as income or labour force status, definitions are not likely to be sufficiently detailed and precise to cover every possible situation explicitly. The existence of a "true value" is even more debatable when survey questions concern the attitudes, beliefs and perceptions of respondents. For such items, it may be necessary to seek a different model of response error. However, most items included in household surveys conducted by government statistical agencies can be characterized as factual or objective, and for such items it is often possible and useful to estimate and analyze survey response errors as defined above.

The idea introduced in Chapter II of dividing non-sampling errors into biases and variable errors is especially useful in connexion with response error. Responses to individual items can be affected in many ways by transient factors such as some preoccupation of the respondent at the time the interview is conducted, the personal characteristics of the interviewer chosen to conduct the interview, or a transposition of digits in recording an answer. They can also be affected by the essential survey conditions that are inherent in the design of the survey and the general setting in which it is conducted.

The variable error component of response error, resulting from transient factors which affect individual responses in a more or less random fashion, is called response variance. The bias component, which depends on the essential survey conditions, is called response bias. Using the idea of independent trials introduced in Chapter II, the response variance is a measure of the variability, for a fixed sample, of the responses obtained in a series of trials carried out under the same essential survey conditions. The difference between the average or expected value

of these responses and the true value is the response bias.

The concept of response variance is relevant even for questions on attitudes, perceptions and beliefs of respondents. Answers to such questions can clearly be influenced by transient or random factors. The extent of such variation would depend on whether responses reflect strongly held opinions on matters that are of considerable importance to respondents, or whether they represent unpremeditated reactions to questions that have little direct interest to them.

It is also possible to conceive of response bias for survey variables of this type, even though "true values" cannot readily be defined for particular individuals. Suppose, for example, that it were desired to survey public opinion of a government's economic policies and programmes. Consider two survey procedures. Under procedure A, interviews would be conducted in the presence of local officials and respondents would be asked to sign their questionnaires. Under procedure B, all interviews would be conducted in private, and no names would be recorded on the questionnaires. One might expect the results using procedure B to be, in some sense, closer to the truth.

The discussion and examples in the sections that follow will be concerned almost entirely with response errors for objective items; however, readers should be forewarned that the avoidance or control of response error in collecting data on subjective phenomena requires special care and attention. Since "true values" cannot strictly be said to exist for subjective items, the concept of bias is usually replaced by the concept of validity, generally interpreted to mean the closeness of a survey expected value to some ideal measurement for the item. Various kinds of validity have been defined for subjective items, including content or face validity, construct validity and predictive validity. A useful discussion of these concepts is given by the Panel on Survey Measurement of Subjective Phenomena (U.S. Committee on National Statistics, 1981, pp. 24-25).

C. Some examples of response bias

This section contains several illustrations of response bias in surveys. The illustrations have been selected from a large number of examples available from surveys conducted in different countries of the world. Sub-section 1 contains examples of bias in the reporting of age, a topic that is included in nearly all household surveys. Sub-section 2 presents some examples of bias in the reporting of vital events - births and deaths.

Failure to remember and report past events correctly is a major source of response bias. Several examples are presented in sub-section 3. Special problems are encountered in longitudinal or panel surveys, leading to the so-called panel or rotation group bias, examples of which are presented in sub-section 4. Examples of response bias arising from the use of proxy or household respondents are given in sub-section 5. Further examples of response bias affecting a variety of topics and associated with various aspects of survey design appear in sub-section 6.

It is easier to demonstrate the existence of response bias than it is to identify its sources or causes and to decide what steps can usefully be taken to eliminate or reduce it. Sometimes the reasons for response bias are fairly obvious, but often a carefully designed methodological experiment will be the only way to pinpoint the sources and to determine with confidence what changes in design and procedures are likely to be most effective in controlling bias.

1. Response bias in age reporting

Systematic errors in age reporting represent probably the best-known and best-documented type of response bias. They are of particular importance because the analysis of data by age group is important to the understanding of most phenomena studied in household surveys. Because so much information on the distribution of population by age is available from censuses, vital registration and other sources, it is usually relatively easy to spot the effects of response bias in age reporting.

Many of the examples which follow are taken from censuses rather than sample surveys. However, the factors associated with bias in age reporting are likely to exist in either case, and the effects are likely to be similar.

A common form of error in age-reporting is preference of respondents for numbers ending in certain digits such as 0 and 5. Table 5.1 shows an index of age preferences at terminal digits 0 and 5 calculated from the results of a 1976 fertility survey in Indonesia (Central Bureau of Statistics, 1978). The figures show, separately for males and females, the ratio of population reported at a given age (such as 15, 20, 25, etc.) to the average population with ages reported in the five years centered on that age (average of ages 13-17, 18-22, 23-27 etc.). For example, the index for age 25 is calculated as the population reported at age 25, divided by one-fifth of the total population who reported their ages in the range 23 to 27. In the absence of response bias, the index for any age would be close to 1.0.

Table 5.1

Indexes of age preferences for terminal digits 0 and 5: Indonesia, 1978

Age	15	20	25	30	35	40	45	50	55	60
Index of heaping:										
male	1.05	1.37	1.64	2.56	2.10	2.05	2.07	1.88	1.68	2.29
female	1.09	1.43	1.75	1.85	1.60	1.65	1.49	1.03	1.24	2.47

Another illustration of the same phemonenon is provided by the following data (Das Gupta and Mitra, 1958) from the Indian National Sample Survey. Table 5.2 shows the full age distribution by single year of age. The figure for any particular age (for example 17) is shown as a percentage of the total population in the corresponding ten-year range (for example 10-19). In addition to heaping at numbers ending with 0 and 5, the data also show a pronounced preference for ages ending with 2. The distortion increases with age.

Table 5.2

Per cent distribution of persons within 10-year age ranges: Indian National Sample Survey

End digit	Ten-year age range							
	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79
0	10.6	15.2	18.7	26.6	31.4	38.6	49.2	51.0
1	10.1	8.1	6.1	4.6	4.5	5.0	4.6	3.8
2	10.9	14.4	13.5	11.8	9.5	9.2	8.9	7.7
3	10.8	8.7	6.9	5.8	5.0	4.5	3.8	4.4
4	9.7	9.4	9.1	5.8	5.6	4.6	4.7	3.0
5	10.9	10.4	18.4	20.4	23.1	19.6	18.7	19.7
6	10.3	11.3	8.4	8.3	6.2	5.0	2.7	3.3
7	8.9	5.8	5.4	4.5	4.9	3.8	2.1	2.7
8	10.2	11.7	9.8	8.4	6.8	6.5	3.6	3.0
9	7.6	5.0	3.7	3.8	3.0	3.2	1.7	1.4
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

This kind of bias arises primarily because many people, particularly in developing countries, do not know their exact ages. It has generally been found that, in given settings, the method of data collection or the questionnaire design can have only a limited effect on age-reporting bias. In several countries, similar patterns of age-heaping have been found in censuses and household sample surveys, even though other major distortions have been less marked in the latter. Digit preference is a cultural phenomenon, as is illustrated by the following United States Census data (Seltzer, 1973).

Table 5.3

Per cent of population with ages ending in each digit 0-9, United States: 1880-1960

(Percentages represent indices of digit preference and were obtained by the Myers' blended method, using starting ages 23-32 and ending at age 99 in all cases.)

Digit of age	1880	1890	1900	1910	1920	1930	1940	1950[a]	1960[b]
All digits..	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0	16.8	15.1	13.2	13.2	12.4	12.3	11.6	11.2	9.9
1	6.7	7.4	8.3	7.7	8.0	8.0	8.5	8.9	9.9
2	9.4	9.7	9.8	10.2	10.2	10.3	10.4	10.2	9.9
3	8.6	9.1	9.3	9.2	9.4	9.4	9.6	9.7	9.8
4	8.8	9.0	9.5	9.4	9.4	9.6	9.7	9.7	10.1
5	13.4	12.3	11.3	11.5	11.3	11.2	10.7	10.6	10.3
6	9.4	9.6	9.4	9.6	9.7	9.6	9.6	9.8	9.9
7	8.5	8.9	9.3	9.1	9.4	9.3	9.6	9.7	10.1
8	10.2	10.4	10.2	10.7	10.6	10.5	10.3	10.2	9.8
9	8.2	8.5	9.7	9.4	9.6	9.8	10.0	10.1	10.3
Summary index[c]	10.4	7.8	4.7	5.6	4.5	4.3	3.0	2.2	0.8

[a]Based on a 20-percent sample.
[b]Based on a 25-percent sample.
[c]The index is one-half the sum of the deviations from 10.0 per cent, each taken without regard to sign.
Source: Shryock and Siegel, 1971, vol. 1, p. 208.

The index of digit preference shown in Table 5.3 is called the Myers index. Its expected value at any digit in the absence of age-heaping is 10.0. The difference from this expected value indicates the relative heaping or deficit for the specified digit. The summary index at the bottom provides an overall index of heaping for each census. This table shows the effect of historical developments on response bias. The gradual establishment of vital registration and the need for people to know and use their birth dates in filling forms and documents for school registration, employment, banking, etc., resulted in the virtual elimination of digit preferences in the reporting of age by the time of the 1960 census.

Table 5.4 shows variations in the pattern of age-reporting of children in censuses for several African countries (Ewbank, 1980). The figures shown are ratios of the number of persons reported at each age to the number reported at the next higher age. In an expanding population and in the absence of

recent marked fertility decline, these ratios would be somewhat greater than 1.0 and would show a smooth pattern if no age-misreporting occurred. With a few exceptions the ratios for age 10 indicate substantial heaping at that age. Many distortions occur at the lowest ages. Several countries, including Ghana, Liberia, Senegal, Swaziland and Togo have very high ratios for age 0 (less than 12 months), while other countries, such as Botswana, Southern Rhodesia (Zimbabwe) and Uganda have very low ratios.

Table 5.4

Ratios of persons reported at age X to persons reported at age X + 1; selected African countries

Country and year		Ages for which ratio is shown													
		0 1	1 2	2 3	3 4	4 5	5 6	6 7	7 8	8 9	9 10	10 11	11 12	12 13	13 14
Botswana	1964	0.56	0.91	0.97	1.00	0.95	1.09	1.00	1.03	1.02	1.01	1.63	0.62	1.55	0.91
Ghana	1960	1.31	0.82	0.90	1.11	1.12	0.94	1.27	1.00	1.25	0.89	1.68	0.66	1.29	1.11
Kenya	1962	1.09	0.62	1.11	0.96	1.25	0.88	1.30	0.78	1.54	0.67	2.26	0.46	1.82	1.24
Kenya	1969	0.86	0.89	1.11	0.99	1.09	0.97	1.20	0.90	1.27	0.86	1.54	0.69	1.36	1.09
Lesotho	1966	0.89	1.07	0.91	1.03	0.92	1.04	1.11	0.93	1.08	0.86	1.38	0.71	1.17	0.98
Liberia	1962	1.48	0.55	0.88	1.08	1.11	1.05	1.28	0.91	1.25	0.92	2.17	0.53	1.58	1.15
Mauritius	1952	0.99	0.93	1.09	1.10	1.04	1.19	0.95	1.01	1.30	0.97	1.11	0.87	1.15	1.06
Mauritius	1962	1.11	0.94	1.01	0.99	1.01	1.04	1.04	1.05	0.96	1.02	0.99	1.03	1.20	1.02
Nigeria	1963	0.80	0.99	0.89	1.09	1.07	1.14	0.89	1.27	0.85	0.96	2.05	0.59	1.58	0.90
Senegal	1961	1.29	0.82	0.91	1.03	1.07	1.09	0.92	1.22	1.30	1.11	1.12	1.04	1.06	1.19
S.Rhodesia (Afr.Pop.)	1969	0.33	1.16	1.09	1.02	0.91	1.12	0.96	1.21	0.78	1.21	1.03	1.17	0.93	0.79
SW. Africa	1946	0.84	0.96	0.79	0.99	1.02	0.97	1.13	0.93	1.23	0.83	1.52	0.75	1.43	1.03
Swaziland	1966	1.20	0.85	0.97	1.08	0.96	1.02	1.11	0.94	1.14	0.94	1.34	0.77	1.18	1.07
Tanzania	1967	1.11	0.84	0.87	1.05	1.11	0.98	1.13	1.07	1.16	0.95	1.76	0.64	1.35	1.06
Togo	1958-60	1.41	0.86	0.72	1.32	0.96	0.99	1.16	1.22	1.21	1.19	1.70	0.71	1.42	1.12
Tunisia	1966	0.96	1.00	1.07	1.06	1.09	0.95	1.10	1.03	1.08	0.97	1.10	0.97	1.11	1.05
Uganda	1969	0.69	0.99	1.05	1.04	1.12	0.99	1.22	0.93	1.24	0.88	1.54	0.68	1.32	1.03
Union of South Africa	1946	0.85	0.94	0.99	1.01	1.07	0.92	1.23	0.83	1.30	0.78	1.62	0.63	1.38	1.07
Zambia	1969	1.08	1.05	1.04	1.03	1.02	1.01	1.06	1.14	0.93	1.10	1.14	1.02	1.16	1.03

Another example of age-misreporting, which illustrates that not all cultures have the same concept of age is reported by You Poh Seng (1959). Table 5.5 shows that in successive censuses in Singapore the proportion of the population classified in the age-group 0-4 was consistently lower than the proportion aged 5-9. This is contrary to expectation, and the explanation lies in the interpretation given to age in the Chinese culture:

"According to the traditional method of age counting a Chinese is one-year old at birth, and thereafter becomes a year older at every Chinese New Year. Since the Chinese New Year is based on a form of lunar calendar, New Year's Day varies from year to year, but normally it falls round about February. An extreme case would be the following: a child is born, say, a week before the Chinese New Year. It is one year old at birth, and one week later, on the occasion of the New Year, it becomes two years old, whereas strictly speaking it is only one week old".

This example also illustrates the point that "response" bias is not always the fault of respondents. Those responsible for designing the censuses apparently were not sufficiently aware that some respondents had a concept of age which differed from the standard demographic concept. Redesign of the census questionnaire and procedures to take account of this difference could probably have controlled this source of bias in age reporting. The Chinese method of counting age is not as complicated as it may sound: age in that system is defined simply as the number of (Chinese) calendar years in which the person has lived since birth.

Table 5.5

Census counts of persons in age groups 0 to 4 and 5 to 9:

	Census						
Age group	1881	1891	1901	1911	1921	1931	1947
0 - 4	5.4	4.8	5.8	5.5	6.6	8.8	12.2
5 - 9	6.2	5.2	6.1	6.5	7.5	9.4	12.6

The response bias in the above examples was so marked as to be obvious without any need for specific comparison with external data or for sophisticated analysis. More elaborate demographic analysis (using stable age distributions and model stable populations, for example) have indicated broad similarities in patterns of age reporting within groups of countries. One such analysis reports that a common pattern of "major distortions is found to characterize surveys in most of Africa and southern Asia while another

pattern of substantial age-heaping, but relatively minor distortions (otherwise), is characteristic of Latin America and the Philippines. However, among the male populations in Africa - South Asia, the similarities in distortion from country-to-country are less for males than for females, and the distortions themselves appear larger on the average. In the Latin American populations, the distortions in the female age distribution are almost always larger than in the male..." (United Nations, 1967).

General statements about patterns of age-misreporting have also been made on the basis of generalizations from experience and previous studies. Caldwell and Igun (1971) note that people in Nigeria "...exaggerate the number of females aged 20-44 and the number of older males, and ... understate the number of persons aged 0-9 and females aged 50-59. It was shown that amongst females aged 15-24, there is a significant tendency for the ages of the unmarried to be understated and those of the married ... to be overstated". Similarly, Ewbank (1980) notes that a typical characteristic of the pattern of age-misreporting in Africa and South Asia is "... a systematic form of individual age-misreporting over a broad range" and "pronounced age heaping by single years". In this pattern there is "... a surplus at 5-9, and a deficit in the adjacent age intervals (10-14 and 15-19) followed by a surplus in the central age of child-bearing (25-34)". As against this, the Latin-American pattern involves preferences for the age groups 25-29 and 35-39 over the groups of 30-34 and 40-44.

Obviously, a major cause of age-misreporting is that many respondents do not know their exact age. This is illustrated by the following example from the Indian National Sample Survey (Das Gupta and Mitra, 1958). Table 5.6 shows the sources of age data in cities, towns and villages in India. The incidence of "hearsay, guess or eye estimate" increases with age. A somewhat surprising result is that the form of age-reporting tends to be more precise in villages than in cities.

Table 5.6

Bases for age reporting by age and type of area:
Indian National Sample Survey

Age groups	Area	Type of evidence: Hearsay, guess or eye estimate	Related with definite or approximate age or events	Definite statement of year of birth	Birth certificate or other document	Total
0 - 6	City	27.1	51.0	21.9	-	100.0
	Other urban	24.6	42.0	31.9	1.5	100.0
	Rural	18.9	34.0	44.2	2.9	100.0
7 - 16	City	41.7	41.7	16.6	-	100.0
	Other urban	33.9	49.3	14.8	2.0	100.0
	Rural	36.7	45.5	15.1	2.7	100.0
17 and over	City	65.8	19.8	13.6	0.8	100.0
	Other urban	53.6	39.0	5.5	1.9	100.0
	Rural	46.2	44.4	6.6	2.8	100.0

Table 5.7 shows data for several developing countries on the form in which mothers reported the ages of their own children (Chidambaram, Cleland and Verma, 1980). Generally, mothers are able to report both the month and year of birth for their most recent child more frequently than for other children. There is a substantial range between countries: all women in the Republic of Korea were able to report the dates of birth of all their children, while only 12% in Bangladesh could do so.

Table 5.7

Bases for ages of children as reported by their mothers in fertility surveys: selected countries

Country	Per cent reporting date of: Last live birth as: Month and year	Year only	Years ago	All live birth as: Month and year	Year only	Years ago
Bangladesh	33	4	63	12	3	85
Fiji	96	4	0	86	14	0
Indonesia	56	8	36	46	11	43
Jordan	84	6	10	67	11	22
Korea, Republic of	100	0	0	100	0	0
Malaysia	95	5	0	86	14	0
Pakistan	90	10	0	80	20	0
Sri Lanka	83	12	5	73	18	9
Thailand	90	8	2	84	13	3
Guyana	93	3	4	91	4	5
Jamaica	93	-	7	91	-	9

NOTE: "-" less than one per cent.

2. Response bias in reporting vital events

A frequent goal in demographic surveys is to obtain a full record of vital events such as births and deaths. It is often found that some respondents do not report all the events; such omissions adversely affect analyses of the data. The level of omissions is thought to depend, among other things, on the type of the event, and the method by which the data are obtained. For example, it may be argued that if women are asked to report the number of births they have had, certain categories of births are more

likely to be omitted than others, among them births occurring a long time ago and, in many cultures, females births and births of children (especially girls) who died early in childhood. One might also expect the level of omissions to be lower in a relatively small-scale intensive survey than in a large-scale survey or census. Table 5.8 compares data on live-births as reported by mothers in small-scale intensive surveys with similar data from large-scale surveys or censuses conducted in the same countries (Chidambaram, Cleland and Verma, 1980).

Table 5.8

Mean number of children ever born by age of mother: comparisons for alternate collection methods in selected countries.

Country	Collection method	Age of mother						
		15-19	20-24	25-29	30-34	35-30	40-44	45-49
Bangladesh	(a)	0.8	2.4	4.2	5.7	6.7	7.1	6.7
	(b)	0.6	1.9	3.5	4.9	5.9	6.2	6.1
Nepal	(a)	0.3	1.4	2.9	4.1	5.1	5.5	5.7
	(b)	0.3	1.1	2.2	3.1	3.7	4.0	4.0
Korea (Rep.)	(a)	0.5	1.0	2.0	3.4	4.4	5.1	5.8
	(b)	0.5	1.0	2.1	3.5	4.5	5.3	5.6

(a) Small-scale, intensive survey.

(b) Large-scale survey or census.

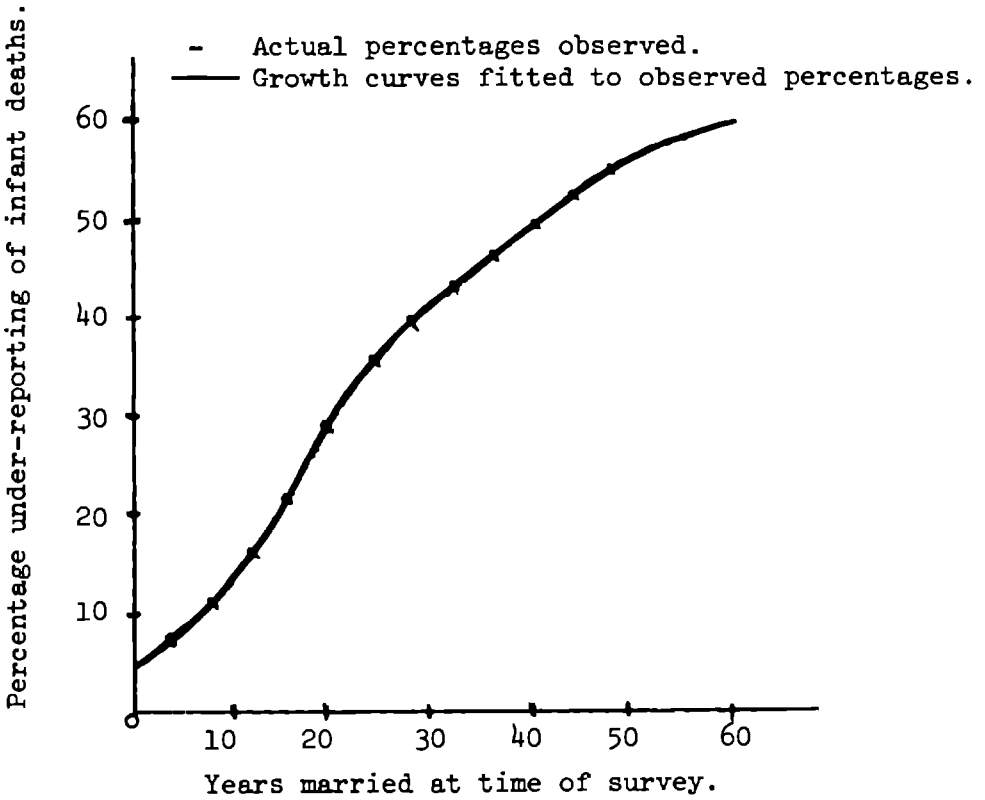
In the absence of sudden changes in the demographic situation, one would expect the mean number of children every born to increase with the mother's age. The data for Bangladesh from both sources show a decrease for age group 45 to 49, which suggests omission of births by respondents in that age group. For both Bangladesh and Nepal, a substantially lower level of omissions is indicated in the more intensive survey, while for

the Republic of Korea, the differences between the two collection methods are minimal. This example shows the relevance of both the social context and the survey methodology in determining the level of response bias. It also shows how plausible hypotheses can be used to detect probable biases in the absence of direct external evidence.

The level of omission can also depend upon the nature of the event. Events which are unpleasant or painful may be omitted more often than those which are not. The study just mentioned notes that "while a carefully designed and implemented live-birth history (in a survey of retrospective fertility) can be expected to net a fairly complete coverage of live-births ... /such/ approach has not succeeded as regards the coverage of non-live births". An illustration of omissions of children who have died is provided by Figure 5.1, which is based on estimates from a survey in India of the percentage under-reporting of infant deaths plotted against the number of years the mother had been married at the time of the survey (Das Gupta, Som, Majumdar and Mitra, 1955).

Figure 5.1

Under-reporting of infant deaths by number of years married: Indian National Sample Survey



Another illustration from India provides evidence that older women were likely to omit reporting of female children (Mahalanobis and Das Gupta, 1954). Among the earlier marriage cohorts, i.e. among women who had married many years ago, there was an increasing tendency to omit some female births. The figures show the sex ratios based on reported births, i.e., number of male births reported per 100 female births reported. Higher values of the sex-ratio indicate higher levels of omission of female births relative to male births.

Table 5.9

Sex ratios for reported births, by year of marriage: Indian National Sample Survey

Period of marriage	Number of couples	Sex ratio
Up to 1909	910	147
1909 - 1919	945	127
1920 - 1929	1 459	143
1930 - 1939	2 757	108
1940 - 1945	2 204	107
1946 - 1951	1 001	107

Table 5.10, which is based on data from the 7th round of the Indian National Sample Survey, also provides evidence of response bias in reporting births (Mazumdar, 1962). When the births occurring in the 12 months preceding the survey are distributed by reported month of occurrence, there is a sharp decline as the number of months elapsed between the birth and the survey increases.

Table 5.10

Distribution of births by number of months between reported month of occurrence and survey: Indian National Sample Survey

Month of birth prior to the survey	Percentage
1st	11.25
2nd	9.62
3rd	9.44
4th	9.18
5th	8.97
6th	9.15
7th	8.40
8th	7.84
9th	7.51
10th	6.97
11th	6.43
12th	5.24
	100.00

The exact nature of the bias is not obvious, however. Although omissions do tend to increase for events occurring farther back in time, it seems unlikely that respondents would fail to report more than half of the births that occurred only 12 months prior to the survey. Another possible explanation for at least some part of this bias would be that some respondents are unable to report month of birth accurately and that there was a tendency to err by reporting births as having occurred more recently than they actually did. This example, then, leads to the subject of recall or memory bias in surveys, which is treated in the next subsection.

3. Recall lapse: omissions of events, displacement of events, incorrect amounts

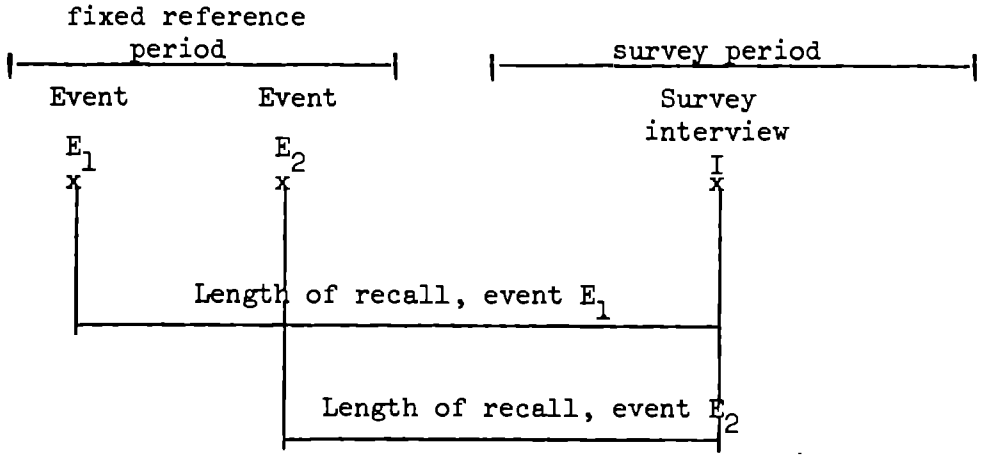
In many surveys, information is collected about certain events that took place during a specified period of time, such as a week, a month or a year. For example, in an income-expenditure survey, data may be collected on income over the past month, and on certain items of consumption, such as

food, over the past week. In a demographic survey, sample households may be asked to report births and deaths during the past year. Such surveys are called retrospective surveys, and the interval for which data on an event or item are sought is called its reference period. The reference period may be the same for all respondents, i.e. fixed in terms of calendar dates, for example from January 1 to December 31 of a given year. Alternatively, the reference period may be defined as a given length of time immediately preceding the interview, for example, the 12 months preceding the date of interview. If respondents in the survey are not all interviewed on the same day, the exact calendar period for which data are collected will vary from one respondent to another. This is referred to as a moving reference period although its duration is constant.

Two concepts related to the reference period are the survey period and the length of recall. The survey period is the time interval during which fieldwork for the survey is done. Depending upon the nature of the survey, its size, and the number of field staff available, the survey period may be only a few days or weeks, or may extend over several months. The length of recall is the time elapsed between the date of a particular event or transaction that occurred during the reference period and the date on which a respondent is asked to recall it. If a moving reference period is used, the maximum length of recall will be equal to the length of the reference period. However, if a fixed reference period is used for all respondents, the maximum length of recall can exceed the length of the reference period. For example, if an event occurred at the beginning of a one-week fixed reference period, and the respondent were interviewed one week after the end of the reference period, the length of recall for that event would be two weeks.

Fig. 5.2

Relationship between reference period, survey period and length of recall



In some surveys, data on the timing of events may be obtained in a different way. For example, in a survey on migration one may ask for the date of the most recent move, rather than whether there was a move in a specified reference period. In a demographic survey, one may ask for the date of the last birth to a member of the household, rather than whether there was a birth during a specified reference period. In this form of questioning, the length of recall may vary substantially from one respondent to another, and the survey and the reference periods are not relevant to the recall process. The problems of recall of the relevant events and of their timing in the two systems of questioning may be quite different in practice. The following paragraphs will be largely concerned with the more commonly used fixed-length reference period.

There are two types of recall errors which affect estimates of the number of events: (i) omission (and occasionally over-reporting) of events, and (ii) errors in location in time of the events reported. These two types of errors are conceptually distinct, though in practice it may be difficult to disentangle their

effect. So far as omission is concerned, it is related to the length of recall as well as to salience of the event to the respondent. The general experience that omissions increase with time may suggest that the length of recall and hence the reference period should be kept as short as possible. However, this reduces the number of events for which information is obtained in each interview, which has the effect of increasing sampling error for estimates based on a fixed number of sample interviews. Thus, in a one-time demographic survey concerned with relatively rare events such as births and deaths, it may not be cost-effective to use a reference period much shorter than one year. In surveys of household consumption, it is usual to use longer reference periods for the less frequently purchased and/or more expensive items (which presumably are also more salient to the respondent, and hence less subject to omission), and to use shorter periods for items consumed more frequently.

The second type of error may be termed event displacement. Some events may be thought by the respondent to have occurred earlier than they actually did (backward displacement), others may be displaced in the forward direction (sometimes called telescoping). It is also useful to distinguish displacement within the reference period (which does not affect the number of events reported in the reference period), and displacement into or out of the reference period.

Error occurs when events are displaced into or out of the reference period. These are sometimes called end effects. Various methods are used in surveys to bound or close the ends of the reference period to avoid such displacements. For example, if the interviewer calls on a respondent on December 20 and asks about events between December 8 and December 14 (fixed reference period of one week) both ends of the reference period are "open"; however, if the questions asked relate to the week immediately before the interview, then only the earlier end is open. If the interviewer also visited the household a week ago to obtain similar information, and the present interview is used only to list the additional items reported

since the last visit, then both ends of the current reference period are closed or bounded. In other words, by design of the study, no events are likely to displaced into or out of the reference period. (The price paid, of course, can be the increased cost of multiple visits.) Similarly, if the reference period is the whole lifetime of the respondent (for example, if a woman is asked the number of children ever born), both ends are closed by definition.

Event displacement within the survey reference period can also be a source of response bias if estimates of the number of events for shorter time intervals are wanted. For example, if the reference period is one year, but respondents are asked to report the month of occurrence of each event, the estimates of events by month may be biased even if those for the full year are not. The distortions in the distribution of births by month that were shown in Table 5.10 were probably due at least in part to this type of bias.

Retrospective surveys are also used to collect information on amounts for items such as income and expenditures. Errors in reporting amounts, like omissions of events, are likely to be affected by length of recall and by the salience of the particular transaction involved. Sometimes errors in amounts may be only variable errors, i.e., their effects will average out and not result in bias; however, as will be shown, this is certainly not always the case. Considerations of "social desirability" may also lead to bias in reporting amounts, as mentioned at the beginning of this chapter in connexion with estimates of purchases of alcoholic drinks and tobacco.

While a shorter reference period and hence shorter average length of recall may result in fewer omissions and more accurate reporting of amounts, it may at the same time be subject to relatively more serious end effects. Hence, a shorter reference period, if unbounded, does not necessarily result in lower relative bias. In addition, of course, there is the larger sampling variability of the results when shorter reference periods are used. Hence, the determination of the best reference period for a particular topic requires

empirical information about the relative biases associated with reference periods of various lengths and kinds. The joint effects of response bias and sampling variance must then be evaluated to reach a decision.

A few additional considerations in the choice of the reference period may be mentioned. If there is a tendency among respondents to heap certain events at a point which coincides with the boundary of the reference period, more serious end-effects can be expected. For example, if there is a marked tendency for respondents to think of children born, say, 9 to 15 months ago as if they were born "one year ago", then the use of a one-year reference period in demographic surveys may be inappropriate, unless some probing is done to minimize the incidence of such heaping. Secondly, if events are distinguished in respondents' memories in some sort of logical "blocks", then the choice of the reference period should take that into account. For example, salaried people may plan their expenditures for a whole calendar month. Monthly expenditures will be interrelated since their sum is limited by the resources available during the month. In such cases, it may be relatively easy to recall monthly expenditures as a block immediately after the end of the month; however, as the expenditure cycle for the next month starts, the old one may be quickly forgotten. This sort of consideration can also be very important in agricultural surveys, where the "blocking" may be according to agricultural seasons.

Recall lapse or memory bias in retrospective surveys has long been recognized as a significant source of error in surveys, and many analyses and methodological experiments have been conducted to identify and measure its effects. Some of these will now be reviewed.

Several investigations of recall lapse were undertaken in India after World War II. In the first round of the National Sample Survey a reference period of one year was used for collecting data on several items. However, this reference period was found to be unsatisfactory for expenditures on food items, so experiments were done to evaluate

reference periods of one week and one month for these items. (Mahalanobis and Sen, 1954) Table 5.11 shows the data on expenditures (in rupees per month per person) obtained with the two reference periods. For each reference period, data were collected for two independent subsamples which could be compared to obtain an indication of the variability in results due to factors other than the reference period. There were consistent differences between the estimates from the two reference periods for several items.

Table 5.11

Mean monthly expenditures per person for selected items, using reference periods of one week and one month: India

Item	Week		Month	
	Subsample No. 1 (Rs.)	Subsample No. 2 (Rs.)	Subsample No. 1 (Rs.)	Subsample No. 2 (Rs.)
Food grains	9.56	9.47	8.87	8.66
Other food items	5.82	6.21	5.06	5.02
Total food	15.38	15.68	13.93	13.68
Betel leaves, tobacco, intoxicants	0.82	0.82	0.68	0.65
Fuel and light	1.54	1.59	1.22	1.17
Monthly items	1.97	2.34	2.34	2.21
Clothing (annual)	1.62	1.73	1.79	1.65
Other annual items	2.39	2.42	2.42	2.19
TOTAL EXPENDITURE	23.72	24.58	22.38	21.55

As either a one-week or a one-month reference period was feasible for food consumption surveys, additional data, external to the regular surveys, were needed to estimate the biases involved. For this purpose, households in the sample were requested to lay in stocks of certain specific food items to be used in the forthcoming period and to restrict their consumption to those stocks. The quantity of each food item was weighed before use, and then once a week afterwards in one subsample and twice a week in another subsample, to

establish the amounts consumed. Data for the same items were also collected by interviewing the households. Some of the results of the experiment are shown in Table 5.12.

Table 5.12

Average amounts consumed for selected food items, based on interviews and weighing with varying frequencies

Item	Interviewing		Weighing	
	Weekly	Monthly	Once a week	Twice a week
Rice	16.21	15.14	15.61	14.87
Pulses	1.56	1.19	1.42	1.19
Sugar	0.23	0.16	0.16	0.14
Salt	1.27	0.76	0.86	0.82

To the extent that results from the weighings were accurate, the interview data based on a one-month reference period appeared to be less biased, on the average, than the data based on a one-week reference period.

In another experiment on household food consumption in West Bengal, India (Ghosh, 1953), field staff were stationed in all food shops in selected villages to record each purchase and to classify it according to whether or not it was made by village residents. Interview data on consumption of specific items of food were obtained with three different reference periods - one day, one week and one year. The data in Table 5.13 showing weekly purchases of food stuff in rupees per household would indicate that if the estimates obtained from food shops were accurate, a reference period of one year for the interview would be the most appropriate for several items.

Table 5.13

Mean weekly expenditures for selected food items, based on interviews with different reference periods and on observation of purchases in shops: India

Item	Reference period for interview data			Estimates based on daily observation in shops
	Day	Week	Year	
Pulse	0.35	0.31	0.23	0.21
M. oil	0.77	0.62	0.48	0.42
C. oil	0.14	0.19	-	0.14
Salt	0.07	0.11	0.08	0.07
Gur	0.07	0.10	0.07	0.02
Pan-supari	0.07	0.12	0.08	0.03
Tea-leaf	0.07	0.05	0.02	0.01
Tobacco	0.35	0.20	0.13	-
Bidi	0.35	0.37	0.29	0.21

The results from the two studies described above seem to contradict each other. In fact there can be room for doubt of the validity of either: the first experiment created a somewhat artificial situation, while the "objective" estimates in the second are questionable. Nevertheless, they indicate efforts to find objective solutions to the problems of choosing appropriate reference periods.

Table 5.14 shows some results from a test of the collection of expenditure data using reference periods of varying lengths in Ghana (Scott, 1973). For most items, the mean number of reported purchases per household per day decreased significantly as the length of the reference period increased. However, it is by no means certain that the data for the one-day reference period are more accurate, because the reference periods were unbounded at the beginning and forward displacement (telescoping) of purchases could have occurred.

Table 5.14

Number of purchases per household per day, by length of reference period: Ghana, urban sample, 1966

Item	Reference period				Index $\frac{7 \text{ days}}{1 \text{ day}} \times 100$	Significance test P
	1 day	2 days	3 days	7 days		
Kenkey	.643	.521	.480	.447	70	.001
Cassava	.517	.595	.570	.538	104	.05
Garden egg	.408	.458	.437	.419	103	Not sig.
Red pepper	.521	.464	.459	.468	90	.001
Plantain	.517	.489	.486	.441	85	.001
Smoked fish	.727	.721	.709	.648	89	.001
Offal	.101	.076	.102	.089	88	.001
Bread	.193	.105	.137	.108	56	.01
Cigarettes	.227	.216	.214	.183	81	.001
Firewood	.282	.214	.211	.191	68	.001

More explicit evidence of displacement effects is available from a study carried out by the United States Bureau of the Census. Neter and Waksberg (1964) report on a complex multi-factor experiment which was incorporated into a national panel survey of household expenditures on residential alterations and repairs.

Part of the experiment was a comparison of bounded and unbounded reference periods of varying lengths, as follows:

(a) Unbounded, 6 months

(b) Unbounded, 1 month

(c) Bounded, 1 month

(d) Bounded, 3 months

Bounding was accomplished in procedures (c) and (d) by recording prior expenditures in a baseline interview conducted at the beginning of the reference period. This information was used during the interview at the end of the reference period to avoid displacement of expenditures into or out of the reference period. For reference periods of more than one month /procedures (a) and (d)/, respondents were asked

to allocate expenditures to the specific months in which they believed they had been made. Thus, in procedure (d), displacement errors were not likely to occur with respect to the entire 3-month reference period, but could occur in allocating the expenditures to months within that period.

Table 5.15 compares estimates of numbers of jobs and amount of expenditures based on one month recall under procedures (a), (b) and (c). For procedure (a), these estimates were based on data reported for the most recent month during the 6-month reference period.

Table 5.15

Jobs and expenditures, estimated from one-month recall, by procedure: United States, February 1960 to March 1961

Size of job (dollars)	Number of jobs (millions)			Expenditures (millions of dollars)		
	Unbounded recall reference period:		Bounded recall, 1 month ref.period	Unbounded recall reference period:		Bounded recall, 1 month ref.period
	6 months (a)	1 month (b)	(c)	6 months (a)	1 month (b)	(c)
Total	213.9	300.6	215.1	11,748	15,343	9,889
Under 10	130.4	192.3	147.8	934	1,221	910
Under 10	89.4	145.1	112.2	389	588	436
10-19	40.9	47.2	35.5	545	633	474
20-99	58.3	78.3	48.2	2,395	3,308	1,935
20-49	40.7	52.7	34.6	1,212	1,584	1,012
50-99	17.6	25.6	13.5	1,183	1,724	924
100 and over	25.3	30.0	19.2	8,418	10,814	7,044
100-499	19.5	25.3	15.7	4,067	5,475	3,107
500 and over	5.8	4.6	3.5	4,351	5,339	3,936

Taking the estimates from procedure (c) as a standard (most accurate), it appears that the use of a one-month unbounded reference period (procedure (b)) led to considerable displacement of jobs and expenditures into the reference period. The results based on the most recent month

of a 6-month unbounded reference period are closer to the standard, but, except for jobs under $10, there still appeared to be some displacement of events into the most recent month of the reference period.

Table 5.16 compares estimates based on one-month bounded recall (procedure (c)) with estimates based on each of the 3 months of the bounded 3-month reference period (procedure (d)). The columns labelled "first-month", "second month", and "third month" show estimates based on the most recent month, the month prior to that, and the earliest month of the 3-month reference period, respectively.

Table 5-16

Jobs and expenditures estimated from one-month bounded recall and each month of a bounded 3-month reference period: United States, June 1960 to January 1961

Size of job (dollars)	Bounded recall with three-month reference period (d)				Bounded one-month recall (c)
	First-month	Second-month	Third-month	Average	
	Number of jobs (millions)				
Total	133.1	70.7	48.4	84.1	116.3
Under 20	82.7	37.7	27.9	49.4	79.1
20-49	23.0	13.7	8.2	15.0	17.9
50-99	11.0	7.2	4.4	7.6	7.3
100 and over	16.4	12.0	8.0	12.1	12.0
	Expenditures (millions of dollars)				
Total	8,743	5,017	3,400	5,750	5,384
Under 20	573	277	199	349	485
20-49	692	420	246	452	528
50-99	762	488	310	520	496
100 and over	6,717	3,833	2,736	4,428	3,875

Presumbably, (c) and the average of (d) are both free from end effects, and the differences between them result primarily from omissions of jobs and incorrect reporting of amounts of expenditures.

Fewer small jobs were reported when the 3-month period was used; however, total expenditures were higher, perhaps as a result of overestimating expenditures on jobs which were reported or perhaps due to sampling error. Clearly, there was a strong displacement effect in the allocation of jobs and expenditures within the 3-month bounded reference period, even for the largest jobs.

Another example from the U.S. Bureau of the Census comes from the National Crime Survey (Murphy and Cowan, 1976). The design of this survey was such that it was possible to compare estimates of victimization rates (incidents per 1,000 persons) based on bounded and unbounded 6-month reference periods. Selected comparisons are shown in Table 5.17. The unbounded rates are consistently higher, indicating substantial displacement of reported incidents of burglary and other types of personal victimization into the reference period.

Table 5.17

Estimated victimization rates based on bounded and unbounded recall: U.S. National Crime Survey

	Total personal victimization rates			Burglary victimization rates		
Quarter	Bounded	Unbounded	% Difference	Bounded	Unbounded	% Difference
I/74	7.89	11.30	43.2	19.23	27.38	42.4
II/74	8.90	12.31	38.3	22.60	33.34	47.5
III/74	9.38	14.88	58.6	26.85	36.62	36.4
IV/74	9.74	13.29	36.4	23.89	31.40	31.4
I/75	8.55	12.17	42.3	20.65	29.13	41.1

Experiments and analyses of this kind have led to fairly widespread adoption of designs using bounded recall and fairly short reference periods for surveys of income and expenditures and other retrospective surveys. These designs are characterized by the following features (Hansen and Waksberg, 1970): "a) Each household must be interviewed for several consecutive time periods. b) After the

first time the interviews are dependent, that is, the interviewer reminds the respondent of what he reported the last time and asks about subsequent expenditure. c) The first report is not included in the tabulation: it is only used for establishing the basis for future dependent interviews."

The literature on recall in household surveys is extensive. Readers who would like to pursue the subject further may consult some of the following references: Cole and Utting (1956), Ferber (1955), Goldberg (1957), Kemsley and Nicholson (1960), McWhinney and Champion (1974), Nisselson and Woolsey (1959), Som (1968), and U.S. National Center for Health Statistics (1972).

4. Response bias in panel surveys: conditioning effects

Some household surveys use a rotating panel design. Households selected for the sample are interviewed at regular intervals, for example weekly, monthly, or quarterly. Also at regular intervals, a fixed proportion of the households (a panel) is dropped and replaced by a new panel of the same size. Each panel is a separate probability sample of the target population for the survey. To illustrate, the sample for a monthly survey might consists of 6 panels, with one of the panels being replaced each month. The households in each panel would be interviewed in six consecutive months and then dropped from the survey.

When this type of design is used, it is possible to estimate population totals separately from each panel. Except for random variation, the samples on which these estimates are based differ only with respect to the number of prior interviews for the sample households. It has been observed frequently that differences between panel estimates are larger than could be expected solely as a result of sampling variability; in such cases the differences are ascribed to panel or rotation group bias.

Simple comparisons of estimates for the different panels do not determine which estimates are biased, or by how much, or what

causes the bias. However, it is generally accepted that part of the explanation of this phenomenon lies in a conditioning or learning effect in which responses in survey rounds after the first one are affected by the experiences of respondents and interviewers in prior rounds. There are many possible effects. Respondents may become familiar with the structure of the questionnaire and answer questions in ways that they know will not lead to a series of follow-up questions. Interviewers may come to expect respondents to have a certain labour force status and fail to probe for possible changes since the prior interview. As respondents become better acquainted with interviewers they may be more inclined to give responses which they think will be socially acceptable. Conditioning could lead to more accurate responses: when respondents know what is coming in the next interview, some of them might keep notes or take other steps that will help them to report more accurately.

Being in a panel survey may even have a direct effect on the respondent characteristics or behaviour which are being investigated in the survey. Questions on labour force status might cause some respondents not in the labour force to start looking for work. Suppose it were desired to conduct periodic surveys, which included medical examinations, to track the health status of the population. While a panel approach would undoubtedly be more efficient for sampling purposes, its use would be inadvisable because persons in the sample would, in all likelihood, develop different patterns of health treatment and status as a result of their participation in the survey. After the first round, they would tend to become unrepresentative of the target population.

An illustration of panel bias is provided by the monthly Current Population Survey (CPS) of the US Bureau of the Census. In the CPS, each sample of housing units selected is divided into eight panels called "rotation groups". Housing units in a rotation group are interviewed eight times: they are interviewed for four consecutive months, dropped from the survey for eight months, and are

then interviewed for four more consecutive months. Each rotation group is a random subsample of the entire sample and in the absence of bias each rotation group would be assumed to have the same expected value. However, systematic differences which are substantially larger than the sampling variability expected, have been observed. Table 5.18 shows that there is a downward tendency in the estimated levels of employment and unemployment depending on the number of months a panel has been in the sample (Bailar, 1979). The table shows a "rotation group index" which is the proportion of persons having the characteristic of interest in a given rotation group divided by the corresponding proportion for the full sample and then multiplied by 100. The expected value of the index in the absence of bias is 100.0.

Table 5.18

Rotation group indices for selected items: United States Current Population Survey

Months in sample	Males		Females	
	Employed	Unemployed	Employed	Unemployed
1	100.8	105.9	102.0	115.4
2	100.0	100.6	100.3	99.8
3	99.9	100.0	99.6	96.8
4	100.2	100.9	100.3	99.9
5	99.9	101.3	100.2	104.4
6	99.7	97.5	99.0	94.6
7	99.6	95.5	98.9	91.3
8	100.0	98.3	99.7	97.8

The existence of panel bias is clear. For example, the estimate of unemployed males from households in the sample for the first time is 5.9% higher than the average (index = 105.9), while the corresponding estimate from households in the sample for the seventh month is 4.5% lower than the average (index = 95.5). The author notes that the sampling error for each index is about 1.0, so that the pattern observed cannot be accounted for by sampling errors.

It is very likely that the above pattern is, at least in part, due to conditioning. However, the phenomenon is complex and its causes by no means fully understood. Note that the most marked differences from the expected value 100.0 in Table 5.18 are for the households which are in the sample for the first time. Regarding this effect, Hansen and Waksberg (1970) note that:

> "From the time that the CPS was converted to a rotating panel operation, in 1947, the staff of the Bureau has been aware that the results obtained varied by rotation group - or by the amount of exposure of the panel to the survey procedures. The most serious effect was on measures of unemployment, and particularly on some subgroups of the unemployed. However, we have found evidence of the same phenomenon in almost every area where rotating panels are used - including health studies, income reports, expenditures data, and anticipated purchases."

> "The estimates of levels of unemployment differ among the eight panels in a consistent way, for reasons that we do not adequately understand. The sizes of the differences vary over time, so that recently we have observed a range twice as great as the range that had been observed earlier -- possibly a result of changed employment conditions. Also, the patterns vary strikingly for various classes of the population and are more serious for population groups that do not have full participation in the labour force, such as women, and persons aged 16 to 25. These panel (or rotation group) differences cause serious difficulties in certain types of analyses."

A number of other studies have documented the phenomenon in different fields. Pearl (1978) has shown its existence in consumer expenditure data; Mooney (1962) has shown that the problem exists for health data; Woltman and Bushery (1975) for crime victimization data; and Hansen et al (1955) and Bailar (1975) for labour force data. Williams and Mallows (1970) discussed the relationships between panel bias and differing levels of non-response for different panels.

5. Response bias and respondent rules

In most household surveys, interviewers are given specific

rules or procedures to follow in determining which person or persons in a household are acceptable respondents for interviews to collect information about members of that household. In general terms, these rules can vary from a very strict procedure, requiring that each adult member of a household must be interviewed separately, even though several visits to the household might be needed, to a very permissive procedure which allows that the information can be obtained from any "suitable" member of a household or perhaps sometimes even from neighbours. An intermediate approach might be to permit the use of proxy respondents if some members of the household cannot be located after a specified number of call-backs. These are general approaches; in practice, rules need to be elaborated to cover special cases, such as persons absent on extended trips, persons who are too ill or infirm to respond, interviews in areas that are especially difficult to reach, and the like.

The particular respondent rules used have important effects on both the accuracy of the data and the cost of conducting the survey. Usually, although not always, one would expect self respondents to provide more accurate information than proxy respondents. The magnitude of response errors introduced by using proxy respondents will depend on the nature of the survey items and of the target population, and can only be determined by experimentation. Greater use of self-respondents will, of course, add to the data collection cost per household, primarily because more call-backs will be necessary. These effects must be balanced against each other in the context of total survey design.

Data from an experiment using different respondent rules in a pilot test for a demographic survey are shown in Table 5.19 (Tuygan and Cavdar, 1975). Under procedure 1, "interviewers were instructed to make special efforts to obtain responses from all eligible women to questions relating to their own children". Under procedure 2, no such instructions were given and as can be seen in Table 5.19, proxy respondents were used more often. The table shows the effect

of the two different procedures on inquiries about children ever born to ever-married women and proportion of children no longer living at the time of the interview.

Table 5.19

Effects of proxy response on demographic items: pilot test for Turkish Demographic Survey

Item	Urban area		Rural area	
	Procedure 1	Procedure 2	Procedure 1	Procedure 2
Per cent of women responding themselves	88	75	66	59
Average number of children ever born to ever-married women	3.00	2.81	5.36	5.07
Proportion of children dead at time of interview	.207	.178	.266	.254

This example shows that it is not always possible to obtain self-response for all or even nearly all sample persons. In this case the authors state that it is difficult to interview women directly in the "rural traditional population", even when special efforts are made. Nevertheless, if self-response is accepted as being more accurate for these items, these special efforts did improve the accuracy of the survey estimates.

Another example, which is not based on a controlled experiment but which suggests that there may sometimes be substantial differences between self and proxy responses, is shown in Table 5.20 (Shah, 1981).

Table 5.20

Per cent of women aged 10 and over in labour force: Pakistan, 1974-75

Type of area	Source of data: Pakistan Fertility survey, 1975 (self-response)	Source of data: Labour Force survey, 1974-75 (largely proxy response)
Urban	16	4
Rural	18	8

In the Fertility Survey, female interviewers were used and they were able to interview women directly. The Labour Force Survey interviewers were men, and they normally obtained information for all household members from a male member of the household. Although some of the differences could have resulted from differences in concepts and definitions, differences in timing of the two surveys and other factors, the author suggests that the main reason for the lower rates in the Labour Force Survey was that many Pakistani men believe that there is some social stigma attached to women working outside the home, and therefore they are likely to under-report labour force participation of female members of their families.

This example shows that comparison of survey estimates with more or less comparable data from other sources (another survey, a census, or data from administrative records) can often shed some light on the existence and sources of non-sampling error.

The final example in this sub-section is based on reinterviews conducted for a sub-sample of the respondents to the United States Health Interview Survey over a 3-year period centered around 1960 (U.S. National Center for Health Statistics, 1973b). In the initial interviews, the use of proxy respondents was allowed when a member of the household was absent; however, in the reinterviews all adults were

self-respondents. In the reinterviews, the same questions were to be asked without reference to the original responses; however, the reinterviewer carried the original responses and, after the reinterview, compared the two sets of responses. When differences existed, the reinterviewer tried, with the respondent's help, to determine the proper response. This last procedure is called reconciliation.

Thus, the data shown in Table 5.21 are based on the initial interviews, with self or proxy respondents, and the independent reinterviews, always with self-respondents, after reconciliation. The first two columns show the proportion with each characteristic. The third column shows the net shift in the results, and the fourth column shows the proportion shifting in either direction between the original interview and the reconciled reinterview.

Table 5.21

Results of initial interviews and reconcilced reinterviews, for self and proxy respondents to initial interview: United States Health Interview Survey, fiscal years 1959-61

Survey item, subject and respondent on original interview, persons with:	Persons in class		Average Net difference rate	Average gross difference rate
	Original	Reinterview		
One or more chronic conditions:				
Adult, self-respondent.....	59.4	64.8	-5.4	6.1
Adult, proxy respondent.....	47.4	58.5	-11.1	14.9
(self-respondent-proxy respondent)	12.0	6.3	5.7	
One or more hospital episodes in past 12 months:				
Adult, self-respondent.....	13.2	13.5	-0.3	0.6
Adult, proxy respondent.....	9.0	10.0	-1.0	1.3
(self-respondent-proxy respondent)	4.2	3.5	0.7	
One or more restricted activity days in past 2 weeks:				
Adult, self-respondent.....	13.5	14.3	-0.8	3.5
Adult- proxy respondent.....	8.2	11.0	-2.8	5.0
(self-respondent-proxy respondent)	5.3	3.3	2.0	
One or more bad days in past 2 weeks:				
Adult, self-respondent.....	6.7	7.1	-0.4	1.7
Adult, proxy respondent.....	4.7	5.5	-0.8	2.0
(self-respondent-proxy respondent)	2.0	1.6	0.4	
One or more time-lost days in past 2 weeks:				
Adult, self-respondent.....	3.1	3.5	0.4	0.8
Adult, proxy respondent.....	4.3	4.7	-0.4	2.4
(self-respondent-proxy respondent)	-1.2	-1.2	-	
One or more hospital days in past 2 weeks:				
Adult, self-respondent.....	0.7	0.7	-	0.3
Adult, proxy respondent.....	0.4	0.4	-	0.2
(self-respondent-proxy respondent)	0.3	0.3	-	

The net difference rates for proxy respondents were generally larger; however, some part of these differences might have been due to inherent differences between persons who were available to report for themselves in the initial interview and those who were not. The net differences rates were larger for proxy respondents for 4 of the 6 items, and the gross difference rates were larger for 5 out of 6, which suggests that larger response biases are associated with proxy respondents. The largest net and gross difference rates, for both self and proxy respondents, are for the items which have the greatest element of subjectivity and are the hardest for other family members to observe: chronic conditions and days of restricted activity.

Additional findings on differences in completeness of reporting for proxy and self-respondents in health surveys are given by Kovar and Wilson (1976), Kovar and Wright (1973), and U.S. National Center for Health Statistics (1977, pp. 15-16).

6. Other illustrations of response bias

Often a carefully designed set of questions will produce more accurate information on a particular variable than will a single question. For example, if an accurate estimate of total household income is needed, a set of questions covering different types of income and/or different members of the household is likely to give a more accurate result than a single question asking for total income of the household.

Marckwardt (1973) has reported on an experimental comparison of fertility data collected from an in-depth interview, using a detailed pregnancy history approach, with data from a "short form", similar to one that might be used in a census. The comparison was made in a national sample survey in Peru (excluding the Lima-Callao metropolitan area) in 1969.

Table 5.22 shows data on mean number of reported live births, by age of woman and degree of urbanization, for the two procedures. If the data from the depth interviews are taken as the standard,

then there appears to be a response bias of approximately 10 per cent, in the direction of under-reporting, in the short interviews.

Table 5.22

Cumulative Fertility, All Peruvian[a] Women 15-49 Years Old, by Degree of Urbanization, Age and Type of Interview

Degree of Urbanization and Age	Mean number of reported live births[b]		
	Short Interview (S)	Depth Interview (D)	(S) + (D)
All places	3.1	3.4	91%
15-24	0.7	0.8	85
25-34	3.3	3.8	89
35-49	5.8	6.4	92
Rural, under 2500	3.5	3.8	91
15-24	0.8	1.0	85
25-34	3.6	4.0	90
35-49	6.0	6.6	92
Towns 2500-19,999	2.9	3.1	93
15-24	0.6	0.6	95
25-34	3.4	3.7	92
35-49	5.6	6.1	91
Cities of 20,000 or more	2.2	2.5	90
15-24	0.4	0.5	82
25-34	2.6	3.0	85
35-49	5.4	5.8	93

a - The Lima-Callao metropolitan areas was excluded.

b - The means corresponding to all ages for each size-of-place category have been standardized on a common age distribution of short and depth interviews, and those corresponding to "all places" have been standardized on a common age-place distribution.

Source: 1969 National Fertility Survey, Sample Survey Center of the Ministry of Labor of Peru.

The differences observed between the two procedures cannot be attributed only to differences in the type and number of questions asked. The use of proxy respondents was allowed for the short interviews, but not for the depth interviews, and this in turn led

to differences in non-response rates for the two procedures.

Jabine and Rothwell (1970) reported on findings from 15 "split-panel" tests of census and survey questionnaires conducted by the United States Census Bureau. Their conclusions about probing questions were as follows:

> "Probing questions can improve accuracy of response. A version of the Current Population Survey questionnaire which included probes about hours worked and about self-employment provided improved statistics for these subjects. Marginal improvement in reports about size of farm were also achieved through probing questions. There are, of course, limits to effectiveness of probing. Additional probing about labour force status in test Number 7 produced no significant changes."

As mentioned earlier in this chapter, responses to questions on sensitive topics such as use of alcohol and drugs or participation in illegal activities are often biased downward. Several methods have been proposed to eliminate or reduce under-reporting on sensitive topics. Perhaps the most ingenious is randomized response, a procedure which permits estimates of proportions at the aggregate level without requiring respondents to disclose to interviewers their responses to specific sensitive items (Warner, 1965).

Tezcan (1977) reports on the use of the randomized response procedure to obtain data on the prevalence of induced abortion in a national sample of Turkish women under age 45. Comparable data were obtained from direct interviews for a parallel sample of women. It was found in this study that "the proportion of women studied with one or more induced abortions in their reproductive lives was found to be 13.9 in the direct questioning sample and 33.1 per cent in the randomized response sample". Even more striking results were obtained in an application of the randomized response technique for the same purpose in Ethiopia. As reported by Chow, Gruhn and Chang (1979), the estimated rate of induced abortion (per cent of women with one or more) based on randomized response was 34.5 per cent, whereas fewer than 0.1 of the women were willing to admit, on direct

questioning, to having had an induced abortion.

It may be of interest to complete this section with an example of response bias which could have occurred even though every respondent answered the question correctly. Tabulations from the October 1980 education supplement to the United States Current Population Survey showed unexpectedly large increases in the proportion of students enrolled in private, as opposed to public schools. Investigation showed that the layout of the response categories to the question "Is it a public or private school?" had been changed from what it had been in previous years. The order of the categories in 1980 was "Private" followed by "Public", the reverse of what it had been. It was hypothesized that some of the interviewers failed to notice or adapt to this change and, therefore, recorded all or some of the responses to this item incorrectly (Siegel, 1981).

D. Response variance

1. Components of response variance

As pointed out in Section B of this chapter, the response variance is the variable component of response error, resulting from transient factors which affect individual responses in a more or less random fashion. If it were possible to repeat, or replicate, a survey using a fixed sample and under the same procedures and other essential survey conditions, the differences between estimates from each replication and their average or expected value over all replications would be entirely due to response variance.

Variable response errors may be either uncorrelated or correlated. In this context, uncorrelated response errors are those that are not affected by the particular interviewers, supervisors, coders or other survey personnel who happen to be associated with a particular element of the sample. The variable part of these uncorrelated response errors is measured by the uncorrelated or simple response variance.

Insofar as individual interviewers, or supervisors, or coders have different average effects on their workloads, they introduce response errors which are "correlated" for all elements of the sample included in that workload. These correlated response errors give rise to correlated response variance, or depending on the category of survey personnel with which it is associated, interviewer variance, supervisor variance, or coder variance.

Another way of looking at correlated response variance, say interviewer variance, is that it results from biasing effects that differ from interviewer to interviewer. In collecting data on household income, for example, one interviewer might tend to forget to ask about income of family members other than the head; a second might tend to record gross rather than net income for self-employed persons; and a third might tend to neglect to ask about income from dividends and interest. Some of these errors overstate income; others understate it. Each interviewer's result has a different expected value under the given essential survey conditions. The correlated response variance measures the variability of these interviewer expected values about their overall mean. The difference between their overall mean and the "true value" is, of course, response bias.

2. Simple response variance

(a) Basic concepts

There is no way, in a single sample survey, to distinguish true variation between individuals (which gives rise to sampling error) from additional variation arising from random factors which affect individual responses (which gives rise to simple response variance). In fact, the usual procedures for estimating sampling error automatically include the effects of simple response variance as well. In order to estimate simple response variance separately, it is necessary to repeat the survey under essentially the same conditions and with the same sample,

or at least a subsample of the units included in the initial survey.

As is true for sampling error, the sample size is the major determinant of the contribution of simple response variance to the mean square error of survey estimates. As the sample size is increased, the effect of the simple response variance will be reduced in about the same proportion.

There are various measures of simple response variance that can be used. Measures used in connexion with attributes may be illustrated by a simple 2 x 2 table, similar to the one shown in Section B of this chapter. As before, the labour force status in the initial survey will be shown in the stub of the table. The heading, however, will no longer show the correct status. Instead, it will show the status as determined by a repetition of the initial survey.

Labour force status in initial survey	Status in re-interview	
	In labour force	Not in labour force
In labour force	a	b
Not in labour force	c	d

One measure of simple response variance is the gross difference rate, that is the proportion of individuals assigned to different categories in the two surveys. The gross difference rate, expressed as a proportion, is

$$g = \frac{b + c}{n} \qquad \text{where} \quad n = a + b + c + d$$

If all individuals were classified in the same way in both surveys, the gross difference rate would be zero.

If the two surveys were, in fact, conducted independently under the same general conditions, g/2 would be an approximate estimate of the simple response variance. In practice, it is not possible to meet this requirement. First, there may be real changes in the situation: certain facts may have changed during the interval between the repeat interviews, or respondents' may change their opinions. Such changes, if not eliminated, will overestimate the extent of response instability. To reduce these effects, the time interval between the two interviews should be as short as possible. On the other hand, the two interviews may not be independent; some respondents may remember the answers they gave during the first interview and try to be consistent. Such effects would under-estimate the actual variability of individual responses. To control these effects, the time interval between the two interviews should be longer, which conflicts with the previous requirement. For factual items which are not likely to change between closely-spaced interviews, such as date of birth, sex, educational attainment of adults, or last year's income, estimates of the simple response variance based on repeated surveys (re-interviews) are likely to be underestimates.

A measure which is useful for comparing response variability for different items is the index of inconsistency. Using the above notation, it is expressed as:

$$I = \frac{g}{2p(1-p)}$$

where p is the average proportion in the two surveys having the specified characteristic. It can be shown that the index

of inconsistency, which can take values between 0 and 1, is a measure of the proportion of total variance for an item that is accounted for by response variance.

A high value of the index of inconsistency may indicate that the item in question is inherently unstable, or it may indicate that the questions, instructions and procedures being used to collect the information need to be improved in order to reduce the level of instability of responses.

Analogous measures of simple response variance for aggregates, such as expenditures and income, can also be developed. However, it is usually easier and more convenient for the analysis of response variance to display such data by class intervals, such as a distribution of households by income class, so that measures for attributes can be used. In such a case, a 2 x 2 table for the analysis would be as shown below.

	Status in replication	
Status in initial survey	In income class $2,000 to $2,999	Not in that income class
In income class $2,000 to $2,999	a	b
Not in that income class	c	d

Repeating the table for each income class, the desired measures of simple response variance can be estimated.

(b) Examples

The first illustration, taken from Seltzer (1973), shows the proportion of persons, by 5-year age group in

reinterviews, who had been classified in different age groups in the initial interview. (The values shown in Table 5.23 are not gross difference rates; they are equivalent to values of $\frac{c}{a+c} \times 100$ in the 2 x 2 tables presented earlier in this section.)

Table 5.23

Reinterview data from two surveys: percentage of persons classified in different age group in initial interview

Age at re-interview	Ghana sample, 1960		Rawalpindi area, 1971*
	Males	Females	
Under 5 years	13	15	20
5-9	27	27	27
10-14	39	31	31
15-19	35	38	37
20-24	48	44	47
25-29	52	58	48
30-34	58	63	59
35-39	66	66	66
40-44	67	60	70
45-49	64	77	64
50-54	66	70	71
55-59	66	85	90
60-64	69	72	68
65-69	66	79	93
70-74	69	71	61
75 and over	41	32	53

*Provisional and unofficial.

As might be expected, the levels of response variability increased with age in both surveys. The lower percentages for the age group 75 and over reflect the fact that this is an open-end interval covering more than 5-years. Errors could occur only at the lower limit of the interval, and a small proportion of the people in the age group (in the reinterview) were close to that limit.

Misclassification in this sense is, of course, a function of the size of the interval. If 10-year age groups had been used, the percentages would have been substantially smaller.

Table 5.24 shows values of the index of inconsistency for selected variables from a fertility survey in Peru (O'Muircheartaigh and Marckwardt, 1980), classified by age, place of residence and level of education of the respondents. For variables such as birth intervals, which depend upon detailed information on birth dates of children, and attitudinal variables such as the desired number of children, the index is high; in one case 68% of the variance is due to response variability. Not surprisingly, the index values are generally higher for the older, rural and less educated respondents.

Table 5.24

Indexes of inconsistency for selected variables by age, residence and education of respondents: Peruvian fertility survey

Variable	Age		Residence		Education	
	25	45	Lima	Rural	None	7 years
First birth interval	56	68	40	63	61	58
Last closed interval	30	33	11	20	17	11
Year of first birth	5	15	1	3	3	4
Children ever born	2	3	1	3	3	1
Month of last birth	4	3	1	4	4	2
Year of last birth	6	3	1	4	4	1
Desired no. of children	45	76	48	68	61	36
Current age	4	27	1	3	5	1
Age at marriage	16	29	11	36	40	7
Year of marriage	10	25	2	6	11	1
Age group	14	68	1	4	7	1
Whether worked	31	41	21	43	53	21
Ever-use of contraception	30	36	34	62	45	37
Births in past 5 years	5	28	5	13	11	6
Sample size	201	158	379	431	309	278

Table 5.25 illustrates the use of the index of inconsistency to compare response variability for different classes of items. The data are based on reinterviews of a sample of persons from the United States 1960 Census of Population. For each topic, the table shows the distribution, by size class, of the indexes of inconsistency estimated for each of several categories under that topic. The column labelled "Total income", for example, shows the distribution of indexes of inconsistency estimated for each of 16 income size classes, separately for males and females, and for both sexes combined (U.S. Bureau of the Census, 1964).

Table 5.25

Distribution of estimated indexes of inconsistency for selected variables, by size and topic: U. S. Census of Population, 1960

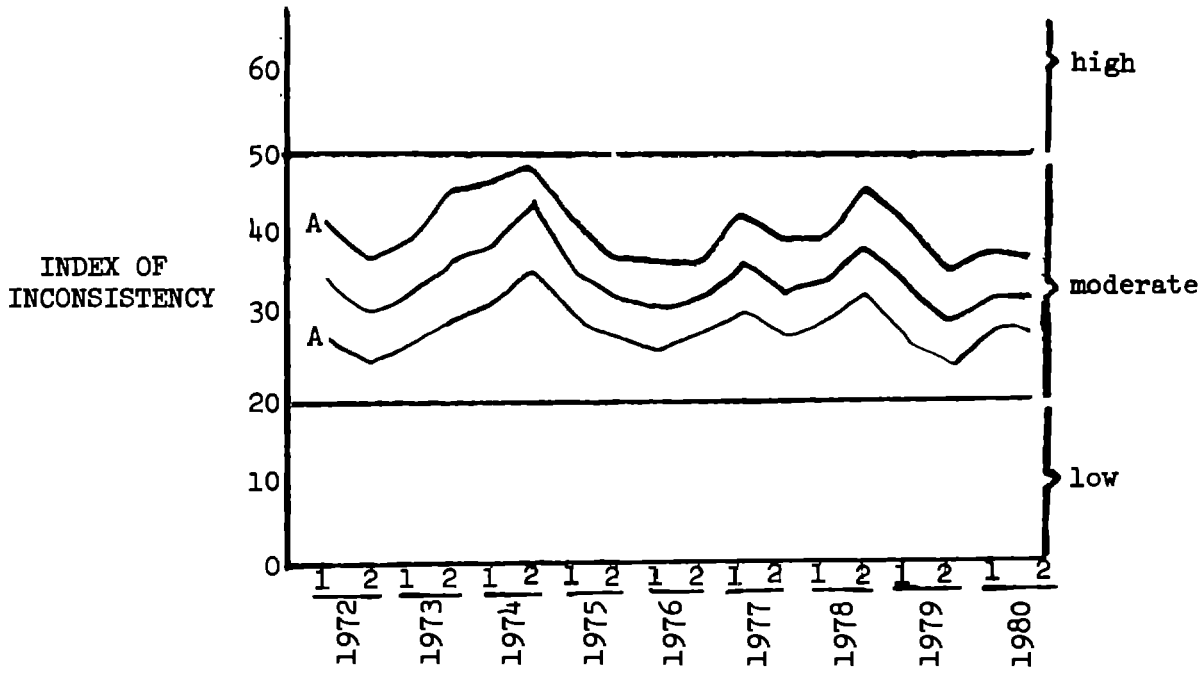
Index of inconsistency (percent)	Total classes	Population characteristic											
		Sex	Color	Age	Mobility	School enrollment	School attainment	Type of school	Children ever born	Employment in 1959	Total income	Self-employment income	Other income
Total classes	322	2	2	144	5	17	15	6	13	6	48	32	32
0 - 09	129	2	2	112	3	1	-	3	5	-	-	-	1
10 - 19	50	-	-	23	2	7	2	3	7	4	2	-	-
20 - 29	29	-	-	9	-	7	1	-	1	2	7	2	-
30 - 39	29	-	-	-	-	2	9	-	-	-	9	2	7
40 - 49	19	-	-	-	-	-	2	-	-	-	14	1	2
50 - 59	34	-	-	-	-	-	1	-	-	-	15	6	12
60 - 69	10	-	-	-	-	-	-	-	-	-	1	8	1
70 or over	22	-	-	-	-	-	-	-	-	-	-	13	9

This table makes it clear that some items are considerably more affected by response variance than are others. Low indexes are observed for sex, colour, age, mobility (residence 5 years ealier), type of school in which enrolled (public or private) and children ever born. Somewhat higher values are observed for other education items and for employment in the previous year. The highest indexes are clearly those for income items.

Figure 5.3 shows recent trends in the index of inconsistency for the number of unemployed persons, as estimated from the United States Current Population Survey, based on semi-annual averages from sample reinterviews conducted since 1972. The band A-A around the trend line represents a 95-per cent confidence band for the point estimates. Note that values of the index up to 20 per cent are called "low", those from 20 to 50 moderate, and those above 50 high. (U.S. Bureau of the Census, 1981).

Figure 5.3

Index of inconsistency for unemployment, unreconciled reinterview sample: U. S. Current Population Survey, 1972-80



Similar charts for other employment items showed the following trends during the same period.

Item	Level of index
Employed in agriculture	low to moderate
With a job, but not at work	moderate
Non-agriculture employment: full-time	low
Non-agricultural employment: part-time	moderate
Not in labour force	low

3. Correlated response variance

(a) Basic concepts

Although the concepts underlying correlated response variance are straightforward, its mathematical treatment is relatively complex. The objective of this exposition is to give a general idea of what it is, how it can be measured, and its implications for the design of household surveys. Readers seeking a more rigorous mathematical treatment may consult Fellegi (1964, 1974); Hansen, Hurwitz and Bershad (1961); Kish (1962); and U.S. Bureau of the Census (1968a).

The nature of correlated response variance associated with interviewers was described by Kish (1962) as follows:

> "Each interviewer has an individual average 'interviewer bias' on the responses in his workload, and we consider the effect of a random sample of these biases on the variance of sample means. The effect is expressed as an interviewer variance which decreases in proportion to the numbers of interviewers."

A similar statement could be made for components of correlated response variance associated with supervisors, editors, coders, key-punchers etc.

An important property of the interviewer variance is that its contribution to the total mean square error is inversely proportional to the number of interviewers employed in the survey. For a fixed total sample workload, the larger the number of interviewers, the smaller will be the effect of interviewer variance on the results. Likewise, the effect of correlated errors introduced by coders would depend on the number of coders employed.

Other things being equal, this suggests that the ideal survey design would be to have a different interviewer for each respondent! In a sense, this is what happens when self-enumeration is used, as has been the case in recent years for censuses in several developed countries. In fact, experiments which demonstrated the effect of interviewer variance on census results had much to do with the trend toward greater use of self-enumeration (see, for example, Fellegi and Sunter, 1974).

However, self-enumeration is not a feasible alternative for most household surveys, and cost factors and other considerations place upper limits on the number of interviewers that can be used in a particular survey. Nevertheless, it is important to be aware that just as sampling errors may be quite large for estimates based on a small number of primary sampling units, the effects of interviewer variance can be substantial for estimates based on the work of a few interviewers. This is a major consideration in deciding whether a survey is adequately designed to produce subnational estimates, say for regions

or provinces. If interviewers are not uniformly well-trained and supervised, interviewer variance can overwhelm other sources of error in subnational estimates, even though its effects may be small at the national level.

The contributions to total error of correlated response variance may or may not be included in the usual estimates of sampling variance. Household surveys generally use multi-stage or clustered samples, hence, the sampling errors are also "correlated" in the sense that individuals coming from the same area or cluster tend to be more alike than do individuals in the population at large. Taking several individuals (or households) from the same cluster reduces the independent information contained in the sample; in other words, it increases the sampling variance. Practical methods of computing sampling variances are generally based on comparing the values obtained from different clusters in the sample: the larger the differences between the clusters, the higher the sampling error will be. (In addition, of course, the size of the sampling error depends inversely on the number of clusters in the sample.) In an analogous way, survey workers, say interviewers, impose their own "clustering" effect on the survey results insofar as errors in responses of individuals interviewed by the same interviewer tend to become somewhat similar due to the influence of that particular interviewer. If the survey arrangement is such that each interviewer is assigned to work in only one sample cluster, then the effect of interviewers will be completely confounded with the effect of clustering on the sampling variance, and usual methods of estimating sampling variance will automatically include the correlated interviewer variance. By contrast, if all interviewers

worked as a team in each cluster, or if the interviewers' workloads were assigned at random, the usual estimation of sampling variance would not include the effect of additional variability due to the interviewers. The effect may be partly included with other intermediate arrangements for assigning workloads to enumerators (see Verma, 1981)

To make separate estimates of interviewer variance or other types of correlated response variance, it is necessary to introduce some degree of randomization or interpenetration of workloads assigned to the particular category of survey personnel involved. For example, in a recent study of interviewer variance in a crime victimization study conducted in 8 cities in the United States, the design was as follows (Bailey, Moore and Bailar, 1978): i) the sample workload in each city was divided into 18 crew-leader assignments, each containing 8 interviewer assignments from a total of 144. Each interviewer assignment consisted of about 80 households; ii) within each crew-leader area, 4 pairs of interviewer assignments were formed; iii) interviewers were assigned to interviewer assignment areas close to their homes; iv) 36 of the 72 assignment pairs were selected at random. Within each of these 36 pairs, housing units in each assignment were assigned in an alternating systematic pattern to the two interviewers. In other words, the workloads of the two interviewers assigned to a particular pair of assignments were interpenetrated.

This design made possible the estimation of the interviewer variance, following a method proposed by Fellegi (1974), using data from both sets of paired interviewer assignments -- those with interpenetration and those without it.

Randomization of assignments generally increases

the cost and complexity of survey operations. In practice there are many constraints which limit the degree to which complete randomization can be effected, though in multi-stage surveys it is frequently possible to select two "interpenetrating samples" within each area and assign them at random to two different interviewers who have been assigned to work in that area. Fortunately, useful results can be obtained even when complete randomization is not possible (Kish, 1962).

So far, there have been only a few studies to obtain separate estimates of interviewer variance and other components of correlated response variance in household surveys. More studies are needed; without adequate information about this important component of error effective application of the principles of total survey design is not possible.

Specific measures of correlated response variance are presented in connexion with the examples which follow.

(b) Examples

Most of the available examples of correlated response errors in surveys have to do with interviewer effects. An early example in anecdotal form was given by Rice (1929). The municipal home for poor people in New York received many applications asking for shelter and food. Two interviewers were sent to inquire about the applicants' situation and ascertain the causes of their state of misery. One of the interviewers reported that alcohol was the cause in the large majority of the cases, while the second interviewer reported that the capitalist society was the main cause. Subsequently, it was found that the first interviewer was a prohibitionist, while the second was a socialist. The findings were thus

in tune with their respective ideologies, and each saw in the given situation what he expected to see.

Of course that was an extreme example. Many studies have shown that responses without larger interviewer effects -- even on attitudinal items, some highly ambiguous and emotionally loaded -- are obtainable in carefully prepared and executed surveys (see, for example, Kish, 1962). Nevertheless, the interviewer effect as a potential source of response error frequently goes unrecognized. A reinterview study following the U.S. Current Population Survey indicated that nearly a third of the individual discrepancies between the original interview and the reinterview arose from interviewers' errors, and a majority of these were due to misrecording of answers (Hansen and Steinberg, 1956). In a comparison of respondents' ages recorded on the questionnaire during the interview with the same information transcribed subsequently from tape recording, it was found that in only 72% of the cases were the results identical (Seltzer, 1973).

Interviewer effects on the variability of survey results were first studied at the Indian Statistical Institute (Mahalanobis, 1946; Sukhatme and Seth, 1952). Soon afterwards, similar studies were made in Great Britain. Kemsley (1965) reported finding significant interviewer effects on response rates in four of six surveys that he examined. Kemsley also studied how these interviewer effects were associated with the kinds of data being collected and the methods of data collection. The results of this analysis are shown in Table 5.26. Kemsley used an analysis of variance approach; the F-ratios in the table were calculated to test, for each item and collection method, whether the observed interviewer effects were significant.

Table 5.26

Distribution of F-ratios by type of data and collection method in expenditure surveys: Great Britain

Type of data and method of collection	No. of items Examined	Number of F-ratios		
		≤ 1.00	≥ 1.00	Significant at .05 level
Expenditure items				
Data taken from respondent records	26	10	16	0
Respondent recall	19	2	17	9
Household characteristics	44	18	26	9

For the first set of items, data on expenditures were obtained from records kept by the respondents; the interviewers did not have much influence on results, and none of the F-ratios was found to be significant at the .05 level. In the second set, the expenditures were reported on the basis of recall, and the interviewer effects were found to be significant for nearly one-half of the variables. The third set consisted of items not requiring a substantial recall effort, but even for these the interviewer effects were found to be significant for several of the variables. A number of other studies on interviewer effects were made in Great Britain; see, for example, Gales and Kendall (1957); Gray (1959); Kemsley and Nicholson (1960); Moser (1951).

To use information on interviewer variances in designing a household survey, it is not enough just to know that significant effects exist for certain items, some estimates of the size of the interviewer variance for different variables is needed. Table 5.27 shows such estimates, taken from the study by Bailey, Moore and Bailar (1978), whose design was described earlier.

Table 5.27

Estimated relvariances for major personal victimizations: Baltimore, 1975

Kind of victimization	Estimated victimization rate	Relvariance of estimate			Ratio: correlated response to sampling
		Total	Sampling	Correlated response	
Total	.1087	.1589	.0988	.0601	.61
Assaultive violence, total	.0604	.2624	.1609	.1015	.63
with theft	.0115	.7531	.6454	.1077	.17
without theft	.0489	.2940	.1912	.1028	.54
Personal theft without assault	.0484	.2216	.1711	.0505	.30

The relvariances shown in Table 5.27 are for a design in which the average interviewer assignment was 80 households. If the same sample design had been used with fewer interviewers and, therefore, larger assignments, the contributions of correlated response variance to total variance would have been larger, and, of course, with more interviewers the reverse would have occurred.

To compare interviewer variance effects for different items and from different surveys (and censuses), it is necessary to have a measure that is independent of the size of interviewer assignment. For this purpose, Kish (1962) uses "roh", which is the ratio of the interviewer variance per element to the total variance per element. Table 5.28 shows ranges of values of roh compiled by Kish from several different sources. The values of "a" in the table are the numbers of interviewers in each study; the larger the number of interviewers, the more reliable the estimates of roh.

Table 5.28

Values of roh for different classes of items from several investigations of interviewer variance

Source of estimates, class of item and number of interviewers	Range of roh
Kish (1962)	
46 variables* in first study (a = 20)	0 to 0.07
48 variables* in second study (a = 9)	0 to 0.05
Percy G. Gray** (1956) (a = 20)	
8 "factual" items	0 to 0.02
Perception of and attitudes about neighbours' noises	0 to 0.08
8 items about illnesses	0 to 0.11
Gales and Kendall** (1957) (a = 48)	
Semi-factual and attitudinal items about TV habits	0 to 0.05
Hanson and Marks (1958) (a = 705): data from U.S. 1950 census	
31 "Age and sex" items	0 to 0.005
19 simple items	0 to 0.02
35 difficult items	0.005 to 0.05
11 "not answered" items	0.01 to 0.07

* Kish states that these variables were "mostly attitudinal".

** Kish states "for these investigations we computed the above values from the published test statistics and this translation involves some uncertainties.

From this table it can be seen that interviewer variances tend to be larger for "more difficult" and for attitudinal items than they are for simple "factual" items. The last item in Table 5.28 indicates that the level of item non-response for some items varies substantially from interviewer to interviewer. Hanson and Marks (1958) noted that the larger interviewer effects observed in the 1950 U.S. Census of Population were for:

> "census questions containing one or more of the following factors -- (a) interviewer 'resistance' to the question, i.e., a tendency on the part of the interviewer to be hesitant about making the inquiry and possibly a tendency to omit or alter the question or assume the answer, or (b) a relatively high degree of ambiguity, subjectivity or complexity in the concept or wording of the inquiry, or (c) the degree to which additional questioning tends to alter respondent replies".

As stated earlier, the impact of interviewer variance depends on the size of interviewer workload. More specifically, the ratio of the total response variance to the simple (uncorrelated) response variance is given by:

$$1 + \text{roh}\ (\bar{n} - 1)$$

where $\bar{n}$ is the average number of units per interviewer assignment. Thus, even for a relatively small value of roh, say roh = 0.01, the interviewer effect doubles the response variance for average workloads of 100. For a higher value of roh, say roh = 0.05, the response variance will be doubled for workloads of size 20. Clearly, the correlated response variance associated with interviewers cannot safely be ignored in the design and evaluation of household surveys.

There have been few, if any studies of correlated response variances associated with personnel other than interviewers -- supervisors, coders, etc. -- in household surveys. Studies in connexion with the 1960 United States Census of

Population (U.S. Census Bureau, 1968a, 1972) indicated that there were significant crew-leader and coder variances for some items, although most of these were small in comparison to the correlated response variances associated with interviewers.

CHAPTER VI

CONTROL OF NON-SAMPLING ERRORS IN HOUSEHOLD SURVEYS

A. Introduction

The purpose of this chapter is to provide an overall view of steps that should be taken (or at least considered) at each stage of a survey in order to control the various types of non-sampling errors that have been discussed and illustrated in previous chapters. Complete elimination of non-sampling errors is, of course, an unattainable goal. "Control" is actually a shorthand expression for identifying various sources of non-sampling error and taking appropriate steps to minimize their effects within the context of total survey design. In general, the steps that are appropriate are those that are feasible and lead to the greatest reduction of error -- sampling or non-sampling -- per unit of cost.

Costs of alternative procedures for controlling non-sampling errors cannot be ignored. In a health survey, the most accurate method of collecting information on current health status might be to hire physicians to conduct a physical examination of each person in the sample. However, even if this were feasible, the cost per interview might be so high that only a very small sample could be used, leading to unacceptably large sampling errors. Because costs have to be considered, the emphasis in this chapter will be on design features and quality control procedures whose costs are relatively low. Nevertheless, higher-cost procedures, such as direct observation, physical measurements, or record checks should not be ruled out of consideration. For some kinds of data, the biases associated with less costly methods may be so large as to make the results of little or no value.

Measures to control the quality of survey results may be applied at any of three main stages of a survey:

(i) The design and preparatory stage:

(ii) The execution stage, i.e., the data collection and processing:

(iii) After completion of the main survey operations.

The discussion in this chapter will cover the design and preparatory and data collection stages, following the sequence of survey activities presented in Chapters III through VII, Part One, of the revised Handbook of Household Surveys (United Nations, 1981). It is not the purpose of this chapter to describe each of these activities in detail; what is attempted here is to identify the kinds of non-sampling errors that may be associated with each survey activity and to suggest appropriate methods of controlling them.

One general question that needs to be considered before proceeding is: who should be responsible for controlling non-sampling errors in a survey? The general answer is simple: the preferred way of conducting a survey is to have one person take full responsibility for all aspects of the survey. This person -- the survey manager -- should be responsible for the quality of the results.

What can the survey manager do to ensure that proper attention is given, at all stages of the survey, to the control of non-sampling errors? One answer is to use this document as a guide to the many kinds of non-sampling errors that can arise at each stage and to use this chapter as a guide to procedures that can be used to control them. To facilitate this, a checklist is provided at the end of this chapter.

Quality control of survey operations is not likely to be a complete success unless the procedures and the need for them are carefully explained to survey workers -- interviewers, coders, data entry clerks and others -- who will be affected by them. Statistical organisations in different countries have different styles of management, ranging from a completely top-down approach to one of wide participation in management decisions at all levels. Whatever approach is used, the manager must make it clear to the staff that both quality and production matter. Preferably, survey workers should be encouraged to submit their own ideas on how to improve the quality of work done. Those who do the "hands on" work in the collection and processing of survey data may often have insights not available to the supervisory and professional staff.

Quality control usually involves "feedback" to interviewers and others about errors they have made. This should be done in a constructive way. It should be pointed out that some errors in large-scale survey operations are inevitable, and that the purpose of feedback is to eliminate systematic errors due to misunderstanding of concepts and procedures. Some of these errors may not be the "fault" of those who make them; they may have resulted from a failure in designing the questionnaire or from inadequate training of field and office staff.

Quality control systems are also likely to be more effective if positive incentives, in the form of extra remuneration or other recognition, are established for those who do good work in each survey operation.

This chapter discusses in detail various measures and procedures for the control of non-sampling errors during specific stages of preparation and execution of a survey, from general planning and selection of topics to be included, to survey design, sampling, questionnaire development and testing, training, fieldwork and data collection, analysis and reporting. While these activities are necessary for any survey of high quality, their potential usefulness, and their feasibility within constraints of available resources, are greatly enhanced when surveys are planned and executed within the framework of an integrated programme. Furthermore, within that framework, there are many things that can be done during and after a particular survey, which will provide valuable information and experience for use in designing and implementing subsequent surveys. These considerations relating to long-range strategies for improving data quality are addressed in Chapter VII.

In addition to this chapter and the chapters cited from the Handbook of Household Surveys, a useful reference for survey managers and others interested in practical techniques for controlling non-sampling errors in surveys is Manual Series No. 2 of the Laboratories for Population Statistics, "Operational Control of Sample Surveys", by Simmons (1972). A simple set of guidelines given in that Manual (page 9) is well worth citing here:

"For any phase of the operation:

(a) Decide what is desired to be accomplished;
(b) Write instructions for doing it;
(c) Do it;
(d) Devise and execute a checking procedure to ensure that is being done correctly and is completed in accordance with the instructions."

The order of the last two steps may be questioned: the checking procedure should normally be devised before performance of the operation and its execution will often overlap the performance of the survey operation that is being checked.

Another useful suggestion by Simmons (1972, p. 14) was that "If circumstances permit, the ideal course is to obtain an independent review of the total plan by one or more competent statisticians who are knowledgeable of the survey objectives, but who were not the responsible original designers." In fact, such reviews could address the entire survey plan prior to its execution, or could address specific aspects, such as the sample design, questionnaire, or data processing plan. For each review, persons with expertise in the appropriate aspects of surveys should be sought.

Before proceeding to the next section, the reader may wish to look at the checklist appearing at the end of this chapter. The checklist is an attempt to list, in a convenient form, the more important questions that need to be considered by survey directors and staff who have a commitment to getting the best quality results with the resources available to them.

The questions in the checklist are grouped by "survey activity", roughly following the organization of survey activities used in the revised Handbook of Household Surveys (United Nations, 1981). For most of the questions, some major implications and possible answers appear in the "Remarks" column of the checklist.

The remaining sections of this chapter discuss the questions appearing in the checklist and possible answers to them in more detail. They are organized in the same way as the checklist, so that Section B covers "General planning", the first activity appearing on the checklist, Section C covers "Determination of topics and items to be included", and so forth.

Two qualifications should be kept in mind in using the checklist. First, it is not intended to cover all questions that need to be considered in planning and executing a household survey, only those that are believed to be particularly relevant to the control of non-

sampling errors. For a more general and comprehensive discussion of each survey activity, the reader should consult the Handbook. Second, not all household surveys follow exactly the same pattern with respect to the inclusion and sequencing of the survey activities shown in this presentation. Therefore, the reader will need to exercise some judgment in deciding which of the survey activities and questions discussed in this chapter are relevant to the particular survey he or she is working on.

B. General planning

The amount of development work needed in preparing for a survey will depend very much on whether the survey will cover topics or target populations not previously investigated by the survey organization. Much of the work needed to develop reliable survey instruments and procedures for new topics can be done at a cost which is fairly low relative to the costs of actual survey operations· however, it may take some time to go through successive stages of design, testing and re-design.

At the very early stages of planning, it is important to decide which of the topics to be covered will require special attention during the preparatory phase of the survey. and to identify the particular problems associated with each of these topics. For some topics, such as expenditures, food consumption and labor force participation, a key question may be the choice of reference period and whether or not to use a panel approach with bounded recall. For some topics, such as contraceptive use or financial assets, sensitivity of the items may be a key issue. For other topics, especially subjective ones, lack of well-defined survey objectives and concepts may be the first problem requiring attention. For certain topics, such as health conditions or area and amounts of crops harvested, many respondents may be unable to give accurate information and some form of direct measurement or observation may be needed. Some topics may incorporate more than one of these issues. Whatever the major questions are, a plan should be laid out for seeking satisfactory answers to them

through pilot studies, pre-testing and other appropriate means.

Among the subjects for which household sample surveys are potentially suited, there are probably very few that have not already been included in one or more surveys. For many topics, standard concepts and definitions have been established by international and regional bodies. Many serious errors can be avoided by reviewing the approaches that other statistical organizations, especially those in similar countries, have used. Obviously, concepts and techniques used elsewhere should not be adopted without a careful evaluation of their suitability. Such an evaluation has two parts: a review of the concepts for compatibility with user requirements, and one or more field tests to see whether the concepts and techniques are appropriate for the survey target population.

C. Selection of topics and items to be included in the survey

The selection of subject matter for a survey is covered in detail in Chapter III.C, Part One of the Handbook. This section discusses those aspects of selecting topics and items that have special relevance to the control of non-sampling errors.

First, the survey planners must ask for each topic and key item: is it feasible, given our resources, to collect data of sufficiently high quality on this topic or item? For some items the answer to the question may be "yes" based on previous experience. Of course, it is not enough simply to note that data on the same topic have been collected previously for the same or similar populations; the results must be evaluated to see if they were acceptable from the point of view of relevance, timeliness and accuracy.

For some topics or items, the answer may clearly be "no". The reason for the answer may be that it is well-known that most respondents are either not able or not willing to give the kind of information wanted. Thus, it would be clear that few respondents in a health survey could describe their illnesses or conditions with sufficient accuracy to permit these to be classified in detail according to the International Classification of Diseases. The collection of data on

financial and other assets (wealth) is known from experience to be exceptionally difficult, at least in more economically developed populations, partly because respondents have difficulty in providing complete and accurate information, and partly because many of them are reluctant to give such information to survey interviewers.

There are other topics, for example, household expenditures or food consumption, for which data have been collected successfully in household surveys, but which are known to require a fairly high level of sophistication and experience from the design, data collection and processing staffs of the survey organization. The survey planners should consider whether their organization is currently qualified to handle such complex topics, or whether it might be better to start with something simpler to acquire the necessary experience.

Sometimes the answer to the question of feasibility will be "maybe" or "don't know". The proposed topic may be a new one or at least new to the target population for the survey. Such topics require special attention in the preparatory stage of the survey, and time may be needed for several stages of testing of survey instruments and collection procedures.

A second important question for the survey planners is whether the set of topics proposed for the survey are compatible and, taken as a whole, can be treated effectively in a single survey. Subsidiary questions are:

(i) how long will it take, on the average, to complete a household interview covering all of these topics, and what impact may the length of interview have on non-response and on the quality of response?

(ii) can the existing interviewing staff (if there is one) do an acceptable job, given the number and complexity of the topics to be covered? How much training will be needed?

(iii) is the data processing staff, including especially those who will be responsible for systems design and programming, sufficiently skilled to process the survey results in a timely manner and to control processing errors at an acceptable level?

A broad review of survey objectives against the capabilities of the survey organization is important. Trying to do too much in one survey has frequently led to poor results.

The choice of reference dates and periods for the data to be collected can have a major impact on the levels of non-sampling error. Several examples of recall errors in household surveys were given in Section C:3 of Chapter V and the subject has been treated at length by Som (1968). The effects of memory bias vary greatly for different topics and different target populations. The Indian National Sample survey, after considerable experimentation, adopted different reference periods for different subjects (Rao and Sastri, 1975):

> "Demographic particulars and housing conditions, for example, are collected as on the date of survey, while data on labour force are obtained with a reference of one week. For most of the information on consumer expenditure of household enterprises, the reference period is one month, while data on vital events, for example, are obtained with a reference of one year."

Relevant experience about recall errors associated with the proposed topics and items should be reviewed by the survey planners. If there are gaps in what is known, some experimentation with varying reference periods may be needed in the preparatory stage, or may even be built into the survey itself.

The review may make it obvious that some of the data requirements cannot be met with a single retrospective survey conducted during a short period of time. If the data requirements cannot be changed, then a different approach may be necessary, such as a panel or longitudinal survey in which the same units are interviewed several times during the reference period (say a year), or a "continuous" survey in in which the reference period is divided into several shorter periods, and different subsamples are interviewed to obtain information covering each of these shorter periods.

Survey planners should, at an early stage, develop fairly precise definitions of each of the key variables. Concepts or variables such as income, labor-force status, education, household, food consumption and medical care, to mention only a few, cannot be translated

into sets of survey questions unless they are defined in considerable detail. Furthermore, there will be no chance of measuring and controlling non-sampling errors unless the concepts are clearly defined.

The initial definitions may have to be modified later to take account of practical difficulties encountered in pre-testing, but they are essential in order to start the development of instruments and procedures for any survey. A detailed discussion of the process of defining the "variables" for topics to be investigated in a survey is available in a set of guidelines prepared on behalf of the Inter-American Statistical Institute for use in experimental and pilot survey activities (Jabine, 1981, Chapter B).

One of the issues that may arise in the selection of survey items is whether to retain an item in the same form in which it was used in a prior survey or to change it in a way that is expected to reduce the non-sampling error associated with it. The main argument for keeping the old item is to maintain the comparability of data over time, with a subsidiary argument being that estimates of change are much less likely than estimates of level to be affected by bias. The latter argument should not be relied on too heavily. The underlying conditions in which the surveys are taken may change, and even a small change in the relative or absolute bias associated with an estimate of level can lead to serious bias in an estimate of change. The choice will depend on the survey planners' view of the prospects for reducing non-sampling error by changing the item: if the prospects are considered good, the change should be made. If continuity of the time series data for an item has special importance, a "bridge" between the old and new items can be established by including both versions in a few rounds before dropping the old item.

Finally, at this stage, it will be useful to do something that is often left until much later, namely to prepare a rough outline of the main tabulations that will be made from the survey data. While this will not necessarily have much effect on non-sampling error in the narrow sense, it may do much to ensure that relevant results are obtained from the survey. Every draft of the survey questionnaire(s) should be carefully checked against the table outlines to be sure that all of the necessary variables can be derived from the questionnaire items.

D. Survey design and structure

Chapter III.D, Part One of the Handbook covers several important aspects of planning a survey, apart from the determination of the subject-matter. Issues that have special importance for the control of non-sampling errors are discussed in this section.

The first issues concern the definition of the target population: what kinds of units shall be included and which of these units should be included?

The main kinds of units of observation in household surveys are persons and households. Other units, such as families, spending units, consumption units, and holdings or other (household) enterprises may be used in connection with certain topics. The survey planners should decide early which particular units of observation will need to be identified for purposes of tabulation and analysis. These units should be carefully defined so that operational procedures can be developed to identify the units of observation that are associated with the selected sampling units.

International recommendations exist for the definition of households and other units of observation (see the Handbook, and United Nations, 1980a). These recommended definitions reflect much practical experience in censuses and surveys and their use will promote the comparability of survey data from different countries. However, there are two other important considerations in deciding what definitions to use in a particular survey. First, the proposed definition must be one that fits the usual living arrangements of the target population and can readily be applied by survey interviewers. Failure to meet this requirement can increase coverage and response errors. Second, the definitions should be consistent with those used in censuses and other surveys of the same population, so that data from the different sources can be more readily compared.

Certain population groups are frequently excluded from national household surveys: these include nomads and others with no fixed residence, persons in certain types of institutions, and some members of the armed forces. In part, these groups are excluded because many

of the survey items do not apply to them or are not of particular interest. The kinds of data needed by policy-makers and other users for nomads, unassimilated tribes and other special groups may be quite different from those needed for the bulk of the population and may require entirely different survey instruments and procedures.

On the other hand, some population groups may be excluded on the basis of cost, even though it would be considered useful to have the survey information for those groups. In a formal sense, such exclusions do not constitute coverage error, provided users are clearly informed which groups have been left out; however, the exclusions can affect the relevance of the survey results.

The method or methods of data collection chosen for the survey clearly influence the types and levels of non-sampling error which may affect the results, and hence the measures needed to control non-sampling errors. As stated in chapter I, this document is primarily about household surveys in which data are collected by face-to-face interviews. Opportunities to use other methods such as direct observation or measurement, self-enumeration or telephone interviewing are rather limited in developing countries as a result of lower literacy and lack of communication facilities. However, these methods should not be ruled out automatically. The Handbook provides the following discussion of the advantages and disadvantages of alternate methods of data collection:[1/]

"Direct interviews

Direct interviews usually achieve higher cooperation and response rates and more complete and consistent data. In highly complex fields or involved multisubject undertakings, there may be no feasible alternative, even where literacy is high. The disadvantages of direct interviewing are the higher costs and manpower requirements and the need for extensive training of field staff and close supervision over the data collected.

Observation or measurement

Alternative techniques using direct observation and measurement are sometimes used in complicated surveys.

1/ This disucssion is from Chapter IV.C, Part One of the Handbook; however, the decision on data collection methods has such a large effect on subsequent stages of development that it seemed appropriate to place it before the discussion of sample design and selection.

An example of observation is a food consumption survey where the interviewer visits every day and records consumption on the basis of questioning the respondent and observing what food has been prepared. An even more intensive approach is where the interviewer actually measures or weighs the ingredients that go into food preparation. Another example of measurement is where survey respondents are examined by health practitioners in the course of a health-related survey.

These intensive techniques are aimed at providing more reliable data than are obtainable through more conventional approaches. These kinds of efforts are usually exceptionally costly and often require highly trained personnel, so that sample sizes are likely to be small. Also, close scrutiny of this kind can influence the behaviour of respondents and result in atypical findings.

Self-enumeration

One alternative in more developed countries and in the more literate populations of developing nations is self-enumeration, that is, completion of survey questionnaires by the respondents themselves. The questionnaires can be distributed by mail or picked up by field agents. Generally, it is advisable to restrict the length of self-enumeration questionnaires in order to avoid confusion and reduce non-response. However, some rather burdensome tasks are sometimes attempted in this way, such as by asking persons to maintain diaries for days or even weeks on such topics as consumption expenditures and use of time. In these more complex applications, it is usually advisable to provide for personal pickup and on-the-spot review of the completed forms by field agents.

The advantages of self-enumeration generally include smaller field costs and lower personnel requirements. Also, in theory at least, there is the opportunity for the various family members to consult each other in providing the information. To the extent that this is done, the survey data can be more accurate than where a single respondent reports for the entire household, as is often the case in interview surveys.

Disadvantages include likely lower co-operation and completion rates and less consistent responses. Where these problems are great, a significant follow-up effort by personal visits or telephone may be required to salvage the information. Also, as indicated, the scope of the survey must usually be limited with self-enumeration, and the volume of editing and coding at the data processing stage is usually greater than for other data collection methods.

Different methods of data collection are not necessarily mutually exclusive: direct observation or measurement and self-enumeration, including the keeping of diaries by respondents, can sometimes be used

in conjunction with personal interviewing to enhance the quality of the information obtained or to collect more detailed data. Insofar as alternatives to direct interviewing are more expensive or taxing, they may be confined to a subsample, employing two-phase sampling and ratio or regression estimates. The usefulness of these approaches depends upon the type of subject matter and circumstances, as well as on the survey objectives. For example, while accurate data on physical disability may require intensive investigations by medically qualified field staff, approximate and more extensive methods using lay-interviewers may meet the initial data requirements of developing countries which have little prior information on the incidence and distribution of physical disability in the population (UNICEF, 1980).

Returning to the assumption that direct interviewing will be the primary data collection method, careful attention must be given to the scheduling and timing of the field operations and the availability of field staff to carry out those operations. Some considerations that relate directly to non-sampling error are as follows:

(i) To avoid large interviewer variance effects, there should be some minimum number (perhaps 10) of interviewers working in each area for which separate survey estimates are planned.

(ii) Use of a full-time or permanent part-time field staff is preferable to use of interviewers recruited on an ad hoc basis for each survey.

(iii) If an ad hoc staff is to be used, this may require scheduling the data collection at certain times of year, e.g. during school vacations when teachers and senior students are available.

(iv) Access to some geographic areas or population sub-groups may be difficult at certain times of year due to flooding, extreme heat or cold, etc.

(v) If there is a fixed reference period (see Chapter V, Part C.3), the survey period should be kept short to avoid recall errors. The importance of this consideration depends of course on the particular topics to be covered.

(vi) On the other hand, a longer survey period provides a greater opportunity to control interviewer error through appropriate supervision and quality control techniques. One way to accomplish this without increasing recall error is to use a rolling reference period, provided this is not in conflict with the survey data requirements.

Final decisions on these questions clearly depend on many factors, most of which are specific to the particular setting in which the survey is to be conducted. In accord with the principles of total survey design, the effects of alternate approaches on sampling errors and costs, as well as on non-sampling errors, must be taken into account in reaching decisions.

E. Sample design and selection

The primary concern in developing a sample design is efficiency with respect to the control of sampling errors. However, the choice among broad classes of designs, especially the mix of area and list sampling at various stages of selection, also has important implications for non-sampling error, particularly coverage error. Furthermore, the execution of the design -- the selection of sample units at each stage -- offers many chances for the introduction of significant bias. Control measures to avoid such biases are necessary.

The nature and control of coverage errors from many of these sources was discussed at length in Chapter III. A recapitulation is included here for the sake of completeness and convenience.

1. Use of interpenetrating samples.

One sample design option that has proved useful for many purposes is the use of interpenetrating or replicated samples, which was introduced by Mahalanobis (1946). The general idea is to use a design in which the total sample consists of several sub-samples, each of which is a separate probability sample of the target population, and can therefore be used to produce separate estimates of totals, means and other statistics. The sub-samples may or may not be selected independently. Separate samples can be selected at any stage of selection. At the one extreme would be a

design in which two primary sampling units were selected independently from each stratum. At the other extreme, the sample households in each ultimate cluster could be allocated to two or more sub-samples by a random or systematic process. The choice depends on the purposes for which the sub-samples are to be used.

Interpenetrating or replicated sub-samples are used for many different purposes including: variance estimation, to allow for partial sample rotation in panel surveys, as a basis for variations in content, to provide flexibility in final determination of sample size, and to permit "built-in" tests of alternate collection methods. Except for the last of these, these uses have relatively little to do with control of non-sampling error.

However, the method can also be used in various ways to control non-sampling errors. Some examples are:

(i) If the assignments of two or more interviewers within primary sampling units are randomized, it will be possible to make more meaningful comparisons of measures of interviewer performance such as completion rates (a measure of non-response), production rates, and error rates based on reviews of completed questionnaires.

(ii) If sub-samples of primary sampling units have been selected, it will normally be possible, using census data which are available by PSU, to make separate estimates of census totals or means from each sub-sample. If there have been any gross errors in the selection process, comparison of the sub-sample estimates may expose these.

Because of their potential value for these and other purposes, the option to use some form of interpenetrating or replicated samples should be considered very early in the development of the sample design.

2. Development and use of area frames.

Commonly, political subdivisions such as regions, provinces, districts, cities and metropolitan areas are used in forming primary sampling units and in grouping the primary units to form strata. The smaller of these units may also be used for sampling within primary sampling units. If the area definitions of these units are clear and remain stable over

time, their use for these purposes should be straightforward.

In some countries, however, there are fairly frequent changes, especially in the smaller units. New units are created, sometimes by splitting existing units into two or more parts, and sometimes by more complicated processes of splitting and recombination. In addition, boundaries of existing units may be altered without creation of any new units. An illustration of how these changes can affect the number of units at various levels is given for Thailand in Table 7.1 (Skunasingha and Jabine, 1981).

Table 7.1 Administrative subdivisions of Thailand by type: 1970, 1980 and 1981

Type of unit	Number of Units			Percent increase	
	1970	1980	1981	1970-1980	1980-1981
Province	71	72	72	1.4	0.0
District	580	698	709	20.3	1.6
First-class	540	621	621	15.0	0.0
Second-class	40	77	88	92.5	14.3
Subdistrict	5,134	5,930	6,080	15.5	2.5
Village	45,538	53,387	54,405	17.2	1.9
Municipal area	120	118	122	-1.7	3.4
Sanitary district	641	709	714	10.6	0.7

For any survey, but especially in countries where the political subdivisions are less stable, it is essential to verify that the names and geographic definitions of the political subdivisions that will be used in the sample design and selection process are up-to-date and accurate. Sometimes checks with local officials may be needed to determine the current situation; up-to-date information on the smaller units may not always be available from a central location.

For sampling within primary units, difficult decisions must often be made. Smaller political subdivisions may be less well-defined: on the other hand, maps available for the creation of other secondary area sampling units may not have adequate detail to provide easily identifiable boundaries for units of the size needed for the survey. Naturally, any maps and other materials from a recent census should be evaluated for possible use. Another important consideration is whether residents of an area can accurately identify the political subdivision in which they reside, if that is to be the basis for their inclusion in or exclusion from the sample.

The survey designers must weigh these factors, decide which general approach is likely to work best, and then do their best to control errors in the execution of the design. This means, in particular, careful attention to field operations for mapping and/or listing and to relevant office procedures, including:

(i) Emphasis of techniques to minimize coverage errors in written instructions and training for these operations.

(ii) Systems for reviewing and evaluating completed work, at least on a sample basis.

(iii) Using unbiased selection procedures at all stages. It is not, in general, unbiased to select a sample of area units and then revise the boundaries of those units to make them more readily identifiable. Any such revisions should be carried out prior to selection, for all potential sample units located within the next higher stage sample units from which they are to be selected.

3. Development and use of list frames.

The use of unmapped political units such as villages as sampling units should be avoided unless there is no alternative. It is usually difficult to obtain a fully up-to-date list, and the possibilities for coverage errors are much greater than when well-defined area units are used.

A more desirable approach is to use area sampling units of the smallest size feasible, and then to proceed with listing of households

within the selected area units and, if necessary, sampling of households from the listings. To minimize coverage error, the selection and interviewing of sample households should be done soon after the listing. The same considerations suggest that census listings should not be used directly as a sampling frame except for surveys taken very soon -- say, not more than six months -- after the census enumeration.

As mentioned in connexion with area frames, the level of coverage error is highly dependent on the quality of the listing operation, so that the development of instructions, training and quality control procedures must be done carefully. One effective method of control that is often possible is to compare household and/or population counts for each area listed with corresponding counts from the most recent census. While precise agreement is not expected, large differences should be investigated and will often turn out to be due to confusion about area boundaries or simply to careless work. If census counts are not available at the lowest level of areas listed, counts from the listings can be inflated to give estimates at a higher level, say, for each primary sampling unit or stratum, and the comparisons made at that level.

4. Choice of ultimate sampling units: rules of association.

The ultimate sampling units for household surveys are usually either housing units or households. Since households are likely to be used also as units of observation, it will normally be more efficient from the sampling point of view to use them as the ultimate sampling unit. However, as described in the Handbook (Part One, p. 37), non-sampling error considerations suggest that there are significant advantages in using the housing unit as the ultimate sampling unit.

In either case, clear-cut rules of association must be established to link units of observation to the ultimate sampling units. Although other possibilities exist, errors of application will probably be minimized by using rules which uniquely link every unit of observation with one and only one ultimate sampling unit.

In associating persons with households, a choice must be made between a de jure (usual residence) rule and a de facto (current location)

rule. The choice should take into account potential effects on the level of coverage error. Once a rule is chosen, interviewers must be trained in its application, and the survey questionnaire used to list household members should include suitable questions to identify all persons who should be associated with the household and to exclude all others.

If the survey is to cover agricultural holdings or other kinds of household enterprises, association rules will also be needed to link these units of observation with households.

5. Centralized vs. decentralized sample selection.

When possible, it is preferable for all stages of sample selection to be carried out at a central location, under close supervision of the staff directly responsible for the sampling procedures. Sampling in field offices or at interviewing sites leaves many opportunities for bias, intentional or otherwise. Selection procedures carried out centrally can be immediately checked and any errors corrected. On the other hand, if sampling is carried out incorrectly in the field, errors will probably not be detected until most of the interviewing has been completed, and it will be difficult or perhaps impossible to make corrections.

Nevertheless, there are often situations where considerations of cost and timing require some decentralization of sample selection procedures, for example, the selection of sample housing units or households from listings. When this is done, the following precautions should be observed:

(i) The procedure should be made as simple as possible.

(ii) Suitable forms and instructions should be provided.

(iii) If possible, the selection should be made by someone other than the interviewer to whom the selected units will be assigned.

(iv) All documentation of the sample selection process should be retained and reviewed by the central staff as soon as possible.

In some circumstances, in spite of the potential bias, there may be substantial savings in cost and time from a design in which the same

individual does the listing, sample selection, and interviewing in a single visit to the listing unit. This would be especially advantageous for listing units in remote locations and in sparsely populated areas, where the associated travel costs are large. It is recommended, however, that this approach not be adopted operationally unless it has been carefully tested and evaluated, and that it be used only where highly qualified field staff are available.

Whether sampling is centralized or decentralized, full documentation of the sample selection at all stages is essential. The sample selection procedures and worksheets should be carefully preserved. Besides providing a basis for review of the selection, they will be needed later on in connexion with the derivation of sample estimates and estimates of sampling errors from the survey data, and will be useful for special studies of the components of sampling error. In addition, they may be needed for sample selection in subsequent surveys.

6. Checking the results of sample selection.

Sample selection operations at each stage should be fully verified immediately, to avoid errors which can be extremely damaging if not discovered until after the completion of field work. One form of verification is simply to repeat the selection procedure. If possible, this should be done independently (using, of course, the identical frames, and the same random starts and sampling intervals) and the resulting selection compared with the original.

A rough check can be made by estimating in advance the number of units to be selected, broken down by region, stratum, or other categories, and comparing these estimates with the counts of units actually selected. This sort of check will probably detect failure to process some significant part of the sampling frame.

Where the sampling units in the frame have associated information which is not used directly to determine their selection probabilities, this information can be used in another kind of check. Values for the units selected are weighted appropriately to produce estimates of population totals, and these are compared with the known population totals.

Differences that exceed what might be expected as a result of sampling error indicate probable errors in the selection process or in the design itself. This technique can be applied at various levels of selection. For example, suppose that households are listed and sampled with equal probability, and that a count of persons is obtained for each household listed. The mean number of persons per household for the selected units can be compared with the mean for all households listed to determine whether the selection may have been biased in favor of large or small households.

A useful discussion of these kinds of sample checks is given by Simmons (1972, pp. 21-22).

7. Re-use of sample units.

Considerations of sampling efficiency often lead to the use of designs with complete or partial sample overlap in successive surveys. The advantages of sample overlap are indisputable, especially when the measurement of change is a major survey objective. However, it should not be overlooked that certain kinds of non-sampling error may be associated with repeated use of the same sample units. Some of the ways in which this can occur are:

(i) If a fixed sample of primary units is retained for a long period, it may become unrepresentative, possibly as a result of deliberate efforts by the government to allocate its resources for various economic and social programmes disproportionately into the sample areas.

(ii) Where a listing and sub-sampling approach is used, coverage bias will be introduced unless the listings are frequently updated.

(iii) Where the same households are interviewed in more than one survey, "panel bias effects", as described in Chapter V Section C.4, may be introduced. The most obvious manifestation would be a gradually increasing refusal rate in successive rounds of interviews. To limit this effect, the burden on respondents, as measured by the number,

length and frequency of successive interviews, should be kept at a level which does not produce a significant increase in non-response. There are also many other potential panel bias effects, and these may be more difficult to identify and control. Where a partially rotating design is used, comparison of estimates from different sub-samples (rotation groups) will provide some indication of the extent of the problem for key survey items.

F. Two strategic decisions on data collection procedures

Two important decisions about data collection procedures are likely to have a major impact on both the cost of the survey and the level of non-sampling errors: the respondent rules and the follow-up rules. Respondent rules specify which members of sample households are acceptable or preferred respondents; follow-up rules specify the number and timing of call-backs and other efforts to obtain a completed interview or at least partial information for each household. The two decisions are, of course, interrelated. A stringent respondent rule (for example, to interview each adult member of the household separately) will mean that, on the average, more call-backs will be needed to complete the interviews. A common practice is to relax the respondent rules (for example, to permit the use of proxy respondents) after one or more visits to a sample household.

Survey planners should make conscious decisions on what rules are most appropriate in a particular survey, and should emphasize observance of the adopted rules in instruction manuals, in training sessions and in supervision of the field work. If this is not done, field staff are likely to take the easiest way out, and the quality of the results will be adversely affected.

These two aspects of data collection procedures will now be discussed in more detail:

1. Respondent rules.

The selection of respondents for household interviews is a matter of considerable importance. Uninformed respondents can be a source of

serious response errors; at the same time, flexibility in the choice of respondent can result in considerable savings in cost, since fewer call-backs will be needed. Much depends upon the kinds of questions to be asked, the type of survey, and the conditions under which it is conducted. In many societies, for example, experience has shown that information on income and economic activity can best be provided by the male head of the household, while information on family composition and certain consumption items, such as food, may be best provided by the female head. In surveys that include questions about individual household members, the issue arises as to whether one person can be interviewed to obtain the information for an entire household (proxy interviewing), or whether each adult member should be interviewed. Alternatively, the information may be obtained from any suitable adult, but seeking as far as possible the participation of other members who may be present at the time of the interviewer's visit (see, for example, Scott and Singh, 1980). When detailed individual data are needed, and especially when attitudinal or sensitive questions are included, there can be little doubt that each household member concerned should be interviewed, preferably in private.

Empirical research is needed to determine the relative costs and advantages of various choices. Hansen and Waksberg (1970), in reporting on the results of some methodological experiments connected with labour force surveys in the United States, reported that:

> "Erroneous reporting by non-self-respondents tends to result in fairly serious errors of the more volatile labour force categories such as unemployed or part-time workers. The effects on more stable labour force items, however, are fairly small. For example, full-time workers are subject to a reporting error of only about 1 per cent."

The authors also discuss the effects of obtaining information on other topics from proxy respondents:

> "For other subjects, however, we have found a marked deterioration in reporting when a self-respondent is not used. Health conditions and disability appear to be particularly affected. A recent analysis of a report on the extent of cigarette smoking also showed sharp differences in quality. We know of no guides to determining rules for selecting the proper respondent that are generally applicable and will produce efficient procedures under a wide variety of circumstances."

Other results from empirical research comparing self and proxy responses in surveys were given in Chapter V, Section C.5.

If reasonable efforts to interview one or more household members are unsuccessful, it may sometimes be possible and desirable to obtain partial information from neighbours or other local sources. This should not be attempted, however, for subjective or sensitive items.

2. Follow-up rules.

Follow-up rules determine the amount and kinds of effort that interviewers shall make to obtain completed interviews for sample households, subject to the prescribed respondent rules. They should cover such matters as:

(i) The conditions under which attempts to interview should be terminated, regardless of the number of visits already made, for example, in the case of refusals. An alternative to termination is to send a supervisor or especially skilled interviewer to try to "convert" the refusal to an interview.

(ii) The number of visits to make to a household before i) relaxing the respondent rules, or ii) abandoning the attempt to complete an interview for that household.

(iii) The timing of initial visits and subsequent visits to sample households. For some households, if all visits are made during normal working hours, there will be little to gain from call-backs after the first visit.

(iv) Variations in rules for different kinds of areas. The cost of call-backs is likely to be much higher in sparsely settled areas and it may not be cost-effective to require several call-backs in such areas.

(v) In situations where each call-back requires travel to and from a cluster of sample households, the rules might take into account the number of households in the cluster for which interviews have not yet been completed, as well as the number of visits already made. For example, work in a cluster might be terminated i) after four call-backs, or ii) after interviews have been completed for 90 per cent or more of the sample households.

There is no simple formula for deciding precisely what rules to adopt. The survey planners should consider the extent to which key data items may be correlated with the number and timing of call-backs needed. Typically, the households more difficult to contact are those which are small, have no young children or elderly persons, and are located in urban areas. Clearly, data on family composition and labour force

are likely to be biased if such households are significantly under-represented among the completed interviews.

The costs associated with alternative follow-up and respondent rules depend very much on the target population and the environment in which the survey is to be conducted. If adequate data on costs are not at hand from previous surveys, they should be obtained from at least one pretest, as discussed further in Section H of this chapter.

G. Development of questionnaires

Anyone who has done much survey work can cite "horror stories" of errors induced by faulty questionnaire design. Many of these disasters could have been avoided if more careful attention had been given to questionnaire development, including appropriate interaction with both data users and potential respondents.

The content of questionnaires is determined by survey data objectives: however, an adequate translation of the latter into the former is a highly demanding task. One of the most difficult problems in survey preparation is the translation of complex ideas and concepts into usable questions. The variables to be measured by the survey have to be transformed into operational concepts and expressed in the form of a logical series of questions which the interviewer can ask and the respondent can comprehend and answer. The structure and format of the questionnaire must facilitate the work of interviewers in asking questions and recording responses, and the work of coders and data entry clerks during the processing stage.

Questionnaire design is discussed in Part One, Chapter IV.C of the Handbook. There are several useful texts available. The Art of Asking Questions by Payne (1951) is a classic; it is concerned mostly with the development of individual questions, especially those dealing with attitudes, knowledge and beliefs. Other useful references are Bradburn and Sudman (1980), Hoinville et al (1979, Chapter 3), Lininger and Warwick

(1975, Chapter 6), Moser and Kalton (1978, 2nd edition, Chapter 13), Sirken (1972), U.S. Bureau of the Census (1979, Part 7-B), and Zarkovich (1966, Chapter 5.3). These authorities are not always in full agreement about particular points; nevertheless, several useful guidelines have emerged and there has by now been enough methodological research to provide experimental support for many of them.

This section lists and discusses those aspects of questionnaire development that are considered to be most directly relevant to the control of non-sampling errors.

1. Qualitative design work.

Assuming that the data requirements for a survey have been reasonably well specified, the first question is whether enough is known about the topics to be covered to start drafting one or more structured questionnaires. If the same topics have been covered in previous surveys of the same population, the answer could be "yes", but if this is not the case, some qualitative design studies should be undertaken first. The primary techniques used are group discussions (sometimes called "focus group interviews") and in-depth individual interviews. These techniques have extremely valuable functions for identifying relevant types of behaviour, attitudes and issues, and developing a better understanding of the framework in which potential respondents think and talk about these matters. The knowledge obtained makes it possible to design a questionnaire that "... avoids forcing respondents' views into a false or irrelevant structure". (Hoinville et al, 1979).

Whether individual or group interviews are used, the interviewer/moderator will have a list of topics based on the survey data requirements, and will ask whatever questions are needed to obtain the desired information, without restrictions as to the order, number and wording of questions. In group interviews, a few introductory questions by the moderator will often generate a spontaneous discussion that covers most of

the topics on the list. Respondents or participants may introduce new related topics and such discussion should be encouraged as long as it is relevant to the survey objectives: however, the interviewer or moderator must be sure that all of the listed topics are covered eventually. Obviously, certain skills are required of interviewers and group moderators, and these can be developed only with practice.

To-date, relatively few government statistical organizations have made use of these techniques: however, their value as a first stage in questionnaire development is becoming better recognized. At a relatively low cost, they make it possible to identify and resolve many problems that would otherwise surface only during the survey data collection phase or even later.

Fuller descriptions of qualitative design methods for surveys are given by Hoinville et al (1979, Chapter 2) and Jabine (1981).

2. Number of questionnaires needed.

There are many reasons why it might be necessary to use more than one questionnaire in a particular household survey. Some of these are:

(i) Versions of the same basic questionnaire for different language groups.

(ii) Questionnaires for successive rounds of a multi-round survey.

(iii) Separate questionnaires for agricultural holdings or other enterprises associated with households.

(iv) Special questionnaires for "rare events", such as births and deaths occurring during a short reference period.

(v) Separate versions for areas for which the data requirements are different, for example, urban and rural areas.

Sooner or later in the development work these needs should become obvious; however, it will be better to ask the question early, so that development of the necessary versions can proceed in an orderly way.

In general, it can be expected that the use of questionnaires tailored to specific needs of respondent groups will reduce the level of non-sampling errors. Additional design and printing costs, as well as more complex processing requirements, must be weighed against any such benefits. Where data for individual households are to be recorded on more than one form and then linked at the processing stage, a careful system of identification and controls will be needed.

The difficulties of constructing equivalent versions of a questionnaire in more than one language should not be underestimated. Different language groups usually have different cultural patterns. As pointed out by Lininger and Warwick (1975), there are:

> "...four interlocking aspects of equivalence in questionnaire design: 1) comparability in the salience and meaning of the concepts, 2) functional equivalence in operational definitions, 3) linguistic equivalence through translation, and 4) comparability in the responses."

For multi-language questionnaires, several special techniques, including back translation, group interviews and interviews with random probes, may be needed to supplement more traditional methods of pretesting for each version.

3. Questionnaires vs. schedules.

A matter of considerable controversy among survey practitioners is whether to use verbatim questions which are supposed to be asked exactly as written (the "questionnaire approach"); or merely to list the items of information needed (the "schedule approach"). The argument in favour of verbatim questioning is that it structures the interview and reduces variability between interviewers in the way in which they put questions to respondents. This can be especially important where responses depend critically on the exact wording of the question asked. On the other hand, the simple listing of items saves space, and allows flexibility to the interviewer to "size up the situation and ask the questions

in a form that is most understandable to particular respondents... The need for flexibility may be especially great when translation into local dialects is necessary" (United Nations, 1981). The important thing is not to be dogmatic about one approach or the other. The appropriate approach will depend upon the context. For many straightforward factual items, the differences may be merely formal, and the simpler and briefer the question, the better it is. Second, merely writing down a detailed question is no guarantee that its meaning will be appreciated by the interviewer or the respondent. The basic issue is that of communication, as noted by Mauldin and Marks (1950):

> "In these communication links (between the respondent and the interviewer), the literal meaning of a statement or question is less important than the particular meaning attached to the statement or question by the respondent and interviewer. In particular, the cultural context of a question is extremely important. This point is illustrated by the question on school attainment. The question in the form: 'What is the highest grade of school you have completed?', interpreted literally should offer no difficulty for the vast majority of people in answering it correctly. However, many people hear the term 'highest grade of school' and immediately think in terms of the highest grade of school they 'went to'. Some people don't even hear the final word in the question, that is, the word 'completed'. We could obtain answers to the above question and then say, 'Did you complete that grade?', and a substantial number would answer 'No'. One's conceptual span is quite limited - even under ideal conditions this is true, but under conditions of the normal survey, respondents do not always devote their undivided attention to the interview. The question we have just quoted demands that the respondent focus on both highest grade and completed. This may be conceived of as a subtle form of the 'double barrelled' question. We have deemed the problem in connection with highest grade of school completed to be sufficiently serious to warrant two questions instead of one. So, we now ask, 'What is the highest grade of school you have attended?' Then, the respondent is asked, 'Did you finish this grade?' This eliminates part of the response bias for this particular item."

Hence, the issue in the above example is not between the questionnaire version "What is the highest grade of school you

completed?" or the schedule version "Highest Grade Completed"; rather the issue is whether better information can be obtained by asking two separate questions, in whichever form.

It might be argued that the schedule approach requires more highly trained interviewers, because they must understand each concept well enough to devise appropriate questions for every situation. However, some reflection will make it clear that interviewers using questionnaires require the same kind of training, because the structured questions will not always elicit acceptable responses and interviewers will have to frame suitable followup questions (probes).

There has been little, if any, research to directly compare these two approaches; for the time being, the choice must be based on the questionnaire (or schedule) designer's judgment as to which is better suited for the particular survey. It is, of course, possible to use the questionnaire approach for some items and the schedule approach for others. It would probably not be appropriate to use the schedule approach for items concerning attitudes, perceptions, or beliefs, since responses to these items are especially sensitive to precise question wording, and usually cannot be tested against any external standard.

4. Accommodation to data processing requirements.

Once a rough draft of the questionnaire is available, the designers should meet with the data processing staff to discuss how the completed questionnaires will be processed, and the implications of the selected processing methods for the design of the questionnaire. Appropriate provisions must be made for any codes that will be entered manually on the questionnaire. The format of the questionnaire should facilitate transference of information from it whether this is to be done by manual transcription or by data-keying direct from the questionnaire. Precoding of response categories is helpful in eliminating many errors that might occur if the codes had to be entered manually by interviewers or coding

clerks. The arrangement of items on the questionnaire should make it easy for transcribers or data entry clerks to proceed from one item to the next.

Once the contents of the survey have been determined, the data processing experts should be fully involved along with the subject matter specialists in the design of questionnaires so as to ensure their processibility. While the guiding principle in the design of questionnaires has to be the collection of most accurate data possible, the requirements and convenience of processing must also be given their due importance. Of course, in the case of serious conflict between data collection and data processing requirements, priority should be given to the former; it should however, be appreciated that there are forms of questionnaire design and coding schemes that lead to simplification of the data processing task without adversely affecting the field work. In fact, a well-designed questionnaire layout can assist both the operations.

A detailed discussion of the advantages and disadvantages of different formats for data processing can be found in the companion NHSCP document "Survey Data Processing: A Review Issues and Procedures" (United Nations, 1982; especially Ch. 2).

5. Review of draft questionnaires.

Every draft of a questionnaire should be carefully reviewed by the designers themselves and by others, preferably including qualified experts from outside the survey organization. Subject matter specialists will review the questions to make sure that they are appropriate for the survey data requirements. In addition, technical features of the questionnaire should be reviewed to spot common faults that are likely to lead to various kinds of non-sampling errors. Important features to check are:

(a) The identification of "cover" section of the questionnaire. Does it include items needed.

(i) To locate the unit and the initial respondent for follow-up or evaluation interviews?

(ii) To assign geographic codes?

(iii) To assign appropriate weights for tabulation and variance estimation?

(iv) To uniquely identify each assigned unit of observation for purposes of control and cross-reference to other forms used in surveys?

(v) To describe the data collection effort (e.g., number and timing of call-backs) and outcome (completed interview or incomplete, with reason) for each unit?

(vi) To control the quality of work done by individual interviewers and data processing clerks?

(b) The manner in which different items are grouped. Items that apply only to selected sub-groups, for example, women of child-bearing age should usually be combined in one section or module. All items on a particular topic, such as education or labour force participation should usually appear together.

(c) Transitions from one section or module to the next. If the questionnaire approach is used, appropriate transition statements should appear on the questionnaire whenever there is a significant change from one topic or frame of reference to another. This is very important when changing reference periods, for example, when switching from questions about purchases in the past week to purchases in the past month.

(d) Skip instructions. The order of questioning for a person or other unit of observation is often contingent on responses to particular questions. Instructions for skipping or answering particular questions should usually be placed immediately following the responses on which they depend. Complex skip patterns should be avoided. Instructions may be verbal or in another form, such as arrows to the next item, but should be easy for the interviewer to follow. Instructions that require turning back to check answers on an earlier page of a questionnaire should be avoided. Making a flow chart of the questionnaire, as described by Sirken (1972), is an excellent method of insuring that proper instructions are provided, and may suggest ways of simplifying skip patterns.

(e) Provisions for recording answers. There are many different ways of recording answers for fixed response categories: check boxes, circling codes or responses, entering numerical or alphabetic codes, entering checks or X's in a column corresponding to the category, etc. There is not much experimental evidence on which to base choices among these methods (for an exception, see Wright, 1980); however, for reasons given earlier, systems requiring coding by interviewers are not recommended. Insofar as possible, the same system should be used throughout the questionnaire. Where similar response categories appear in different items (for example, "yes" and "no"), they should be in the same order and, if precoded, have the same codes. Where open-end responses are required, adequate space should be allowed for the maximum length of response expected. Where amounts are required, the units to be used should be clearly indicated, for example, hectares, thousands of dollars, days, etc. If more than one type of unit can be used to report an item, there should be a separate space (or possibly a set of check boxes) to record the units used.

(f) Provisions for distinguishing answers of "zero" or "none" from non-response. Where entries of amounts are called for, the recording format should make it possible to distinguish answers of zero or none from cases where the item was overlooked or not answered for some other reason. The illustration in Figure 7.1 is taken from the U.S. Bureau of the Census (1979).

Figure 7.1 -- Two methods of asking for amounts

Method A

Last month, how much income did you receive from --

(a) wages or salary? _____

(b) pensions or annuities? _____

(c) unemployment insurance? _____

Method B

1) Last month, did you have any income from --	(If yes) How much?
(a) wages or salary? ☐ yes ☐ no	(a) _____
(b) pensions or annuities? ☐ yes ☐ no	(b) _____
(c) unemployment insurance? ☐ yes ☐ no	(c) _____

Method B is likely to produce more accurate data, as it requires the interviewer to make a specific entry for each item.

(g) Provisions for coding and data entry. These were discussed earlier. Questionnaire drafts should be reviewed by the data processing staff to determine whether their requirements have been met, to the extent they are not in conflict with other relevant considerations.

(h) Clarity of questions. Individual questions should be reviewed for clarity and lack of ambiguity. Instead of asking "How old are you?", it is better to ask "How old were you at your last birthday?" or "What is the date of your birth?" (Less precise questions may be used for respondents who are not able to answer exactly.) Most questions about events in the past should have specific reference periods, such as "last year, that is, in 1981...", "in the past seven days", "since the beginning of this year", etc. Unfamiliar expressions should be avoided. As pointed out by Lininger and Warwick (1975), some respondents might be confused by the question "What is your marital status?", but would be able to say whether they are married, widowed, divorced, separated, or have never been married. Questions on income from self-employment must make it clear whether gross or net income is wanted: however, the terms "gross" and "net" may be unfamiliar to many respondents, and other ways of asking the question should be developed. Subjective questions on attitudes, perceptions and beliefs require special care: small changes in wording may produce quite different responses. For example, in an experiment described by Rugg (1941), the question was asked: "Do you think the United States should not allow public speeches against democracy?". In another survey, the words "not allow" were replaced by "forbid", and the number answering "yes" to the question dropped by 20%.

(i) Difficulty of questions for respondents. Each item should be reviewed for the difficulty of the task that it imposes on respondents. Does it require respondents to recall events or transactions that happened over a long period of time? Does it require them to recall events that may not have had any particular importance to them? Does it require them to grasp concepts with which they are not likely to be familiar?

If the answer to any of these questions is "yes", it does not necessarily mean that the item should be eliminated. However, it does mean that steps should be taken to aid and encourage respondents in the tasks they are asked to perform. There are several things that can be done:

(i) Break the item down into several parts. If the objective is to obtain total income of the household, ask separate questions for each adult member and/or for different sources of income. If the object is to identify household enterprises, ask separate questions about different types of income-producing activities carried on in or near the home.

(ii) Use probes (follow-up questions) to obtain more accurate information. Jabine and Rothwell (1970) report on an experiment in which data on hours worked in the previous week were substantially improved by asking followup questions about time lost due to illness and other causes, overtime work, and work at a second job.

(iii) Instruct interviewers to use verbal reinforcement techniques to encourage respondents in their recall tasks. The reinforcement consists of the interviewer making a statement such as "That's the kind of information we need" or "We're interested in that" when the respondent reports an event or condition which is

responsive to the question asked. In a controlled experiment to evaluate the effects of this technique, respondents receiving verbal reinforcement reported about 25 percent more chronic and acute health conditions than did those not receiving reinforcement (U.S. National Center for Health Statistics, 1977).

(iv) Use interviewing aids. Several suggestions are given in the Handbook:

> "A calendar might be shown to the respondent when inquiring about dates. Certain cards (sometimes called 'check-lists' or 'prompt cards') containing a range of possible answers might be given to the respondent to review in answering a question. An example is a list of chronic health conditions for use in asking about a person's health. Pictures of different sizes of containers or an actual set of measuring cups or spoons might be used in asking about quantities of food used, etc."

To help place events in time, it may also be useful to give the interviewer a calendar showing the dates of events that might be well-known to respondents in various localities, such as traditional festivals, local holidays, school openings and closings, elections, floods, and changes in government.

(j) Single questions incorporating multiple dimensions. Such questions are sometimes called "double barreled". When they are used, it is generally not clear what the answers mean; consequently, they should always be avoided. An example of such a question would be "What type and amount of life insurance coverage do you have?" This question actually has three dimensions: first, whether or not the individual has any coverage: second, for those who do, the type or types of coverage; and third, the amount of coverage. Each of these dimensions should be covered by separate questions.

(k) Sensitive questions. The reviewer should look for questions that are likely to be sensitive to some respondents, leading to non-response or inaccurate response. Topics that are sensitive vary in different areas and cultures; some examples are income, assets, religion, political affiliations and beliefs, sexual practices, and participation in illegal activities. Sensitive questions that are not essential should be dropped. If they are essential, some methods of obtaining the best possible results include:

(i) Careful placement of sensitive items. They should not be placed near the start of the questionnaire. Most authorities suggest putting them close to the end, but not at the very end. They should be placed in the section of the questionnaire containing related items, if there any.

(ii) Use of "warm-up" questions that are related, but of a less-threatening nature.

(iii) For questions on socially undesirable behaviour, use of language that is familiar to the respondent to describe the activities in question. Start by asking whether respondent has ever engaged in the behaviour before asking about current or recent behaviour.

(iv) Use of special techniques such as "randomized response" for especially sensitive items. Keep in mind, however, that this requires particularly well-trained interviewers, and that reductions in bias are achieved at the price of increased sampling errors.

(l) The response categories used in multiple-response questions. The categories should be exhaustive, that is, they should cover all possible answers to the question. If there are some rare categories or if it is not possible to anticipate all the responses, an "other"

category should be included, with an instruction to the interviewer to record the answer in the space provided.

The response categories should generally be mutually exclusive, that is, they should be defined so that no respondent properly belongs in more than one category. This rule does not have to be followed in all cases. For example, one might use the question "Which of the following things have you done to try to find work? (check all that apply)", followed by a list of relevant activities. This may appear to be a more "economical" approach from the point of view of the questionnaire designer; however, it will not necessarily focus the respondent's attention on each of the alternatives given, and it is considerably more difficult to process responses from questions that permit multiple responses.

6. Testing of alternatives.

The experience of others does not always provide sufficient guidance as to what formats or specific questions are likely to be most successful in controlling non-sampling errors. At this stage, if there is still considerable doubt about the proper approaches to use for key variables, consideration should be given to a carefully designed test of alternative treatments in a pretest, or even in the survey itself.

7. Quality control of printing.

Copy sent to the printer and/or proofs received from the printer should be reviewed with great care to spot any typographical errors. When the printed questionnaires are received, they should be subjected to a sample inspection procedure. As stated in the Handbook:

> "It is not uncommon to find that certain batches of the printed form may be too light, or smudged, or missing certain sections, etc. A sufficient sample of each batch should be examined to be sure that these problems do not exist. Otherwise, it may be discovered too late that there is an inadequate supply of usable forms."

H. Pre-testing

A pre-test may be largely a waste of time and effort or it may contribute significantly to the quality of the survey results, depending on the care with which it is planned and carried out. A useful discussion of pre-testing appears in the Handbook (Part One, Chapter IV, C.3). This section highlights certain aspects that are especially relevant to the control of non-sampling errors.

1. Specification of objectives.

As described in the Handbook, there are many possible purposes that can be served by pre-tests. To insure that each pre-test is suitably designed, the main objectives should be listed, in priority order. Under each objective, there should be listed the kinds of data needed to achieve that objective, so that plans can be developed to capture those data during or immediately following the pre-test.

Refinement of questionnaires and data collection procedures will nearly always be among the main pre-test objectives. Some other possibilities are: to estimate unit costs and components of sampling variance, to provide a basis for testing processing procedures, to decide between alternative versions of questions or questionnaire formats, and to test the feasibility of collecting certain kinds of information.

2. Sample size and design.

As for any regular survey, the appropriate sample size and design for a pre-test will depend on its objectives. Sample sizes are typically small, say, from 50 to 300 households. The data collection is usually limited to a few primary sampling units, chosen purposively to reflect a variety of environments for data collection. It is better, if possible, to use probability sampling within the selected primary units.

The selection of interviewers and field supervisors to work in the pre-test is also important. One strategy would be to use only field supervisors as interviewers, so that they can gain experience that will be useful to them in training and supervising interviewers in the main survey. On the other hand, it might be considered

better to use a group of interviewers with education and experience more typical of those who will work in the survey. In either case, there are benefits from making the number of interviewers as large as possible, subject to the constraint that each should have a sufficient workload to justify the training and travel costs.

In considering whether to include formal tests of alternatives in a pre-test, remember that even a sample of 300 households, with as many as 20 interviewers, may not be large enough to discriminate between alternatives with the desired precision. This is especially true if the main issues relate to a small sub-set of the population, such as unemployed persons or migrants. If the differences between treatments are expected to be relatively small, an experiement built into the main survey may be the only satisfactory way to measure them.

3. Scheduling pre-tests.

The need for allowing sufficient time to analyze and apply the findings of pre-tests may seem obvious. Nevertheless, important benefits of pre-tests are often not realized, usually because the analysis has not been planned in enough detail, and insufficient time and resources have been committed to it. A key deadline is the date when the final questionnaire must go to the printer; all pre-test results relevant to the questionnaire must be available before that date.

4. Capturing information from pre-tests.

As in any scientific experiment, the methods for observing and recording relevant data about pre-test operations and results must be planned before the pre-test begins. There are several useful techniques for capturing relevant data and these should be used to full advantage, subject to personnel and other resource constraints. The following should all be considered:

(a) Field observers. Available professional staff should be detailed to observe training and interviewing and to prepare written reports based on their observations. These reports will be most useful if observers are given a list of points

to watch for as they observe and a standard outline to follow in preparing their reports. In addition to observing substantive aspects of interviews, observers can be asked to record the time spent by interviewers on various activities.

(b) Use of the questionnaire as a recording instrument. As described in an earlier section of this chapter, operational details such as the time of the initial visit to the household, the number and timing of callbacks, and the duration of the interview can be recorded on the questionnaire. In addition, one or two open-end questions can be put at the end of the pre-test questionnaire about the respondent's reaction to it: what items were difficult to answer and why, and what was the respondent's general reaction to the interview? Items can also be included in which the interviewer is asked to record his or her impression of the interview and to describe any particular difficulties that were encountered.

(c) Debriefing sessions for interviewers and professional staff. During or immediately following the data collection, meetings can be held involving interviewers and members of the survey design staff to discuss what the former have learned about the questionnaire and interviewing procedures. For maximum benefit from such sessions, the format should be similar to that described earlier for group discussions in the early stages of questionnaire design. The moderator should prepare an outline of the topics to be covered and should be skilled in eliciting relevant comments from all of the participants.

(d) Tape recording. Tape recording of actual interviews and of debriefing sessions can be quite useful. Preparation of transcripts from a small number of interviews can give all members of the survey staff important insights into the problems involved in communicating with respondents. Permission should, of course, be obtained from all persons who are to be recorded in an interview or debriefing session.

(e) Interviewer and supervisor reports. Interviewers and field supervisors may be asked to prepare reports to indicate problems encountered and suggestions for improvements in the questionnaire and procedures. They may be asked to complete structured questionnaires, or to complete narrative reports, following a standard outline, or both.

(f) Reports on processing of pre-test results. Reports of errors found in manual and computer edits, and of errors committed in manual coding and data entry operations should be prepared and analyzed to identify problems that might be overcome by changes in questionnaires or in data collection and processing instructions.

(g) Formal tabulations and analysis of pre-test results. If the pre-test has included any formal experiments, the tabulations necessary to compare alternative treatments must be prepared and analyzed.

I. Interviewer selection and training

The quality of field work can be greatly affected by the amount of care taken in the selection and training of interveiwers.

If the survey organization has a permanent field staff in the selected sample areas, there will probably be no need to recruit new interviewers for a particular survey. A well-trained and supervised permanent field staff is, of course, one of the keystones of an integrated household survey programme. Having a permanent staff does not guarantee that the quality of field work will be high, but it provides a better opportunity to achieve that goal. Achieving high quality with an *ad hoc* staff assembled for a particular survey is much more difficult.

However, if there is no permanent staff, or if it needs to be supplemented, or if there are special requirements for the survey, interviewers must be selected. What criteria should be used? There are no simple answers to this question: clearly, the attributes of "good" interviewers will vary depending on the cultural and physical environment in which the interviews are to be conducted and, to some extent, on the content of the survey.

Some minimum requirements with respect to education are usually established; interviewers must be able to read and write with facility and to achieve a good understanding of the survey concepts, definitions and procedures. Ability to read maps and to make appropriate changes and notations is often important and can be tested. Important personality traits include self-confidence, poise, tact, resourcefulness and a good appearance. Where necessary, ability to work in the field under more or less primitive conditions should be taken into account. Under some circumstances, it may be desirable to exclude one sex or certain age groups from consideration.

Once the general criteria have been established, there are several techniques that can be used in the selection process. Lininger and Warwick (1975) mention application forms, personal interviews, psycological tests, and references. Application forms can serve a dual purpose: to provide the necessary background data and to provide a rough sample of the applicant's handwriting and ability to follow instructions. Personal interviews can be used to rate applicants on the desired personality traits. A simple intelligence test may be useful to eliminate clearly unqualified candidates; on the other hand, there seems to be no evidence that standard personality tests are of much value. With respect to references, a study by Hauck and Steincamp (1964) found that evaluations by outside referees of applicants' self-confidence and of their appearance, manner and poise were significantly related to interview completion rates. None of these techniques by itself is likely to do a good job of selection. But used together, they may at least be effective in identifying those applicants who are considerable more likely to do good work.

Applicants should be given a candid, unglamourized description of the kinds of work to be done and the expected working conditions in the field. Attrition, which can have serious adverse effects on quality, is likely to be lower if the persons selected enter the job with realistic expectations

Interviewer training is covered in Part One, Chapter V.C of the Handbook and will be discussed at greater length in a companion NHSCP study on training. Within a somewhat more specific context, the World Fertility Survey documents on training (1975b, 1975c and 1975d) provide

useful guidelines to survey organizers on planning and execution of training programmes for office and field survey staff. Some aspects of training that are especially important in the control of non-sampling errors include:

(a) Active participation by the trainees. The Handbook notes that:

> "Training is usually most effective when the trainees have to participate actively, through such means as oral question and answer periods, test exercises, mock and practice interviews and the like. There is probably nothing so tedious, and probably so ineffective, as straight lecturing. Audio-visual aids such as recordings, slides or motion pictures, etc., cannot only be useful as teaching devices but can serve to relieve the monotony. Careful preparation by the trainer is an absolute necessity or the whole process can degenerate into chaos."

In order to permit active participation, the size of classes must be kept small; fifteen might be taken as a maximum, and ten would probably be better.

(b) The use of standardized materials. There should be an interviewer's manual for every survey. If manuals for prior surveys contain standard materials on listing procedures, interviewing techniques and other features common to all surveys, it may be sufficient to prepare a supplement covering the new features of the current survey, especially the subject matter. If the training is to be conducted in several locations by different persons, a formal training guide is very desirable, to indicate in detail the topics to be covered and the practice exercises, quizzes, etc. to be included in each session.

(c) Balanced content. There is often a tendency for survey content -- the concepts and definitions -- to be over-emphasized at the expense of other aspects of the survey in an attempt to provide exhaustive coverage of all problems of interpretation that might be encountered by interviewers. It is probably more effective to concentrate discussions of content on the more common types of situations to insure that most interviewers will be able to handle those correctly. In addition, training should cover the objectives

of the survey and the procedures to be used for quality control of field work, so that interviewers will know what to expect and will be aware that the quality of their work is a matter of concern. If trainees have not been previously exposed to training in the use of effective interviewing techniques, this should be included. The fact that an interviewer has worked in several surveys does not necessarily mean that he or she is using effective techniques. A source of useful ideas about various aspects of interviewing -- introductions, methods of asking questions, use of probes, and methods of recording answers -- is the Interviewer's Manual of the University of Michigan Survey Research Center (1976).

(d) Evaluation of trainees. The effectiveness of the training and the extent to which individual trainees have mastered the essential elements should be evaluated by rating performance of the trainees on practice exercises, quizzes and examinations of various kinds. Those who are clearly incapable of doing the work in the field should be eliminated from consideration, reassigned or given additional training. Those whose performance in training is marginal should be identified to field supervisors so that they can receive special attention in the early stages of data collection.

J. Dress rehearsal

The Handbook does not make any distinction between a "pre-test" and a "dress rehearsal"; all such activities are covered in a short section on "pre-testing" (Part One, Chapter IV.C, pp. 70-71). Other works on survey development do distinguish various phases of testing. Jabine (1981) makes a distinction between "informal" and "formal" pre-tests. The former come first and are deemed essential in preparing for any new survey. The formal pre-test, just prior to the survey, is regarded as optional. If the survey itself is looked at as an experimental or pilot survey, a formal pre-test may not be necessary.

Simmons (1972, pp. 21-22) says that "a final dress rehearsal operation is advisable before entering into formal full production of the official set of survey data". He goes on to say that it "... should be a close

approximation to a trial or experimental conduct of the final survey itself, except carried out in miniature and over a limited period of time".

The principal ways in which a dress rehearsal can contribute to the control of non-sampling error are:

(i) Providing a basis for further improvement of survey instruments, manuals and training materials. Major revisions cannot be expected at this stage; however, some refinements may result from testing with personnel and procedures that closely approximate those to be used in the survey.

(ii) Identifying and correcting any weaknesses in the planned administrative and logistic features of the survey.

(iii) Providing a realistic body of data, i.e., a set of completed questionnaires, for use in testing data processing procedures. The specifications and instructions for editing, coding, data entry, computer editing and tabulation should be available in draft form by this time; for further development, each of these operations should be tested using a set of completed questionnaires, large and representative enough to identify any major gaps in the procedures.

If a dress rehearsal is carried out, its contribution to the control of non-sampling errors will depend on many of the same factors discussed in Section H of this chapter under "Pre-testing": clear specification of objectives, scheduling to permit use of findings, and development of a plan for capturing relevant information from the dress rehearsals.

K. Data collection

The organization and execution of the field work for a household survey are covered in detail in Part One, Chapter V of the Handbook. Throughout that chapter there are many references to procedures designed to minimize the effects of non-sampling errors. Some of these have already been discussed under "Interviewer Selection and Training" in Section I of this chapter. Those that apply directly to the data collection will be covered in this section.

The policies and procedures discussed in this section must be decided on before the actual start of field operations. The basic decisions on method of interviewer payment, production and quality standards and formal quality control systems are needed early so that these features of the field operations can be provided for in the survey forms, manuals and training materials.

1. Staffing and method of payment.

Field supervisors should play a major role in controlling the quality of the field work. An important factor affecting their ability to do so is the ratio of interviewers or other field staff to supervisors, i.e., the number of interviewers for whose work the average supervisor is responsible. This ratio should be kept small enough for the supervisor to be able to spend a significant amount of time, both in the field and in an office setting, observing, reviewing and checking the work of each of his or her interviewers. Other supervisory functions that are relevant to quality include "conversion" of refusals, training of replacement interviewers, and reporting to the central staff on problems not covered by instructions. The appropriate ratio in a particular survey will depend on such factors as the prior experience of the interviewing staff, the size of the assignment area, the nature of transportation and communication facilities available, and the proportion of time that supervisors may be required to spend on matters not directly related to the survey.

The two basic methods of payment to interviewers are fixed hourly or daily rate and "piece rate", that is, a fixed amount per completed interview or per sample unit assigned. Other things being equal, a piece-rate system tends to promote production at the expense of quality, and a fixed-rate system tends to do the reverse. Since both production and quality are important, it is necessary to supplement whatever basic payment scheme is adopted with a system of standards and controls for production and quality. Quality standards for interviewers should incorporate measures such as completion rates, edit failure rates and reinterview error rates. If supervisors are

diligent in monitoring the performance of each interviewer against such standards, and taking appropriate steps to reward good performance and correct poor performance, quality will be maintained regardless of which system of payment is adopted.

2. The field supervisor's role.

It is almost impossible to over-emphasize the importance of the field supervisor in controlling the quality of field work. If the work sent back to the central office for processing is of low quality, very little can be done at that stage to overcome the errors made in the field, short of sending the work back to the field -- a costly and time-consuming option.

Interviewers will take their cues primarily from their supervisors. No matter how much emphasis is given to quality in training sessions and in instruction manuals, if interviewers receive no feedback from supervisors on the quality of their work, they will soon adjust their priorities accordingly.

To help supervisors understand the importance of this role, quality control should be strongly emphasized in their training and in manuals or other written instructions covering their duties. All too frequently, the training of supervisors concentrates mostly on logistic and administrative matters -- receipt and shipment of materials, payment and allowances for interviewers, and the like. These subjects are all important, but do not deal directly with the supervisor's central responsibility, which is to see that the work is done on schedule and that standards of quality are met.

More specifically, the instructions to supervisors should cover three elements:

(i) A clear understanding of the kinds of quality-related problems that require communication with the central survey staff, and a well-defined procedure for obtaining answers to the questions that arise.

(ii) A comprehensive treatment of the procedures that will be used to control the quality of work done by interviewers, and of the supervisors' responsibilities for these procedures.

(iii) A description of the criteria which will be used by higher-level field staff or central staff to evaluate the performance of field supervisors.

3. Problem resolution procedures.

Field supervisors, especially in new surveys, frequently encounter unexpected problems in carrying out their work. Some examples are:

(i) A particular sampling area turns out to be inaccessible during the survey period due to floods or lack of cooperation from local officials.

(ii) An interviewer resigns or is ill, and there is no one immediately available to take on his or her assignment.

(iii) A sampling area turns out to have a much larger number of households than expected (this might turn up during a listing operation).

(iv) The map provided for an area sampling unit appears to be seriously inaccurate in some respect.

(v) The questionnaire and/or instruction manual do not make it clear how to deal with a particular situation encountered. For example, in a survey on use of health services, the local area has certain facilities or health personnel who do not fit into the standard categories provided for.

In all of these examples, failure to adopt the best practicable solution may have an adverse effect on quality. Several of the problems listed require, for their solution, technical understanding of sampling or subject matter issues that may be available only from a regional office or from the central survey staff.

To avoid ad hoc solutions by unqualified persons, an effective system for resolution of these problems must be established before the field work begins. The important attributes of the system are: a clear understanding, by the field supervisors, of the kinds of problems that must be referred; an effective system for communicating with the regional or central office; a procedure for channelling each problem to the person who is best qualified to devise a solution; quick "turnaround" in responding to the field; and a procedure for quickly notifying all field staff of any resulting procedural changes that are of general applicability.

4. Formal quality control procedures.

Several methods of quality control of field work are described at length in the Handbook (Part One, Chapter V.D). These might be regarded as falling into three major categories: observation of interviews (including listening to tape recording of interviews), review and editing of completed questionnaires (including preliminary hand tallies), and reinterviews for a sub-sample of assigned units.

Each of these three methods serves different purposes, and it is recommended that all be used in every survey, if resources permit. The nature and scope of their application will depend on factors such as the experience of the field staff and the novelty and complexity of the topics covered in the survey.

The amount of observation of interviews and, to some extent, reviews and edits of completed questionnaires should be greatest for new interviewers and for surveys covering new topics. However, it should not be assumed that experienced interviewers always maintain a high level of performance. In a study at the US Bureau of the Census (Stevens and Bailar, 1976), interviewers' performance was measured by the rates of non-response and probing. (The latter rate refers to the recording of answers without actually asking the necessary questions: for example, in a survey on income, some interviewers may obtain the annual income simply by multiplying the income for the first month by twelve or for the first quarter by four, rather than by obtaining and summing up income for individual months or quarters of the year). The performance was evaluated against the interviewers' experience, defined as the number of times they had participated in the income survey. It was found that interviewers who worked for the survey for the first time had the highest response rates, while the length of experience did not seem to be correlated with non-probing. In another study, the U.S. Census Bureau (Rustemeyer, 1977) used mock interviews to study the performance of interviewers at different levels of experience. It was found that the new interviewers were more careful in introducing and closing the interview, while the more experienced interviewers

performed better in recording and transcribing the required information and notes. Findings of this kind suggest that the work of experienced interviewers should continue to be observed and reviewed on a sample basis.

For each method of quality control used, specific forms and instructions must be developed. Thus, for supervisory observation of interviews, there should probably be a form to be completed by the supervisor for each interview observed. This form would serve as a checklist for supervisors to remind them of all of the aspects of the interview that need to be evaluated. It will create a record that will allow the supervisor to compare the work of different interviewers more objectively and to monitor the performance of individual interviewers over time.

The instructions to the supervisor for the observation programme should cover the number and scheduling of the observed interviews for each interviewer, what specific things to watch for during interviews, how to use the observation form, what to do with the form following the interview, how to use the results of the observation to help interviewers improve their work, and other pertinent aspects of the activity.

Similar forms and instructions are needed for review/edit and for reinterview activities. Preparing these materials calls for considerable work by the survey planning staff, but it is the only way to develop effective quality control systems. Merely giving supervisors a vague instruction to pay attention to quality is not enough: it is necessary to tell them in detail how to do it, how much to do, and how to use the findings from the various monitoring techniques.

The review and edit of completed questionnaires is especially important at the start of the data collection period, so that any systematic errors resulting from misunderstanding of instructions or faulty instructions can be corrected early. Careful scheduling of interviewer assignments will be needed to assure that they get feedback from early reviews before doing large numbers of additional interviews.

As field supervisors receive completed questionnaires and review all or a sample of them, they should routinely record the outcome for each sample unit assigned as eligible-complete, eligible-incomplete (by reason), or ineligible (by reason). From time to time, completion rates (see Chapter IV, A.2) should be calculated for each interviewer, and corrective action taken where the rates are below standard.

Reinterviews of a sample of units can be used both for quality control and for the evaluation of survey results. What they measure will depend on the specific procedure used. If the reinterview is fully independent of the original interview, without reconciliation of differences, the results will be primarily an indication of the level of response variance, including the interviewer component. If the differences are reconciled to determine which of differing responses is correct (occasionally a third response will prove to be the correct one), the results can be used to estimate response bias, of which a part, but not necessarily all, can be attributed to poor interview performance. Thus, interviews with reconciliation are probably a more suitable basis for quality control.

The cost of reinterviews is high and the time required to conduct them and to process the results makes them generally unsuitable for a quick, early measure of interviewer performance. Consequently, they cannot be recommended for quality control purposes in a one-time survey; their use in this way is generally restricted to continuous or periodic surveys. Examples are given for labour force surveys by the U.S. Bureau of the Census (1963, 1968a and b) and for health surveys by the U.S. National Center for Health Statistics (1973b).

A special form of check, which could be undertaken separately or done at the same time as reinterviews with households for which interviews were completed, is the "vacancy check". There may be a temptation for some interviewers to classify some occupied sample housing units for which interviews are inconvenient or difficult to obtain as "vacant", thereby eliminating the requirement to obtain interviews for those units. This practice can readily be

controlled by checking a few of the units classified as vacant to determine whether they were in fact vacant when the interviewer was there.

For any of these quality control methods, the objectives and procedures must be fully explained in advance to the interviewers, so they know what to expect. Feedback on errors identified should be presented in a constructive way: the intention should not be to criticize, but to work together to find ways of avoiding such errors in future interviews. The whole idea of quality control will be much more acceptable if the system includes rewards for good performance.

5. Evaluation of supervisors' performance.

Because of the key role of field supervisors in maintaining the quality of field work, their performance, as well as that of their interviewers, should be evaluated. Their performance in training sessions prior to a survey can and should be evaluated. They should be fully trained in the operations they will be supervising and should demonstrate understanding of their own special functions.

Under most circumstances, there will not be much chance to evaluate the work of supervisors during a survey in time to have any significant effect on the quality of field work in that survey. The evaluation, which consists in large part of aggregating information on the performance of interviewers working under them, will, however, be very useful when the supervisors are part of permanent field staff that will be working on future surveys. Therefore, it will be discussed in the next chapter, which deals with long-range strategies for improvement.

6. Capturing information for purposes of survey assessment and long-term improvement.

This topic is mentioned here only as a reminder, it will be discussed in Chapter VII. In a certain sense, each survey can be thought of as a "pre-test" for future surveys. Therefore, the survey planning staff should insist that provisions be made to collect relevant information on costs, error components and other relevant design variables during the data collection phase and indeed during

all phases of a survey. With respect to data collection, this implies several techniques that have already been discussed under "Pre-testing" in Section H of this chapter, such as having professional staff observe the field work, using the questionnaire to record information about call-backs and duration of interview, holding debriefing sessions for interviewers and supervisors, tape recording of a limited number of interviews, and asking field staff to prepare narrative reports describing any difficulties encountered during the data collection and offering suggestions for improvements in questionnaires, control forms and procedures.

L. Data Processing and Analysis

Procedures for the assessment and control of errors in the data processing and analysis phases are not discussed in this document. Nevertheless, for completeness, quality control measures following the data collection phase have been included in the checklist given below. A separate study in the present series, namely Survey Data Processing: A Review of Issues and Procedures (United Nations, 1982) provides an overview of the data preparation and processing task in the context of programmes for continuing collection of statistical data. More specifically, that study addresses in detail the various technical and operational problems of organization and implementation of data processing activities for household survey programmes undertaken by national statistical agencies in developing countries.

M. CHECKLIST FOR CONTROL OF NON-SAMPLING ERRORS IN A HOUSEHOLD SURVEY

SURVEY ACTIVITY	QUESTIONS	REMARKS
B.* General planning	a. Have the proposed topics been investigated in prior surveys: (1) By this organization? (2) By other statistical organizations?	If "no" to a(1), more time will be needed for planning and preparation. If "yes" to a(2), try to benefit from their experience.
C. Selection of topics and items to be included.	a. Has it been demonstrated that the desired items can be obtained with reasonable accuracy from household survey respondents?	If "no" for some items, testing will be necessary.
	b. Consider the probable length and complexity of the questionnaire that will be needed. Is it likely to:	Type of resulting error:
	(1) Lead to low response rates?	Non-response error.
	(2) Be too much for available interviewers to handle?	Response error.
	(3) Overtax facilities available for data-processing and analysis?	Data-processing error.
	c. For what reference periods are data to be obtained?	For longer periods a multi-round survey may be necessary.
	d. Are the key concepts clearly defined?	If "no",will be impossible to measure or estimate some components of non-sampling error.
	e. To what extent is retention of items used in prior surveys desirable?	There may be a conflict between improvement of quality and the desire for data that are comparable over time, as well as data-processing convenience.
	f. Has at least a rough outline of the survey tabulations been prepared?	This is a good way to avoid leaving out items needed to fulfill the purposes of the survey.

*Item numbers B-L refer to the topics discussed in the corresponding sections of this chapter.

SURVEY ACTIVITY	QUESTIONS	REMARKS
D Survey design and structure	a. What are the units of observation?	Units consisting of groups of persons should be clearly defined. Review international recommendations.
	b. Should any population groups be deliberately excluded from the target population?	Consider the effects of exclusion on the relevance of survey results.
	c. What methods of data collection will be used?	For some topics direct observation or measurement may be the only way to get sufficiently accurate results.
	d. Over how long a period will the interviewing be done? How many interviewers will be needed?	These decisions can significantly affect the level of response error, especially interviewer variance and bias.
	e. How is the scheduling of interviews likely to affect the levels of non-sampling error?	The timing of interviews may affect the ability to contact respondents in some areas, and the ability of respondents to answer questions accurately.
E. Sample design and selection.	a. Will the sample design incorporate interpenetrating samples?	Comparison of results among sub-samples is useful to spot some kinds of non-sampling error.
	b. What kind of area frame will be used? What is the quality of maps available?	Be sure that up-to-date information on political sub-divisions is available. Develop quality control procedures for all field and office operations to prepare and/or update maps.
	c. What kinds of list frames will be used?	If possible avoid use of village lists. Check the adequacy of forms, instructions, training and quality control procedures for field listing operations.
	d. Are the ultimate sampling units clearly defined, as well as the rules of association that link units of observation with them?	Make early decisions on dwelling unit vs. household and de jure vs. de facto coverage.

SURVEY ACTIVITY	QUESTIONS	REMARKS
	e. Will part of the sample selection process be carried out in the field?	Provide for full documentation of the selection process and review of the results. Retain documents for later use.
	f. Does the operational plan include checks on the sample selection process?	At each stage compare actual and expected sample counts. Compare estimates based on sample counts with known totals.
	g. Will samples for subsequent surveys overlap the sample for this survey?	If listings are used, some procedure for updating will be needed. Be aware of possible effects of panel bias, and plan to measure these effects.
F. Data collection: two strategic decisions.	a. What respondent rules will be used?	For some topics, information from proxy or household respondents will be less accurate.
	b. What follow-up procedures will be used?	Interviewers should have clear instructions. Follow-ups reduce non-response, but increase costs.
G. Development of questionnaires.	a. Is enough known about the topics to start design of structured questionnaires?	If the topic is new, try unstructured individual and/or group interviews first.
	b. How many different questionnaires are needed?	Survey may require regional or language versions, or supplemental questionnaires for persons with specified characteristics, etc.
	c. Should a verbatim question or a "list" approach be used?	There has been little research to guide this decision. Sometimes both approaches can be used on the same questionnaire.
	d. How will the questionnaires be processed?	Review needs with those responsible for coding operation (if any) and data entry.

SURVEY ACTIVITY	QUESTIONS	REMARKS
	e. Has the draft questionnaire been carefully reviewed for: (1) complete identification section? (2) appropriate groupings of items (modules)? (3) smooth transitions from one module to the next? (4) skip patterns clearly indicated, and as simple as possible? (5) method of recording answers clearly and adequate space provided? (6) clear distinction between "zero" or "non" as opposed to "don't know" or non-response? (7) suitable provisions for coding and data entry? (8) questions that are vague or unclear? (9) questions that may be too difficult for many respondents to answer? (10) questions incorporating more than one idea? (11) sensitive questions that are likely to cause problems? (12) multiple-response questions for which the response categories are not mutually exclusive and exhaustive?	Questionnaires should be reviewed by several people after each revision. Check each of the items listed.
	f. Are there alternatives that should be tested in a pretest or in the survey itself?	The size of the pre-test may not be sufficient for this purpose.
	g. Has a plan been developed for quality control of the final printed questionnaires?	Check for systematic printing errors, such as blank or omitted pages in some batches.
H. Pretesting	a. Are the main objectives of pretest(s) clearly specified?	Possible objectives include: - refining questionnaires and procedures - estimating cost factors - comparing alternate approaches.

SURVEY ACTIVITY	QUESTIONS	REMARKS
	b. Are pretest sample size and design adequate for the main objective?	Consider both the number of respondents and the number of interviewers participating.
	c. Does the scheduling of the pretest(s) allow enough time to analyze and make proper use of the findings?	Some results are needed before the date the final questionnaire must go to the printer.
	d. What methods will be used to capture relevant information from the pretest(s)?	Consider: - observers - time recording procedures - interviewer and respondent "debriefings" - tape recording of interviews and debriefings - written reports by interviewers and supervisors - formal tabulation and analysis of data.
I. Interviewer selection and training.	a. Will survey be done by existing interviewing staff or will it be necessary to select and train new interviewers?	If some new interviewers will be hired, establish selection criteria. Prior experience will influence amount and content of training needed.
	b. Are there any special requirements related to the content of the survey and the target population?	Consider travel requirements for field work, possible needs to match interviewers and respondents with respect to sex, language spoken etc.
	c. What are the logistics for training supervisors and interviewers who will do the training, how many sessions, how long and at what locations?	Training in small groups, with active participation by trainees, is much more effective.
	d. What should the training sessions cover?	Basic items include: - objectives of survey - content and concepts - interviewing techniques - practice interviews - quality control.

SURVEY ACTIVITY	QUESTIONS	REMARKS
	e. Have the following documents been prepared: (1) Supervisor's manual?	Should place major emphasis on quality control functions.
	(2) Interviewer's manual? (3) Training guide?	Especially important if several different people will be training interviewers. For example, practice exercises for training.
	(4) Other? f. What criteria will be used to decide which of the potential interviewers have successfully completed training?	Persons who are clearly unfit for interviewing should not be given field assignments.
J. Dress rehearsal.	a. Is a formal dress rehearsal necessary? What are its objectives?	For a continuing survey, it may be possible to treat the first survey period (say 3 months) as a dress rehearsal.
K. Data collection	a. What will be the ratio of field supervisors to interviewers?	It should be high enough to allow the supervisor to spend time in the field observing and reviewing the work of each interviewer.
	b. What system of interviewer payment has been adopted? Is it supplemented by appropriate standards of production and quality?	
	c. What kind of training programme has been established for supervisors?	Supervisors' training should strongly emphasize their role in quality control of field work.
	d. Is there a formal procedure for problem resolution and quick transmittal of new instructions to the field?	In spite of best previous efforts, the questionnaire and procedures may still have some "bugs".
	e. Have plans been developed for quality control systems in each of the following categories: (1) Supervisory observation and review of interviewers' work in field?	Each quality control system requires: - clearly defined objectives - appropriate form or other means of capturing performance data

SURVEY ACTIVITY	QUESTIONS	REMARKS
	(2) Field office review/edit of completed questionnaires? (3) Reinterview of a sub-sample of units?	- standards of acceptable performance - procedures for corrective action when work is not acceptable.
	f. Does the central office have a system for evaluating supervisory performance?	Objective criteria should be used as much as possible.
	g. What can be done during the data collection phase to improve performance in subsequent surveys?	Examples are: - debriefing sessions of central office staff with interviewers and/or supervisors - narrative reports or questionnaires to be filled by interviewers and/or supervisors - solicitation by central office of field suggestions for improvement of quality - "built-in" experiments with alternate collection procedures.
L. Data Processing and Analysis: 1. Receipt and control of questionnaires for data processing.	a. Have materials been received from each field office?	
	b. Are all assigned sample households or area units accounted for?	
	c. Do counts of sample households agree (within tolerance) with preliminary estimates?	Follow up with field for PSU's whose counts are outside pre-established tolerance limits.
	d. Does each questionnaire have correct and unique identification?	It may be desirable to assign separate batch and serial numbers for data processing.
	e. Does the batch control form have provision for: (1) Batch control counts? (2) Recording desired information for each manual processing step?	As a minimum, identify operator and show start and completion date for each step.

SURVEY ACTIVITY	QUESTIONS	REMARKS
2. Manual edit	a. Has a clear set of instructions been developed and pretested?	
	b. Has a suitable form been developed for recording errors found by editors?	Data on the edit form will be used for quality control (QC) of edit operation and for feedback to field.
	c. Will it be possible to distinguish editors' entries on questionnaires from those made by interviewers?	Editors should use a pen or pencil of a distinct color, and should never erase interviewers' entries.
	d. Is there a formal procedure for problem resolution and quick transmittal of new instructions to editors?	
	e. Are edit clerks qualified by experience and specific training for this operation?	
	f. Has a plan been developed for verification and quality control of editors' work?	See item K.e for essential elements of a QC system.
3. Manual coding	a. How many separate coding operations are needed?	For complex items such as occupation and industry, a separate operation may lead to better quality.
	b. Has a clear set of instructions been developed and pretested?	
	c. Can special reference aids be made available to coders?	For example, a list of industry codes used in establishment surveys for the large employers in each locality.
	d. Is there a formal procedure for problem resolution and quick transmittal of new instructions to coders?	
	e. What training do coders need before starting on production?	

SURVEY ACTIVITY	QUESTIONS	REMARKS
	f. Has a plan been developed for verification and quality control of coders' work?	See item K.e for essential elements of a QC system. An independent verification scheme is preferable.
4. Data entry	a. Are edit facilities of key-entry equipment being used to full advantage?	
	b. Have keying and associated edit software been adequately pretested?	Tests should cover suitability of output for subsequent processing steps.
	c. Have provisions been made for periodic checks of data entry equipment and software?	
	d. Has a plan been developed for verification and quality control of data entry clerks' work?	
5. Computer processing: pre-tabulation phase	a. Has a complete file of records been generated?	Check record counts against control counts from prior stages.
	b. Do procedures call for retention of back-up tapes at each stage?	
	c. Has a fixed routine, using manufacturer's test decks, been established for computer maintenance?	
	d. Has the overall computer edit and correction/imputation system been adequately tested?	A sample of records should be examined by analysts on a before and after basis to evaluate the results of each processing step.
	e. Do edit runs include the production of "diaries" showing frequency and types of errors found, and disposition?	Diary outputs are a valuable source of information for identifying questionnaire deficiencies and weak spots in earlier collection and processing operations.
	f. Are corrected tapes re-edited until no further errors are detected?	

SURVEY ACTIVITY	QUESTIONS	REMARKS
6. Estimation and tabulation	a. Has a procedure been developed and tested for imputation of data for missing cases and items remaining after edits?	Generally, imputation of missing data should be kept to a minimum.
	b. Have weighted population and household counts been checked against independent estimates?	Large differences may indicate improper application of weighting procedures.
	c. What kinds of checks are planned for output tables?	Tables should be reviewed for: - internal consistency - inter-table consistency (agreement of comparable totals) - significant changes from prior data - "reasonableness" based on expectations and data from other sources.
	d. Have sampling errors for major variables been estimated and reviewed?	Sampling errors determine the limits to which survey data may be reasonably disaggregated for tabulation and reporting.
7. Dissemination	a. Have all transcriptions from computer output to final publication copy been verified 100 per cent?	
	b. Have all calculations based on the computer output been verified?	
	c. In addition to basic tables and text, will the publication contain other elements needed by users for their analysis and understanding of the data?	The publication should include: - a description of the survey design - explanation of concepts and definitions used - information about sampling errors and available data on non-sampling errors.

SURVEY ACTIVITY	QUESTIONS	REMARKS
	d. Has the descriptive and analytical text been reviewed to insure that conclusions presented are justified by the basic data, taking all sources of error into account?	
	e. Have all sections of the report been adequately proofed and reviewed by subject-matter analysts?	

CHAPTER VII

ASSESSMENT OF ERRORS AND LONG-RANGE STRATEGIES FOR IMPROVEMENT OF QUALITY

A. Introduction

Chapter VI covered in some detail measures and procedures for the control of non-sampling errors during the various stages of preparation and execution of household surveys. While these activities are necessary for any survey of high quality, their potential usefulness is greatly enhanced when surveys are planned and executed within the framework of an integrated programme. Obviously, the success of such efforts will depend to a considerable degree on what has been learned in earlier surveys about the effectiveness and cost of specific survey procedures. Hence, it is not only useful but also essential that, in a continuing programme of household surveys, information needed for the long-range improvement of data quality is systematically gathered and documented during each survey. Building up data quality is a long-term process requiring an understanding of the nature and sources of errors in surveys and an assessment of their magnitude and effect on survey results. A more rational allocation of resources, especially the identification of major inefficiencies in design and procedures, require quantitative information on at least the main components of survey errors.

This chapter provides an overview of the techniques and procedures which may be used to capture information relevant to the quality of survey data and improvement of survey design, and outlines some long-term strategies for improving the quality of the data. Such information is necessary, for example, for:

(i) survey management and its operational and quality control;

(ii) assessing the quality of the data and cautioning users of the limitations of the information produced; and

(iii) obtaining qualitative and quantitative information on the basis of which procedures and design of future surveys can be improved.

The costs of monitoring and evaluation have to be seen in the context of their contribution to the enhancement of quality and usefulness of the substantive information produced by the surveys, in accordance with the general principles of "total survey design" sketched in Chapter I. Their significance is greatly enhanced in the context of a continuing programme of surveys, where the integration of the data produced from various surveys, greater responsiveness to user needs and policy relevance, and gradual improvement of survey facilities and procedures are fundamental requirements. At the same time, continuity in survey activities tends to make continued survey monitoring and evaluation not only more necessary but also more feasible and cost-effective. For example, operational and quality control measures can be implemented more thoroughly and more economically given the permanent and regularized arrangements and procedures in a continuing programme. Similarly, evaluation studies requiring special fieldwork can be more easily accommodated -- both organizationally and financially -- within the primary survey operations.

B. Accumulation of information on data quality

There are many opportunities for the accumulation of information about non-sampling errors during a survey programme. Firstly, a great deal of useful information can be obtained as a by-product of routine survey management, monitoring and control. Since basically these techniques of obtaining information on data quality require no addition or modification to the operations necessary for management in any case, these may be termed "informal techniques". Secondly, more quantitative information may be obtained through special investigations involving additional data collection, or modification of the design of the primary survey, or both. These are aimed at obtaining more quantitative assessment of the nature and magnitude of survey errors, and may be termed "formal techniques". Finally, quality of survey results at the aggregate level can be investigated by using analytical methods, involving internal validation on the basis of substantive considerations, as well as comparison with external sources including data from other related surveys in the programme. Each of these groups of techniques is discussed in turn in the following sub-sections.

1. Informal techniques

Much can be learned about non-sampling errors through summarization and analysis of the survey questionnaires and responses, as well as of records developed for operational purposes during data collection and processing. For this, plans must be made to create and preserve, in easily accessible form, the appropriate records and documents, and to summarize and analyze the information they contain. These records can provide very detailed information on sources and nature of errors as well as on possible means for their control. At the same time, financial data (or its surrogate in the form of time spent on various operations) can provide a quantitative basis for more efficient and cost-effective design of future surveys.

Perhaps first priority should be given to the analysis of outcomes of the data collection operation. Using records for the sample of housing units or other units initially assigned for field work, tabulations should be prepared showing the numbers of units in each of the following categories:

Eligible for interview;

Interview obtained;

Interview not obtained (by reason);

Ineligible for interview (by reason);

Eligibility not determined (by reason).

Where possible, the data should be cross-classified by geographic area, supervisor, type of area (e.g. urban or rural), and other relevant variables. Appropriate rates, such as completion rates or refusal rates can be calculated and compared for the different sub-groups.

Analysis of the data can suggest various possibilities for improvement. If completion rates are low overall, it may, for example, be necessary to consider changing the call-back rules, or to institute checks to see that existing rules are being followed. If completion rates are low in some areas, this may indicate a need for better supervision in those areas, or it may suggest that greater efforts are needed to inform local officials and the public about the survey. If a significant number of units are classified as "unable to locate", this may suggest that better maps are needed, or that the listing operation needs to be done more carefully.

Scrutiny of detailed data on "sample outcomes" does not always make

it clear how to solve problems, but it is an excellent way to find out where the problems are in the areas of non-response and, to some extent, coverage.

Other kinds of analyses based on questionnaires and/or operational records might include:

(a) Analysis of item non-response. These could be done from final tabulations, from a sample of the original questionnaires, or at some intermediate stage, depending on the kinds of imputation procedures used in the survey and on the purpose of the analysis. High non-response rates for particular items can occur for various reasons. An item may be too difficult for some respondents to answer, it may be highly sensitive, or it may be overlooked frequently by interviewers as a result of faulty questionnaire structure or inadequate training. Like completion rates, item non-response rates help to identify problems; further study may be needed to suggest solutions.

(b) Analysis of edit reports. Ways to improve questionnaires and interviewer training may be suggested by statistical analysis of interviewer errors discovered in questionnaire edits or reviews performed by field supervisors or by central office staff prior to data entry. The analysis will be most effective if:

(i) the editors or other reviewers have been given clear instructions on what items to check and what kinds of errors to look for, and

(ii) the results of the edits or reviews have been recorded on a structured form.

The summarization of the edit records may be done centrally, using all or a sample of the forms on which the errors have been recorded. Alternatively, field supervisors could be asked to prepare and submit summary reports. The latter approach can be helpful to the supervisors if they summarize the data by interviewer and compare their error rates.

(c) Analysis of quality control records from coding and data entry operations. Quality control systems require the recording of errors found in various operations in order to determine corrective actions, such as the rejection of a particular batch of work or increasing the inspection level for a particular coder or keyer. It may be useful to analyze the same records, summarized for the entire operation, to see what particular questionnaire items are causing the most difficulty, and what kinds of errors are being made. This analysis may suggest future improvements in questionnaire design and in instruction materials and training for coders and data entry clerks.

(d) Analysis of computer edit and consistency check results. Computer "cleaning" of raw data files proceeds in two ways: through the "rejection" and manual correction of records that fail certain tests, such as range or consistency checks, and (less frequently) through automated correction of appropriate fields in records failing the same kinds of tests. In either case, the processing system should generate summary records showing the frequency of failure and, if appropriate, the distribution by reason for failure for each test performed. These records should be reviewed to identify the most common edit failures and to try to determine why they occurred and how they can be reduced in future surveys. Some of the problems identified may be corrected by revising data collection instruments and procedures. Others may suggest changes in operations subsequent to data collection, including specific consistency checks in the instructions for field edits or widening the acceptable limits for a test that is part of the computer edit.

Insofar as they are not part of the quality control system for the current survey, these analyses do not necessarily have to be carried out while the survey is under way. If they are to be performed later, however, arrangements must be made to ensure that the necessary records will be available when needed. With respect to questionnaires, even if no specific studies have been planned, it would probably be useful to select and retain for at least five years a small, well-defined sample of the questionnaires

from each survey for possible future analyses.

There are also several extremely valuable ways of obtaining qualitative information about non-sampling errors during the data collection. These techniques were mentioned in Chapter VI in connexion with pre-tests, where the findings could affect the quality of the data collected in the main survey. These techniques are relatively inexpensive, and should also be used in the main survey as much as possible to obtain information that can be used in planning future surveys. They include:

(i) Field observation and reports by members of the survey professional staff.

(ii) Inclusion of one or two items at the end of the questionnaire for interviewers to note particular problems that occurred in the interview.

(iii) Debriefing sessions for interviewers and field supervisors. Unlike those in the pretest, these sessions would be held on a selective basis, rather than for all persons participating in the data collection.

(iv) Tape recording of a small number of interviews.

(v) Final reports by interviewers and field supervisors, with special emphasis on identifying questionnaire items that had frequent problems and describing those problems.

2. Formal evaluation techniques

Formal evaluation studies attempt to estimate the magnitude of selected components of non-sampling error. Three groups of techniques are described below:

(i) Modification of the design of the primary survey operations so that comparisons between different sub-sets within the survey yield estimates of various components of error or isolate the effect of different procedures on survey results. Frequently, these techniques are based on the use of interpenetrating sub-samples.

(ii) Re-interview of a sub-sample of the respondents, followed by a reconciliation of differences between the original and the re-interview so as to identify sources and nature

of errors at the micro (individual respondent) level. The re-interview may be designed to be a repetition of the original interview conducted under the same essential conditions, in which case it is aimed at measuring the variable component of the error; alternatively, the re-interview may take the form of an in-depth interview, aimed at obtaining more accurate data, and hence an assessment of bias in the original interview.

(iii) Micro-level matching and comparison with external independent sources of data, so as to identify coverage and content errors in the survey.

These formal techniques, if used effectively, can provide information that is more directly applicable in the design of future surveys. At the same time, they may call for substantial additional resources and place significant demands on the technical staff of the survey organization. However, for reasons indicated earlier, the undertaking of special, formal investigations of the type described in this section is likely to be more within the reach of available resources and facilities of an organization undertaking a continuing survey programme.

(a) The use of interpenetrating samples

The idea of interpenetrating samples was introduced in Chapter VI along with some discussion of different uses for the technique and different levels at which it can be applied. There are two basic ways in which interpenetrating samples can be used for the assessment of non-sampling errors. One is to provide quantitative estimates of variable components of non-sampling error, using the same survey procedures in each of the two or more interpenetrating sub-samples. The other is to compare the effects of alternate survey procedures, for example, two or more different versions of a questionnaire, on non-sampling error. In the latter case, the study is often referrred to as a "split-panel" test or experiment.

When the technique is used to estimate variable components of error, using a fixed set of survey procedures, the data collection and processing operations for the sub-samples can be designed to

different degrees of symmetry and completeness. A very complete design might call for the random allocation of equal numbers of respondents to each interviewer, with the work programme of supervisors, editors and coders who are also potential sources of response variation arranged in an "orthogonal design" so as to permit separate estimation of correlated errors from each source. At the other extreme, the state of completeness of "randomization" may consist simply of contrasting results of different interviewers without actually ensuring randomized allocation. In practical situations, complete randomization is seldom feasible organizationally, and random allocation to interviewers also increases travel costs. Nevertheless, even a reasonable degree of randomization in allocation can provide valuable quantitative information on variable components of error (Kish, 1962). However, insofar as the procedures used in the various independent parallel sub-samples are the same, comparison of their results does not provide any indication of the biases in the data. Bias can be measured only through comparison with some other more accurate source of data.

The use of this technique to estimate variable components of non-sampling error in household surveys has been rather limited; some examples were given in Section D.3 of Chapter V. Studies in connection with household surveys have all related to interviewer variance; a review of the literature has not turned up any efforts to estimate correlated response variance due to supervisors, editors or coders.

On the other hand, there are many examples of the use of inter-penetrating samples for methodological experiments built into household surveys. Some advantages and disadvantages of such experiments are (Jabine, 1981):

> "The major advantage of built-in experiments is that they provide relatively large samples at low marginal cost. Samples of 30 to 50 cases and 3 interviewers per treatment, as recommended for pretests in the most recent draft of the UN Handbook of Household Surveys, Part I,

p. 61) may be enough to provide some rough qualitative information where treatments differ substantially, but are not nearly large enough to detect important but relatively small differences. Imaginative use of experiments built into ongoing surveys is probably the only way for an organization with limited resources to conduct useful quantitative methodological research.

There are also some disadvantages to built-in experiments. The Survey staff should be prepared to deal with the situation where survey results turn out to differ substantially by treatment. Should the data for all treatments be pooled, or should the data for inferior treatments, judged by whatever criteria have been established, be discarded? Obviously, an experiment built into a survey is not the place to test a radically different procedure about which very little is known. In general, substantial technical resources will be needed if such experimentation is to succeed in its objectives. Members of the survey staff who are more results-oriented may resent the additional complications introduced into what is already a complicated undertaking. In summary, while it will be tempting to the methodologically-oriented statistician to build as much research as possible into the pilot survey, such efforts should not be allowed to compromise the primary objective, which is to produce usable data on the topics selected."

Like any scientific experiment, a methodological test built into a survey must be carefully planned. Major elements of the plan include:

(i) Definition of the study population. This will normally be the target population for the survey, or some subset of it.

(ii) Definition of the "treatments", that is, the alternatives to be tested. These could cover almost any aspect of the survey, such as alternate methods of training interviewers alternate versions of the questionnaire, alternate respondent rules, or alternate procedures for use in manual processing operations.

(iii) Sample design and allocation of treatments to the sample. There must be some element of randomization in the allocation of treatments to respondents and to interviewers (or to clerks, if the alternatives to be tested are part of the processing phase of the survey) The allocation of

treatments must be carefully coordinated with the sample design for the survey.

(iv) Specification of evaluation or decision criteria. The variables to be observed and analysed in order to evaluate and make a choice among alternative treatments should be specified in advance. As noted by Jabine (1981):

> "For most experiments of the type discussed here, these variables will relate to costs and to nonsampling error. It should be possible to judge in advance whether different treatments are likely to have substantially different costs. For example, a minor variation in wording of a particular question is not likely to affect interviewing costs, but a difference in respondent rules could have a substantial effect. If costs are important, then record-keeping procedures to provide data on cost by treatment must be established. Nearly all design features and variations thereof are likely to affect the level of nonsampling error, and plans should be made to capture relevant information. Some useful data can be obtained by calculating various kinds of operational error measures separately for each treatment. Thus, information on completion rates, item non-response rates, errors found in manual and computer edits, etc. by treatment will be useful. More direct measures of nonsampling error may require the introduction of special procedures, such as conducting reinterviews for a sample of households or comparing the survey responses with information obtained from participant observer studies, or from administrative records. For some items, the survey results themselves, tabulated by treatment, may provide an indication of differences between treatments in the completeness of reporting. For example, if it is known from other studies that certain kinds of income are likely to be under-reported, then the treatment which produces the larger estimated amounts of income will be preferred."

Like the planning, the execution and analysis of built-in experiments require careful attention. The plan for allocation of treatments to interviewers or clerks must be scrupulously followed. The statistician who designed the

test should be consulted whenever circumstances such as resignation or illness of interviewers or loss of primary or secondary units from the sample require adjustments of the randomization procedure. An essential part of the analysis is the estimation of sampling errors, so that confidence intervals can be constructed for the estimates of differences between treatments.

Several examples of built-in experiments have been discussed in earlier chapters. Jabine and Rothwell (1970) presented findings from "split-panel" tests of questionnaires conducted by the U.S. Census Bureau, several of which were built into household surveys. Neter and Waksberg (1964) describe a complex multi-factor experiment that was built into a national panel survey of expenditures for home alterations and repairs. Among the features tested were recall periods of varying lengths, with and without bounding by data from prior interviews, and alternate rules for the choice of respondents. Some of the results were presented in Section C.3 of Chapter V.

(b) Reinterview studies

The use of reinterview studies for quality control in continuing surveys was discussed in Chapter VI; their use for evaluation will be discussed here. Like other evaluation studies, reinterview studies serve both to inform users of some of the limitations of results from the survey they relate to and to provide useful data for the design of future surveys.

In reinterview studies, a sub-sample of households or individuals included in the survey are interviewed for a second time, by a different interviewer, shortly after the initial interview. There are two basic types of reinterview. The first type, whose

primary object is to measure response variance, is a replication of the initial interview, using the same questions although usually only for selected items. The second type, which attempts to measure response bias, depends on the use of specially trained or qualified interviewers and a series of special probes for the items included.

For the replication approach, the reinterviews should be conducted independently of the initial interviews, that is the reinterviewers should not have access to the original answers before or during their reinterviews. The pairs of answers for individuals and households may then be treated as "independent" observations of the same variables, obtained under the same set of general survey conditions, and hence can be used to estimate response variance for these items. Actually, of course, they are not completely independent observations, because reinterview respondents may remember what they told the initial interviewer. In general, lack of independence should lead to underestimation of response variance.

Frequently, when the replication approach to reinterviews is used, the original interview and reinterview responses are reconciled, that is, they are compared, and where there are differences a determination is made, by asking additional questions as needed, as to which response is correct. The response obtained from the reconciliation process will sometimes differ from both the initial interview and reinterview responses. The data obtained from reconciliation, that is, the differences between the reconciliation values and initial interview (or reinterview) values may be used to estimate response bias for the variables investigated.

Reconciliation can be done by the reinterviewer, immediately on completion of the reinterview, while still in the sample household; however, the reinterviewer must be under strict instructions:

(i) not to look at the responses from the initial interview until the reinterview is complete, and

(ii) not to change any of the initial interview or reinterview entries as a result of the reconciliation.

Results of reconciliation should be recorded in a separate section of the reinterview form. Even with these precautions, experience has shown that estimates of response variance and related statistics such as the index of inconsistency, are usually lower when reconciliation of the differences is undertaken at the same time as the reinterview, than when it is not.

The objective of the probing type reinterview is to obtain the most accurate information possible for each of the variables selected for investigation. The differences between the reinterview and initial interview responses then provide a basis for estimating response bias. The variety of probing questions that can be used for this purpose is limited only by the ingenuity of the investigators. Where the information to be checked deals with time spent on various productive activities, one method of probing might be to ask for a day-by-day accounting of activities during the reference period. Similarly, if the initial survey asked for total amount spent on food during the past week, the reinterviewer might obtain disaggregated information by asking about each commodity purchased. Another possibility would be for the reinterviewer to ask permission to inspect the foods currently stored in the dwelling unit, to determine if any that might have been purchased during the reference period were overlooked.

Like replication reinterviews, probing reinterviews should be conducted independently of the initial interviews. There should always be a reconciliation of the initial interview and reinterview responses; the use of intensive probing technques is no guarantee that the reinterview responses will always be more accurate than the initial ones. The reconciliation process will improve the accuracy of the estimates of response bias, and will also provide an indication of the effectiveness of the particular probing approaches used in the reinterviews.

A reinterview study designed to produce overall estimates of response variance and/or bias should probably have a minimum sample size of 300 to 400 households. Where reinterviews are also conducted for quality control purposes, the sample sizes may be larger, since some of the work of each interviewer must be included.

(c) Record check studies

Record check evaluation studies compare survey data for specific households or individuals with information on the same topics from record-keeping systems external to the survey. If the data from records are believed to be sufficiently accurate to serve as a standard, the comparisons will provide a basis for estimating either coverage or response bias for the survey items covered by the record-keeping system. The technique can be used either in the survey itself, or in pretests.

The key requirement in the design of record check studies is to locate record systems that (1) contain information for some of the key variables included in the survey (2) contain information believed to be accurate and (3) cover a substantial proportion of the households in the survey target population, or some important sub-group of that population. Some types of record systems that might be relevant to household survey topics include:

(i) Records of public utilities, such as telephone or electric companies, or customer billings and payments.

(ii) Records kept by employers of hours worked and wages or salaries paid.

(iii) Records of government transfer payments, such as pension and welfare benefits.

(iv) Income tax records.

(v) Registers of persons licenses to practice certain professions or occupations.

The probability of the existence of potentially usable record systems is greatest for the urban areas of the more developed countries. In less developed areas, the coverage and quality of

record systems is less likely to be adequate for use in record checks.

Other important requirements for the conduct of record check evaluation studies are:

(i) Gaining access to the individual records. Permission to use the records will have to be obtained from the custodian of each record system used. Some custodians may require that waivers be obtained from survey respondents, giving their permission for the survey organization to seek access to their records in the system.

(ii) Locating the records of specific survey respondents. This will require a matching operation. The matching will be greatly facilitated if the characteristics of the record system to be used are known in advance of the survey, so that the identifiers needed for matching purposes can be obtained in the survey. In addition to names and addresses, identification or account numbers, if they are used, will be especially helpful.

Normally, the information from records in the system is obtained following the survey, when the identification of the sample households and individuals is known. A different technique, which is especially useful in connection with pretests, is the reverse record check. In a reverse record check, a sample of households or persons is selected from the record system, and included in the interviewer assignments. This approach has two important advantages:

(i) The sample can be designed to include households or persons with particular characteristics that are relevant to the survey topics.

(ii) The records needed for the study can be pulled or copied at the time the sample is selected so that there is no need for a difficult matching operation. Matching is guaranteed, except when the interviewer cannot locate the selected household or person.

A problem in doing a reverse record check in an actual survey is that the coverage of the record system used will probably not be identical to the survey target population, so that a more complex sample design would be needed to insure survey coverage of households or persons not included in the record system.

The major advantages of record checks in general are accuracy and relatively low cost. It is not safe to assume that data in record-systems are completely accurate, but accuracy is an important requirement for most record systems. Transactions are frequently recorded at the time they occur, in contrast to a retrospective survey or reinterview in which the degree of ability to recall past events is an important determinant of accuracy. The cost of obtaining the record check information tends to be low compared to, say, reinterviews, since the information is obtained from a single location or at most from a small number of locations where the files are kept.

Record check studies have certain disadvantages, some of which have already been mentioned:

(i) They do not provide any information about bias for those members of the survey target population who are not covered by the record system.

(ii) There may be conceptual or definitional differences between the survey and record system data, making direct comparisons difficult or impossible for some variables.

(iii) They depend on the willingness of respondents and record-system custodians to grant access to the records.

(iv) Except for reverse record checks, a difficult matching operation may be required. This may partially offset the savings from not having to conduct reinterviews, and may limit the usefulness of the results.

Several record checks have been conducted in connection with household surveys, especially in the areas of health (U.S. National Center for Health Statistics, 1972, 1973a) and crime victimization

(U.S. Committee on National Statistics, 1976). More recently, reverse record checks have been used to evaluate the accuracy of reporting of certain kinds of welfare payments in surveys designed to obtain information on income and participation in government programmes (U.S. Social Security Administration, 1980, Chapters 10 and 11).

3. Analytical methods of evaluating the quality of survey results

Survey tabulations can be examined and analyzed in many different ways to gain insight into their accuracy and how it may have been affected by various kinds of errors. Some methods of analysis are purely internal, depending only on data from the survey. Others are based on comparisons with aggregate data available from other sources. Insofar as possible, these analyses should be carried out as soon as the survey tabulations are available, so that any obvious failures in the data processing can be corrected before the results are released to the public in published or other form. Information on less serious errors and those which cannot be corrected for the current survey will become part of the body of data available for use in designing future surveys.

There are several possibilities for internal validity checks, as they are called in the revised Handbook of Household Surveys, (United Nations, 1981). Purely arithmetic checks can be made to see whether lines and columns of tables add to their respective totals and whether per cents have been computed correctly. Other checks may be used to determine whether the results are in accordance with prior expectations or common-sense judgements about how a particular set of data should behave. For example, the distribution of the survey population by age and sex can be examined by the construction of age pyramids, the calculation of age-specific sex ratios, and the calculation of preference indexes for terminal digits of age. These measures can be looked at for the total survey population and also for various subgroups defined geographically or in other ways.

There are certain patterns that can be expected for other survey items. As stated in the Handbook:

> "... birth rates would be expected to decline in an orderly manner for each successively older female age group. Income would be expected to rise with increasing education or occupational skill, etc."

Similarly, labor force participation rates for males would be expected to follow a unimodal distribution by age, peaking for middle-aged adults and gradually declining thereafter.

Failure of the survey results to be in accord with these prior expectations does not necessarily mean that the data are in error; however, it does suggest the need for more intensive review of the processes used to obtain the results.

Usually, there are also opportunities to make external validity checks. The survey results can be compared with aggregate data from other sources which differ in organization, methodology or in any other way. If there are no a priori grounds for preferring one estimate to another, the divergence between two or more estimates may provide a basis for at least some appraisal of the margin of uncertainty. Generally, however, there may be a number of reasons to expect one source of data to be more accurate than the other. As noted by Mahalanobis and Lahiri (1961), in such comparisons

> "... a complete absence of a priori preferences is not always a reality. For example, if it is known that a particular agency has had a long experience in a particular field, then the results thrown up by it may be accepted to have higher validity. Or, when one survey is carried by temporary ad hoc staff and another survey by a whole-time permanent statistical staff, then some may be inclined to accept the results thrown up by the latter to have higher validity. For the same reason one may accept an intensively supervised well-conducted sample survey by qualified, experienced and well-trained investigators to have higher validity compared to a complete enumeration conducted under usual census conditions."

However, the authors note that there may be difficulties in such a priori evaluation; for example:

> "... the assessor may not be in possession of full background information about the agency, or about the conditions under which the census data, for example, may have been collected or processed. Sometimes comparisons between census and sample check or those between interpenetrating samples covered by different agencies may provide corroborative evidence in support of certain 'feelings' based on previously available background information which may even be of vague and inadequate nature."

Insofar as different sources are not equally preferred, differences between a less-preferred source and a source believed to be more accurate would be indicative of biases in the former.

Of all possible sources for external validity checks, data from an earlier round of the same survey, or an earlier survey conducted under the same general conditions, are probably the most useful. Many population characteristics, such as the distribution by race or ethnic categories or by age, sex and marital status, are relatively stable, and sudden departures from earlier patterns and trends would suggest the possibility of error. An interesting example occurred in the annual education supplement to the United States Current Population Survey, when the enrollment of students in private schools at secondary and university levels in 1980 showed very substantial increases over the figures for 1979. A careful review of the reasons for this change showed that it was apparently due to a reversal of the order of the response categories "Private" and "Public" from those that had been used in prior years and in another part of the questionnaire in the same year.

Another important source of data for external validity checks is the latest population census. Frequently, survey estimates are adjusted to current estimates of population by age and sex that are obtained by projecting the corresponding census totals to the survey reference date. The unbiased or unadjusted survey estimates by age and sex should always be compared with the independent estimates based on the census; their ratios should be close to 1.00. High or low ratios may suggest coverage or age-reporting problems in the census or the survey, or inaccuracy in the projections of census data. This type of comparison was discussed in Section F.3 of Chapter III.

It may also be useful to compare other items that are common to the census and the survey, taking into account changes that are known or believed to have occurred between the census and survey reference dates.

There may also be other sources of data that can be used in external validity checks. Household survey data on school enrollment can be compared with data from the school system. Data on certain kinds of employment can be compared with data collected from employers in economic censuses or surveys. Survey data on births can be compared with information from the vital registration system. In making such comparisons, as pointed out in the Handbook, allowances must be made for differences in concepts and coverage, differences in timing, sampling variation, and known deficiencies in the sources used for comparison.

The general philosophy and justification for this type of comparative analysis was stated clearly by Simmons (1972):

> "It will be readily understood that comparitive analysis does not require a close match between survey results and the other evidence. The other evidence may be outdated, weak, or wrong. Indeed, if the survey was justified and well-conducted, the evidence from it is likely the best that can be found. But the comparisons should be made. For in some instances, they may uncover flaws in tabulation or estimation that can be corrected, or at least recognized and treated in subsequent analysis. More often, they will focus attention on new findings, contrasts, contradictions, and highlights for which consumers may expect some explanation, and on which final reports should comment."

Persons who use analytic techniques for the evaluation of survey data should have some knowledge of the basic methods of demography. The Methods and Materials of Demography (Shrvock, Siegel and associates, 1971) provides a comprehensive treatment of the analysis of demographic data from censuses and surveys, and includes many examples from developing countries.

C. Long-range strategies for the improvement of quality

Improving the quality of survey data is a slow and long-term process, requiring careful planning and allocation of resources, both human and material. A continuing and integrated programme of surveys offers the best environment for such a process. Also necessary is a commitment on the part of the survey organization to setting and achieving high standards of quality, taken in its broad sense to include the relevance, timeliness and accuracy of survey results.

Some key elements of an integrated household survey programme that are relevant to the long-term improvement of quality are discussed briefly in this final section.

1. Common infrastructural facilities.

The basic idea of an integrated household survey programme is to develop infrastructure facilities that are essential for the planning and execution of household surveys and to design and schedule individual surveys in such a way that the same facilities can be used in each survey. This reduces the average cost per survey and at the same time makes possible continuing development and improvement of these survey-taking facilities.

Among the most important infrastructure facilities are: a professional staff trained in substantive and methodological work; experienced field supervisors; at least a core of permanent field interviewing staff; a well-mapped and up-to-date sampling frame; and dependable data processing and dissemination facilities. All of these facilities, as has been made clear in previous chapters, have important implications for the control and long-term reduction of non-sampling errors.

The preparation and maintenance of an adequate sampling frame is often one of the most expensive elements in sample surveys in developing countries. Many problems of coverage and non-response result from the use of out-of-date and poor quality frames. The cost advantages of using a common frame and possibly a common master

sample for a series of individual surveys are obvious; better quality is also likely to result from this approach.

Poor questionnaire design is a major source of non-sampling error. Substantial returns in the improvement of quality may result from the establishment of a small unit specializing in questionnaire development, consisting perhaps of 1 to 3 professionals and a few highly experienced interviewers. This unit would do considerable field work, specializing in the use of qualitative methods of questionnaire development such as in-depth individual and group interviews, as described in Chapter VI, Section G.

2. Planning for the assessment of quality

Assessment of the quality of the data collected should be treated as an integral part of each survey. It would be a mistaken policy to consider the mere collection of more data as the primary task, and to relegate evaluation to a secondary position. Rather than being a subsidiary operation which competes for time and resources with the primary objective of collecting more and more data, evaluation of data and procedures should be regarded as an activity that will result not only in better data but also in making the process of collection itself more efficient and cost-effective.

The previous section of this chapter covered various informal and formal methods of gathering information during surveys about the magnitude and effects of non-sampling errors. No statistical organization, no matter how advanced, is likely to have the resources to use all of these methods in a particular survey. However, many organizations could do much more than they are now doing along these lines with a relatively small additional effort. It is suggested that the following guidelines be observed in planning for the assessment of quality:

(a) Consider the assessment of quality to be an essential and integral part of any continuing survey programme. It should not be assumed that, once a certain level of proficiency in conducting surveys is achieved, efforts at evaluation and methodological development can be de-emphasized. The process must be continuous because the needs of users and the economic, social and technological environments in which surveys are conducted are constantly changing.

(b) Make organizational and personnel arrangements necessary to plan and conduct evaluation and methodological development work. The activities which have been discussed, while they require the full co-operation of those responsible for survey operations, also call for objectivity and the ability to concentrate on the work of evaluation and assessment, free of operational responsibility for the surveys. Evaluation and methodological development activities should, therefore, be placed outside of the unit that is in charge of survey operations, and the individual in charge of these activities should be given adequate authority and resources to work on an equal footing with other members of the survey team, subject to the same overall direction.

(c) Prepare a formal plan for evaluation and methodological development work. Those who conduct surveys are fully aware of the need for a detailed plan to insure that the many interrelated activities involved proceed on schedule and fit together to produce a smooth, efficient operation. Formal planning is also necessary for evaluation and methodological development activities. In part, this is for the same reasons, that is, to insure that the necessary forms, instructions and other materials will be ready when needed, and that personnel will be available and properly trained in their functions. The timing of built-in experiments and formal evaluation studies must be very closely coordinated

with that of the survey operations. In addition, good research practice in conducting formal experiments demands that the outcome measures and at least a general plan of analysis be specified in advance. If this is not done, the chances of drawing useful conclusions are much smaller.

Formal plans for evaluation methods and development activities are needed at three levels:

(i) for each individual activity carried out in connexion with a survey, for the reasons just given;

(ii) for the survey as a whole, to insure that the highest priority needs are met, and that the overall programme of evaluation and methodological development is compatible with the resources available; and

(iii) for future survey methods research and development activities based on experience to-date.

Findings from specific experiments and evaluation studies will almost certainly indicate a need for further improvements. Following each survey, these results should be reviewed, a list of areas needing further attention prepared, and recommendations drafted for additional research and development activities, with an indication of priorities. This amounts to an updating of the long-term evaluation and methodological development agenda following completion of each survey.

(d) <u>For formal evaluation studies, use simple designs and adequate samples</u>. The techniques of experimental design have become very sophisticated. Multi-factor experiments are common-place and computer software is available for analysis of various complex designs. This is fine for controlled experiments in a laboratory or at an agricultural research station. However, controlling the conduct of a complex experiment as part of a national survey, with interviewers working in remote locations, is another story. Simplicity is strongly recommended. Basically, this means resisting the temptation to try to explore too many different issues in a single experiment or evaluation study.

The injunction to use adequate samples covers two main points:

(i) use probability samples as much as feasible in order to permit generalization of the findings to a well-defined population; and

(ii) where quantitative results are needed, use samples large enough to produce reliable estimates of components of error, unit costs, treatment effects, or whatever is being measured. It is generally not very useful simply to be able to say that there is a statistically significant difference in outcomes between two procedures, especially if they have different costs. A reasonably reliable estimate of the size of the difference is needed. A corollary to this recommendation is that large samples, usually a minimum of several hundred, are needed; therefore, this kind of research is better done as part of a survey, rather than in the pretesting stage or as a separate experiment.

3. Documentation and reporting.

There can be no development of enduring instruments and skills in survey-taking without a systematic accumulation of experience and knowledge of the conditions under which the surveys are conducted, the sources of difficulties in implementation of each enquiry, the shortcomings in the data collected, the sources of errors, and the cost-effectiveness of the options available in survey design and of measures necessary for the measurement and control of errors. In Chapter I, reference was made to the idea of a catalogue of problems and survey methodology. Because of the critical importance of this concept efforts to control and reduce non-sampling errors in household surveys, it seems fitting to end this study with a further reference to it.

The catalogue or archive should include general background material on the conditions and resources available for survey work in the country. This might cover such topics as languages spoken in different areas; regional variations in economic structure and social customs; the kinds of maps available at different levels of detail; the kinds of administrative records maintained for individuals and housing units by national and local authorities (for example, birth and death records); and identification

and description of all organizations that conduct household surveys.

Another component of the collection might be a set of publications of regional and international organizations containing recommendations and guidelines, especially those relating to concepts and definitions that are common to most household surveys. Some of these publications may also contain useful information about techniques used by other countries with similar problems.

Reference was also made in Chapter I to the literature on survey methods. Numerous citations of books, manuals, reports and articles on survey methods have been included in this document; copies of those items that appear most relevant should be obtained and made part of the collection. For many items, this can be done by writing directly to the authors. A conscious continuing effort is needed to find out what publications are available, obtain them, catalogue them and make them available to the survey staff.

The most important element of the catalogue or archive, however, will be the materials relating directly to the surveys conducted by the organization. For each survey, the following kinds of materials should be obtained and preserved in a permanent collection:

(i) Questionnaires, other forms, instruction manuals, training materials, procedures manuals and specifications from all phases of the survey.

(ii) Descriptions of the overall survey design and the design and procedures for each phase of the survey.

(iii) Operating records, such as cost and production records and quality control records, for each phase.

(iv) Documents describing the procedures and findings resulting from use of the informal and formal evaluation techniques described in Section B of this chapter.

Too often, materials of this kind, which have potential value to aid both in the interpretation of survey results and in the design of future surveys, are lost because of failure to document results while survey operations are underway and failure to develop a plan for preserving them.

The survey organization may find it helpful to establish one or more series of memoranda for purposes of internal documentation of survey methodology. A single series might be called "Survey Methodological Documentation Memoranda"; alternatively, separate series might be established for various purposes, e.g.:

(i) Survey planning memoranda, to describe developmental activities for the survey programme and for each survey;

(ii) Survey operations memoranda, to provide records and analyses of costs, production and quality control findings for the basic survey operations;

(iii) Survey evaluation and research memoranda, to describe the design, procedures and results of evaluation studies and built-in experiments.

Survey staff should be encouraged and, if necessary, assigned to prepare such internal memoranda. Individual memoranda should usually be fairly short, directed toward a specific aspect of the survey activities, informal in style, and prepared on a timely basis, as soon as relevant data and other information are available. Elaborate tables and charts are not necessary; working tables or even computer printouts might often be used to supplement the text. There should be an internal review process, primarily to ensure accuracy and objectivity of reporting. Although the series is established for internal purposes, it will also be useful as source material for the preparation of external publications and presentations. In particular, a summary of the methods used to conduct the surveys and limitations of the survey

results, as determined through formal evaluation studies and by other means, should be prepared for inclusion in final survey reports.

Other media for the dissemination of findings from evaluation and methodological studies are also available and their use by the survey staff should be encouraged. The publication or presentation of such reports in professional journals or bulletins and at meetings of professional and technical associations has several benefits:

(i) It ensures a wide audience and often some feedback from others working on similar problems;

(ii) Publication brings recognition to the authors and contributes to their professional development. It also brings recognition to the organization to which they belong;

(iii) Preparing such articles and presentations is a form of documentation, to ensure that important information on design, procedures and results will be preserved.

Administrators of survey organizations may sometimes be reluctant to allow public dissemination of information on non-sampling errors because of a feeling that this will reflect unfavourably on them and on the work of their organization. In the short-run, such openness may indeed cause some problems for the survey administrator. In the longer term, however, a policy of openness should prove to be beneficial. First, the survey staff will realize that they cannot afford to be complacent about the quality of survey results, and will have an incentive to strive constantly for improvements in quality and efficiency. Second, users will eventually come to realize that the only survey organization they can trust is one which attaches importance to the measurement and control of all kinds of errors and which is willing to tell its users whatever it knows about the quality of its product.

ANNEX A

BIBLIOGRAPHY

This bibliography has two parts. Part 1 is an annotated recommended reading and reference list. It is a short list of articles, reports and manuals which are judged to be particularly useful for those wishing to learn more about the topics covered in this document. Included are two bibliographies (reports a.3 and a.13) which may be used to supplement Part 2 of this bibliography.

Many books on survey methodology have been published. Some of these are excellent and include material which is quite relevant to the measurement and control of non-sampling errors. Only four such works are listed in Part 1; these were chosen because of their special relevance to surveys in developing countries.

Part 2 of the bibliography is an extended list of references, consisting primarily of those items that have been specifically referred to in the text.

Part 1

Short List

a. Articles, reports and manuals

1. American Statistical Association.
1978 ff. Proceedings of the Section on Survey Research Methods. Annual.

Compendia of papers on survey methods presented at the annual meetings of the American Statistical Association, starting in 1978. Prior to 1978, most papers on household survey methods were included in the annual Proceedings, Social Statistics Section.

As one example, the article "The effect of the question on survey responses: a review", by Kalton and Schuman, in the 1980 Proceedings gives an excellent summary of past research on the responses effects of variations in question wording and format.

2. Brooks, Camilla A. and Barbara Bailar.
1978. An Error Profile: Employment as Measured by the Current Population Survey, Statistical Policy Working Paper 3, U.S. Office of Federal Statistical Policy and Standards.

This document describes, in detailed, quantitative terms, the sources of non-sampling error in the Current Population Survey that affect national statistics of employment. The purposes of this document, as stated in the preface, were:

1) To illustrate how an error profile is created in an effort to encourage government statisticians to provide error profiles for the major recurrent survey statistics;

2) To compile in a single document the sources of error and the information that is available about these sources of error and their impact;

3) To illustrate the need for controlled experiments to measure non-sampling errors because of the lack of knowledge of the impact of these errors;

4) To stimulate the development of a mathematical model that will reflect the ways in which the errors from different sources interact.

3. Dalenius, Tore.
1977. Bibliography on non-sampling errors in surveys; International Statistical Review 45: 71-89 (A to G), 187-197 (H to Q), 303-317 (R to Z).

The author cites the need for additional research and development work on non-sampling errors to permit statisticians "...to strike a rational balance between sampling and non-sampling errors". The bibliography, which was published in three consecutive issues of the International Statistical Review, was "... developed with the aim of providing an aid for this research and development work".

4. Fellegi, Ivan and A.B. Sunter.
1974. Balance between different sources of survey errors -- some Canadian experiences. Sankhya: Indian Journal of Statistics. 36 (Series C): 119-142.

(Author's summary) A simple model for the allocation of resources to the reduction of survey error components, both sampling and non-sampling, is developed in order to demonstrate the nature of the decision-making process in determining this allocation.

A number of examples drawn from surveys carried out by Statistics Canada are used to demonstrate the application in real situations of the ideas contained in the model.

5. Hansen, M.H., W.N. Hurwitz and M.A. Bershad.
1961. Measurement errors in census and surveys. Bulletin of the International Statistical Institute, 38 (2): 359-374.

This was the first full presentation of the "Census Bureau error model", which permits the analysis of different components of sampling and non-sampling errors and their relative contribution to total survey error. Special attention is given to the measurement and control of the correlated component of response variance. An illustration is given from the 1950 United States Census of Population and Housing.

6. Jabine, T.B. and B.J. Tepping.
1973. Controlling the quality of occupation and industry data. Bulletin of the International Statistical Institute 45 (3): 360-89.

7. Mahalanobis, P.C. and D.B. Lahiri.
1961. Analysis of errors in censuses and surveys with special reference to experience in India. Bulletin of the International Statistical Institute, 38 (2): 401-433.

This paper, which was presented at the same meeting as the previous one, gives several illustrations of techniques used to evaluate survey results. All of these techniques relay on comparisons among alternative independent estimates. Special emphasis is given to the use of interpenetrating samples.

8. Neter, J. and J. Waksberg.
1965. Response Errors in Collection of Expenditures Data by Household Interview: An Experimental Study. Technical Paper No. 11, U.S. Bureau of the Census.

This report describes the design and findings from an experimental study of response errors in data on expenditures for home alterations and repairs collected in a large-scale household survey. This was the first study to provide a direct comparison of response effects with bounded and unbounded recall. Many other factors were also studied experimentally, including length of recall, length of reference period, choice of respondent and extent of probing on specific expenditure items. This is a notable example of the benefits of carefully designed methodological research during the early stages of a survey on a new topic.

Some of the findings are presented in Chapter V, Section C.3 of this document.

9. Simmons, Walt R.
1972. Operational Control of Sample Surveys, Laboratories for Population Studies, Manual Series No. 2. Chapel Hill; University of North Carolina.

This manual gives guidelines for establishing control systems to insure that the execution of a survey is faithful to its design. It gives attention to "...validation of the design; to control over data collection, editing, and processing; and to the preparation of estimates and the final report". Both formal procedures and more subjective control mechanisms are presented.

10. United Nations
1981. Handbook of Household Surveys (draft revision) DP/UN (INT-79-020/2) with two addenda. New York.

This manual, a substantial revision and expansion of the 1964 version, is the basic reference document for the National Household Survey Capability Programme. It has been issued as three separate documents. The first contains the overall introduction and Part One, "General Survey Planning and Operations". The second document (addendum 1) contains Part Two, "Issues in

Survey Content, Design and Operations". This section is organized by subject matter. It discusses data requirements, concepts and methodological issues in household surveys covering the following topics: income, consumption and expenditure; employment; food consumption and nutrition; agriculture; health; and education and literacy (a section on demographic characteristics is to be added). The third document (addendum 2) contains Part Three, "Selected Issues from Regional Survey Experience". This section was prepared in collaboration with the UN regional commissions.

11. United Nations
1982. Survey Data Processing: A Review of Issues and Procedures (draft) DP/UN/INT-81-041/1.

Like the present document, this is one of a technical studies for the National Household Survey Capability Programmes. It addresses the various technical and operational problems of organization and implementation of data processing activities for household survey programmes undertaken by national statistical agencies in developing countries.

12. U.S. Bureau of the Census.
1974. Standards for the Discussion and Presentation of Errors in Data. Technical Paper No. 32.

This guide was developed to help Census Bureau employees meet their responsibilities for informing data users of the important limitations of estimates disseminated in publications or by other means. While primary emphasis is on sampling errors, many of the standards apply equally to the presentation of information about non-sampling errors.

A slightly revised version of the Standards was published as Part II of the September 1975 issue of the Journal of the American Statistical Association (Vol. 70. No. 351). A Spanish translation was issued in 1978 by the Inter-American Statistical Institute (Document 7432 Esp.).

13. U.S. Bureau of the Census.
1974. Indexes to Survey Methodology Literature. Technical Paper No. 34.

This publication is an index to some 2,500 articles, reports, manuals, memoranda and other documents on selected aspects of survey methodology that were contained in the Census Bureau's Survey Methodology Information System at the time of publication. The publication includes bibliographic, author, organization and key-word-in-context (KWIC) indexes to the collection.

Although no updates to these indexes have been published, the Survey Methodology Information System was maintained and added to through early 1982, at which time the Census Bureau concluded that it could no longer afford to support it. Plans for disposition of the existing collection are uncertain at this time.

14. U.S. Committee on National Statistics.
1982. Final report of the Panel on Incomplete Data.

This report is expected to be an authoritative and comprehensive reference on the sources, effects and treatment of incomplete data in surveys. For further information, see Chapter IV, Section G.

15. U.S. National Center for Health Statistics.
1977. A Summary of Studies of Interviewing Methodology. Data Evaluation and Methods Research Series 2, No. 69, U.S. Department of Health, Education and Welfare.

(Author's summary) A summary of methodological studies designed to test the effectiveness of certain questionnaire designs and interviewing techniques used in the collection of data on health events in household interviews and to investigate the role of behaviours, attitudes, perceptions, and information levels of both the respondent and the interviewer.

16. Verma, Vijay.
1981. Assessment of errors in household surveys. Bulletin of the International Statistical Institute. Vol. 49, 1981.

Provides a summary of various types of errors and reviews some of the basic issues in the assessment of errors in household surveys in developing countries.

b. Books

1. Casley, D.J. and D.A. Lury
1981. Data Collection in Developing Countries. Oxford: Oxford University Press.

2. Hursh-Cesar, Gerald and Prodipto Roy, eds.
1976. Third World Surveys: Survey Research in Developing Countries. New Delhi: MacMillan Co. of India, Ltd.

3. Som, R.K.
1968. Recall Lapse in Demographic Enquiries. Bombay: Asia Publishing House.

4. Zarkovich, S.S.
1966. Quality of Statistical Data. Rome: Food and Agriculture Organization of the United Nations.

In many respects, the present document may be taken as an update and expansion of Zarkovich's book. The book was based on material presented by the author at various FAO-sponsored seminars and training seminars. Its aim was "...to spread awareness of the quality problem of statistical data and to promote interest in quality checks as a source of guidance on the adequate uses of the data and on the ways and means of improving the methods used".

Part 2

Extended List

Anderson, John E.
1979. Fertility estimation: A comparison of results of contraceptive prevalence surveys in Paraguay, Sao Paulo State, Brazil and El Salvador. Proceedings of the Social Statistics Section, American Statistical Association, 532-537.

Anderson, R., J. Kasper, M.R. Frankel, et al.
1979. Total Survey Error. San Francisco: Jossey-Bass.

Bailar, Barbara A.
1975. The effects of rotation group bias on estimates from panel surveys. Journal of the American Statistical Association, Vol. 70: 23-30.

1979. Rotation sampling biases and their effects on estimates of changes. 43d session of the International Statistical Institute, Manila.

Bailar, Barbara A. and C. Michael Lanphier
1978. Development of Survey Methods to Assess Survey Practices. Washington: American Statistical Association.

Bailey, L., T.F. Moore and B. Bailar
1978. An Interviewer Variance Study for Eight Impact Cities of the National Crime Survey Cities Sample. Journal of the American Statistical Association, Vol. 73: 16-23.

Barnes, R., and F. Birch
(nd) The census as an aid in estimating the characteristics of non-response in the General Household Survey. Office of Population Censuses and Surveys, Study NM 1, London.

Bennett, C.M. and R.E. Hill
1964. A comparison of selected personality characteristics of respondents and non-respondents to a mailed questionnaire. Journal of Educational Research, Vol. 58. No. 4: 178-180.

Bergman, L.R., R. Hanve and J. Rapp
1978. Why do some people refuse to participate in interview surveys? Statistisk Tidskrift.

Bradburn, Norman M. and Seymour Sudman
1980. Improving Interview Method and Questionnaire Design. San Francisco: Jossey Bass.

Brooks, Camilla A. and Barbara Bailar
1978. An Error Profile: Employment as Measured by the Current Population Survey. Statistical Policy Working Paper 3. Office of Federal Statistical Policy and Standards. U.S. Department of Commerce.

Caldwell, J.C. and A. Igun
1971. An experiment with census-type age enumeration in Nigeria. Population Studies, Vol. 25.

Casley, D.J. and D.A. Lury
1981. Data Collection in Developing Countries. Oxford: Oxford University Press.

Chapman, David D. and Charles E. Rogers
1978. 1978 Census of Agriculture - Area sample design and methodology. Proceedings of the American Statistical Association Section on Survey Research Methods, 141-147.

Chapman, D.W.
1976. A survey of non-response imputation procedures. Proceedings, Social Statistics Section, American Statistical Association, 245-251.

Chevry, G.
1949. Control of a general census by means of an area sampling method. Journal of the American Statistical Association, Vol. 44: 373-379.

Chidambaram, V.C., J.G. Cleland and V. Verma
1980. Some aspects of WFS data quality: a preliminary assessment. World Fertility Survey Comparative Studies, No. 16. The Hague: International Statistical Institute.

Chow, L.P., Walter Gruhn and Wen Pin Chang
1979. Feasibility of randomized response technique in rural Ethiopia. American Journal of Public Health, Vol. 69: 273-276.

Cobb, J.M., S. King and E. Chen
1957. Differences between respondents and non-respondents in a morbidity survey involving clinical examination. Journal of Chronic Diseases, Vol. 6.

Cole, D.
1956. Field work in sample surveys of household income and expenditure. Applied Statistics, Vol. 5: 49-61.

Cole D., and J.E.G. Utting
1956. Estimating expenditure, saving and income from household budgets. Journal of the Royal Statistical Society, Vol. 119, Series A, 371-392.

Cornfield, J.
1942. On certain biases in sampling of human populations. Journal of the American Statistical Association, Vol. 37: 63-68.

Dalenius, Tore.
1957. Sampling in Sweden. Stockholm: Almqvist and Wiksell.

1977. Bibliography on non-sampling errors in surveys. International Statistical Review 45: 71-89 (A to G), 187-197 (H to Q), 303-317 (R to Z).

Das Gupta A. and S.N. Mitra
1958. A technical note on age grouping. The National Sample Survey, No. 12, New Delhi.

Das Gupta A., R.K. Som, M. Majumdar and S.N. Mitra.
1955. Couple fertility. The National Sample Survey, No. 7, New Delhi.

Demaio, T.Y.
1980. Refusals: who, where and why? Public Opinion Quarterly, Vol. 44.

Deming, W.E.
1953. On a probability mechanism to attain an economic balance between the resultant error of response and the bias of non-response. Journal of the American Statistical Association, Vol. 48: 743-772.

Dunn, J.P. and R. Hawkes
1966. Comparison of non-respondents and respondents in a periodic health examination program to a mailed questionnaire. American Journal of Public Health, Vol. 56: 230-236.

Ewbank, D.C.
1980. Implications for demographic analysis for various levels and patterns of age misreporting. Panel on Data Collection of the Committee on Population and Demography, National Research Council (manuscript).

Fellegi, Ivan P.
1964. Response variance and its estimation. Journal of the American Statistical Association, Vol. 59: 1016-1041.

1973. The evaluation of the accuracy of survey results: some Canadian experiences. International Statistical Review, Vol. 41: 1-14.

1974. An improved method of estimating the correlated response variance. Journal of the American Statistical Association, Vol. 69: 496-501.

Fellegi, Ivan and A.B. Sunter
1974. Balance between different sources of survey errors - some Canadian experiences. Sankhya: the Indian Journal of Statistics, 36 (Series C): 119-142.

Ferber, R.
1955. On the reliability of responses secured in sample surveys. Journal of the American Statistical Association, Vol. 50: 788-810.

Finkner, A.L.
1950. Methods of sampling for estimating commercial peach production in North Carolina. Agricultural Experimentation Station Technical Bulletin, No. 91.

Gales, K. and M.G. Kendall
1957. An enquiry concerning interviewer variability. Journal of the Royal Statistical Society, Vol. 120, Series A: 121-138.

Ghosh, A.
1953. Accuracy of family budget data with reference to period of recall. Calcutta Statistical Association Bulletin, Vol. 5: 16-23.

Goldberg, S.A.
1957. Non-sampling errors in household surveys: a general review of some Canadian work. 32d session of the International Statistical Institute, Stockholm.

Gower, A.R.
1979. Characteristics of non-respondents in the LFS (labour force survey). Statistics Canada (internal memorandum).

Gray, P.G.
1959. Examples of interviewer variability taken from two sample surveys. Applied Statistics, Vol. 8: 73-85.

Gray, P.G. and T. Corlett
1950. Sampling for the Social Survey. Journal of the Royal Statistical Society, Vol. 113, series A.

Gray P., and F.E.V. Gee
1972. A quality check on the 1966 ten per cent sample census of England and Wales. Office of Population Censuses and Surveys. London: Her Majesty's Stationery Office.

Haase, K.W. and R.W. Wilson
1972. The study design of an experiment to measure the effects of using proxy responses in the National Health Interview Survey. Proceedings of the Social Statistics Section, American Statistical Association: 289-293.

Hansen, M.H. and W.N. Hurwitz
1946. The problem of non-response in sample surveys. Journal of the American Statistical Association, Vol. 41: 517-529.

Hansen, M.H. W.N. Hurwitz and M.A. Bershad
1961. Measurement errors in censuses and surveys. Bulletin of the International Statistical Institute, 38(2): 359-374.

Hansen, M.H., W.N. Hurwitz and T.B. Jabine
1963. The use of imperfect lists for probability sampling at the Bureau of the Census. 34th session of the International Statistical Institute, Ottawa.

Hansen, M.H. W.N. Hurwitz and H. Nisselson and J. Steinberg
1955. The redesign of the Current Population Survey. Journal of the American Statistical Association, Vol. 50: 701-719.

Hansen, M.H. and J. Steinberg
1956. Control of errors in surveys. Biometrics, 462-474.

Hansen, M.H. and J. Waksberg
1970. Research on non-sampling errors in censuses and surveys. Review of the International Statistical Institute, 317-332.

Hanson, R.H. and E.S. Marks
1958. Influence of the interviewer on the accuracy of survey results. Journal of the American Statistical Association, Vol. 53: 635-655.

Hauck, M. and S. Steincamp
1964. Survey Reliability and Interviewer Competence. Bureau of Economic and Business Research, University of Illinois.

Hendricks, W.R.
1956. The Mathematical Theory of Sampling. New Brunswick: Scarecrow Press.

Hirschberg, David, Frederick J. Scheuren and Robert Yuskavage
1977. The impact on personal and family income of adjusting the Current Population Survey for undercoverage. Proceedings of the Social Statistics Section, American Statistical Association. 70-80.

Hoinville, Gerald, Roger Jowell and associates
1979. Survey Research Practice. London: Heinemann Educational Books.

Horvitz, D.G. and K.M. Wolter
1977. Total survey design. Advances in Health Survey Research Methods, U.S. Department of Health, Education and Welfare Publication No. (HRA) 77-3154.

Huang, E.T., C.R. Hogue and C.T. Isaki
1980. Comparisons of multi-frame with single frame sample designs using registration and voting survey data. Proceedings of the American Statistical Association Section on Survey Research Methods, 711-715.

Hursh-Cesar, G. and Prodipo Roy (eds)
1976. Third World Surveys: Survey Research in Developing Countries. New Delhi: MacMillan Co. of India, Ltd.

Indonesia Central Bureau of Statistics
1978. Indonesian Fertility Survey 1976. Principal Report. Jakarta.

Jabine, T.B. and Naomi Rothwell
1970. Split-panel tests of census and survey questionnaires. Proceedings of the Social Statistics Section, American Statistical Association: 4-13.

Jabine, T.B. and B.J. Tepping
1973. Controlling the quality of occupation and industry data. Bulletin of the International Statistical Institute, 45(3): 360-389.

Jabine, Thomas B.
1981. Guidelines and recommendations for experimental and pilot survey activities in connexion with the Inter-American Household Survey Program. Inter-American Statistical Institure: Washington.

Kalsbeek, William and Judith Lessler
1978. Total survey design: effect of non-response bias and procedures for controlling measurement errors. In Health Survey Research Methods. Second Biennial Conference. Department of Health, Education and Welfare Publication No. (PHS) 79-3207.

Kemsley, W.F.F.
1965. Interviewer variability in expenditure surveys. Journal of the Royal Statistical Society, Vol. 128, Series A, 118-139.

Kemsley, W.F.F. and J.L. Nicholson
1960. Some experiments in methods of conducting family expenditure surveys. Journal of the Royal Statistical Society, Vol. 123, Series A, 307-328.

Keyfitz, N.
1979. Information and allocation; two uses of 1980 census. The American Statistician, Vol. 33: 45-56.

Kish, L.
1962. Studies of the Interviewer variance for attitudinal characteristics. Journal of the American Statistical Association, Vol. 57: 92-115.

1965. Survey Sampling, New York: Wiley.

Kish, L., R.M. Groves and K. Krotki
1976. Sampling errors for fertility surveys. WFS Occasional Paper No. 17. The Hague: International Statistical Institute.

Kish, L. and I. Hess
1958. On non-coverage of sample dwellings. Journal of the American Statistical Association, Vol. 54: 509-524.

1959. A "replacement" procedure for reducing the bias of non-response. American Statistician, Vol. 13, No. 4: 17-19.

Kovar, Mary G. and Ronald W. Wilson
1976. Perceived health status - How good is proxy reporting? American Statistical Association, Proceedings of the Social Statistics Section. Part II: 495-500.

Kovar, Mary G. and Robert A. Wright
1973. An experiment with alternate respondent rules in the National Health Interview Survey. American Statistical Association, Proceedings of the Social Statistics Section, 311-316.

Lessler, J.T.
1974. A double sampling scheme model for eliminating measurement process bias and estimating measurement errors in surveys. Institute of Statistics Mimeo Series No. 949. University of North Carolina, Chapel Hill.

1980. Errors associated with the frame. Proceedings of the American Statistical Association Section on Survey Research Methods, 125-130.

Lininger, Charles A. and Donald P. Warwick
1975. The Sample Survey: Theory and Practice. New York: McGraw-Hill.

Lubin, B., E. Levitt and M. Zuckerman
1962. Some personality differences between responders and non-responders in a survey questionnaire. Journal of Consulting Psychology, Vol. 26: 192.

Lundberg, G.A. and O.A. Larsen
1949. Characteristics of hard-to-reach individuals in field surveys. Public Opinion Quarterly, Vol. 13: 487-494.

Lyberg, L. and E. Rapaport
1979. (unpublished) Non-response problems at the National Central Bureau of Statistics, Sweden.

MacDonald, A.L., P.M. Simpson and A.M. Whitfield
1978. An assessment of the reliability of the Indonesian Fertility Survey data. WFS Scientific Report No. 3. The Hague: International Statistical Institute.

Mahalanobis, P.C.
1946. Recent experiments in statistical sampling in the Indian Statistical Institute. Journal of the Royal Statistical Society, Vol. 109, Series A: 325-370.

Mahalanobis, P.C. and S.B. Sen
1954. On some aspects of the Indian National Sample Survey. Bulletin of the International Statistical Institute, Vol. 34, pt. 2.

Mahalanobis, P.C. and A. Das Gupta
1954. The use of sample surveys in demographic studies in India. Proceedings of the United Nations World Population Conference, Vol. 6.

Mahalanobis, P.C. and D.B. Lahiri
1961. Analysis of errors in census and surveys. Bulletin of the International Statistical Institute 38(2): 359-374.

Marckwardt, Albert M.
1973. Evaluation of an experimental short interview form designed to collect fertility data: the case of Peru. Demography 10: 639-657.

Marks, Eli S. and John C. Rumford
1976. The 1974 Post-Enumeration Survey of Liberia - a new approach. Proceedings of the American Statistical Association Social Statistics Section (II), 578-583.

Mauldin, P.W. and E.S. Marks
1950. Response errors in census research. Journal of the American Statistical Association, Vol. 45: 425-438.

Mazumdar, M.
1962. Vital rates. The National Sample Survey, No. 54, New Delhi.

McCord, H.
1951. Discovering the "confused" respondent: a possible projective method. Public Opinion Quarterly, Vol. 15: 363-365.

McNeil, John M.
1981. Factors affecting the 1980 census content and the effort to develop a post-census disability survey. Presented at the Annual Meeting of the American Public Health Association.

McWhinney, Isabel and Harold E. Champion
1974. The Canadian experience with recall and diary methods in expenditure surveys. Annals of Economic and Social Measurement 3: 411-437.

Mooney, H.
1962. Methodology in Two California Health Surveys. Public Health Monograph No. 70.

Moser, C.
1951. Interview bias. Review of the International Statistical Institute, Vol. 19: 28-40.

Moser, C.A. and G. Kalton
1978. Survey Methods in Social Investigation (revised edition). New York: Basic Books.

Murphy, L.R. and C.D. Cowan
1976. Effects of bounding on telescoping in the National Crime Survey. Proceedings of the Social Statistics Section, American Statistical Association, 633-638.

Seltzer, W.
1973. Demographic Data Collection: a Summary of Experience. New York: The Population Council.

Shah, Nasra M.
1981. Data from tables used in presented at weekly seminar of East-West Population Institute, October 28. Honolulu.

Shryock, H.S. Jr., and J. Siegel
1971. The Methods and Materials of Demography. Washington, D.C.: U.S. Bureau of the Census

Siegel, J.S.
1974. Estimates of coverage of population by sex, race and age: demographic analysis. 1970 Census of Population and Housing Evaluation and Research Program, PHC (E)-4, Washington, D.C.

Siegel, Paul M.
1981. Estimates of private school enrollment for October 1980. Unpublished memorandum to Arthur J. Norton, U.S. Bureau of the Census.

Simmons, Walt R.
1972. Operational Control of Sample Surveys. Laboratories for Population Statistics, Manual Series No. 2. Chapel Hill: University of North Carolina.

Singh, M.P. and R. Tessier
1975. Method Test Panel (MTP) Phase III; feasibility test for complete non-proxy procedures. Methodology and Analysis, Statistics Canada.

Sirkin, Monroe, G.
1972. Designing Forms for Demographic Surveys. Laboratories for Population Statistics Manual Series No. 3. Chapel Hill: University of North Carolina.

Skelton, V.C.
1963. Patterns behind "income refusals". Journal of Marketing, Vol. 27.

Skunasingha, Annie and T.B. Jabine
1981. The administrative sub-divisions of Thailand and their relation to household surveys. Unpublished report.

Som, R.K.
1959. On recall lapse in demographic studies in India. International Population Conference, Vienna.

1968. Recall Lapse in Demographic Enquiries. Bombay: Asia Publishing House.

Stevens, J.A. and B.A. Bailar
1976. The relationship between various interviewer characteristics and the collection of income data. Proceedings, Social Statistics Section, American Statistical Association, 785-790.

Sukhatme, P.V.
1947. The problem of plot size in large-scale yield surveys. Journal of the American Statistical Association. Vol. 42: 297-310.

Sukhatme, P.V. and G. Seth
1952. Non-sampling errors in surveys. Journal of the Indian Society of Agricultural Statistics, Vol. 4: 5-41.

Tezcan, Sabahat
1977. A comparative study of induced abortion in Turkey using the randomized response technique versus direct questioning. Disseration for Doctor of Public Health degree. University of North Carolina.

Thailand
1977. Survey of Fertility in Thailand, Country Report No. 1. Bangkok: National Statistical Office and Chulalongkorn University.

Thomsen, Ib and E. Siring
n.d. On the causes and effects of non-response: Norwegian Experiences. Central Bureau of Statistics, Norway (unpublished).

Turkey
1980. The Turkish Fertility Survey 1978: First Report. Ankara: Hacettepe University.

Turner, Anthony G., Henry F. Woltman, Robert Fay and Beverley Carlson
1979. Sample survey design in developing countries - three illustrations of methodology. Bulletin of the International Statistical Institute, 47.

Tuygan, Kuthan and Tevfik Cavdar
1975. Comparison of self and proxy responses related to children of ever-married women. In Laboratories for Population Statistics Scientific Report Series, No. 17: 22-28.

UNICEF
1980. Surveys of the disabled. Social Statistics Bulletin, Vol. 3, No. 7.

United Kingdom, Office of Population Censuses and Surveys
1978. Methods of reducing non-response and its effect on the quality of statistics. Statistical Commission and Economic Commission for Europe, Conference of European Statisticians, Meeting on Problems relating to Household Surveys document (CES/AC.48/5).

United Nations

1956. Demographic Yearbook.

1967. Methods of estimating basic demographic measures from incomplete data (Manuals on methods of estimating population: Manual IV). ST/SOA/Series A/42.

1977. Economic Commission for Africa. Study of special techniques for enumerating nomads in African censuses and surveys. (E/CN.14/CAS.10/16).

1980a Principles and Recommendations for Population and Housing Censuses. Statistical Papers, Series M. No. 67. Statistical Office.

1980b The National Household Survey Capability Programme: Prospectus.

1981 Handbook of Household Surveys (draft revision) DP/UN/INT-79-02o/2, with 2 addenda. New York.

1982. National Household Survey Capability Programme. Survey Data Processing: A Review of Issues and Procedures (draft) DP/UN/INT-81-041/1.

U.S. Bureau of the Census.

1963. The Current Population Survey Reinterview Program, Some Notes and Discussion. Technical Paper No. 6. Washington, D.C.

1964. Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Accuracy of Data on Population Characteristics as Measured by Reinterviews. Series ER60, No. 4. Washington, D.C.

1968a. Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Effects of Interviewers and Crew Leaders. Series ER60, No. 7. Washington, D.C.

1968b. The Current Population Survey Reinterview Program, January 1961 through December 1966. Technical Paper 19. Washington, D.C.

1972. Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Effects of Coders. Series ER60, No. 9. Washington, D.C.

1974a. Standards for the discussion and presentation of errors in data. Technical Paper No. 32.

1974b. Indexes to survey methodology literature. Technical Paper No. 34.

1978. An overview of population and housing census evaluation programs conducted at the U.S. Bureau of the Census. Census Advisory Committee of the American Statistical Association.

1979. Popstan: A Case Study for the 1980 Census of Population and Housing. Part B, planning and preparation for Popstan census. Washington, D.C.

1981. Report of CPS reinterview for 1980. Unpublished memorandum by Schreiner, Irv and Ellen Katzoff. Washington, D.C.

U.S. Committee on National Statistics
1976. Surveying Crime. Report of the Panel for the Evaluation of Crime Surveys. Washington: National Academy of Sciences.

1979. Symposium on Incomplete Data, Preliminary Proceedings. Panel on Incomplete Data. Washington: Social Security Administration.

1981. Surveys of Subjective Phenomena: Summary Report (Turner, Charles F. and Elizabeth Martin, eds.) Panel on Survey Measurement of Subjective Phenomena. Washington: National Academy Press.

1982. Final Report of the Panel on Incomplete Data (forthcoming).

U.S. National Center for Health Statistics
1972. Optimum recall period for reporting persons injured in motor vehicle accidents. Data Evaluation and Methods Research, Series 2, No. 50. U.S. Department of Health Education and Welfare.

1973a. Net differences in interview data on chronic conditions and information derived from medical records. Data Evaluation and Methods Research, Series 2, No. 57. U.S. Department of Health, Education and Welfare.

1973b. Quality Control and measurement of non-sampling errors in the Health Interview Survey. Data Evaluation and Methods Research, Series 2, No. 54. U.S. Department of Health, Education and Welfare.

1977. A summary of studies of interviewing methodology. Data Evaluation and Methods Research, Series 2. No. 69. U.S. Department of Health, Education and Welfare.

U.S. National Commission on Employment and Unemployment Statistics
1979. Counting the Labor Force. Washington: U.S. Government Printing Office.

U.S. Social Security Administration
1980. Reports from the Site Research Test, Income Survey Development Program. U.S. Department of Health and Human Services.

University of Michigan, Survey Research Center
1976. Interviewer's Manual (revised edition). Ann Arbor: University of Michigan.

Verma, Vijay
1977. Sample designs for the World Fertility Survey. Bulletin of the International Statistical Institute, Vol. 46, No. 3.

1979. Training, the key to quality data. POPULI, Vol. 6, No. 1.

1980. Sampling for National Fertility Surveys. World Fertility Survey Conference, London.

1981. Assessment of errors in household surveys. Bulletin of the International Statistical Institute, vol. 49.

Verma, V., C. Scott and C. O'Muircheartaigh
1980. Sample design and sampling errors for the World Fertility Survey. Journal of the Royal Statistical Society, Series A, Vol. 143, Pt. 4.

Ware, H.
1977. Language problems in demographic field work in Africa: the case of the Cameroon Fertility Survey. WFS Scientific Report No. 2. The Hague: International Statistical Institute.

Warner, S.L.
1965. Randomized response: a survey technique for eliminating evasive answer bias. Journal of the American Statistical Association, Vol. 60: 63-69.

Williams, W.H. and C.L. Mallows
1970. Systematic biases in panel surveys. Journal of the American Statistical Association, Vol. 65: 1338-1349.

Wolter, Kirk M. and David A. Pyne
1978. Optimum sample allocation to different modes of enumeration. Proceedings of the American Statistical Association Section on Survey Research Methods, 414-419.

Woltman, Henry and John Bushery
1975. A panel bias study in the National Crime Survey, Proceedings of the Social Statistics Section, American Statistical Association.

World Fertility Survey
1975a. Manual on Sample Design. Basic Documentation. The Hague: International Statistical Institute.

1975b. Training Manual. Basic Documentation. The Hague: International Statistical Institute.

1975c. Supervisors' Instructions. Basic Documentation. The Hague: International Statistical Institute.

1975d. Interviewers' Instructions. Basic Documentation. The Hague: International Statistical Institute.

1977. Guidelines for the Country Report No. 1. Basic Documentation. The Hague: International Statistical Institute.

Wright, Patricia
1980. Strategy and tactics in the design of forms. Visible Language XIV 2: 151-193.

You Poh Seng
1959. Errors in Age Reporting in Statistically Underdeveloped Countries. Population Studies, Vol. 3: 164-182.

Zarkovich, S.S.
1966. Quality of Statistical Data. Rome: Food and Agriculture Organization.

Zitter, Meyer and Edith K. McArthur
1980. Census undercount: the international experience. In Proceedings of the 1980 Conference on Census Undercount. U.S. Bureau of the Census: 164-180.

Printed in U.S.A.

40735-July 1991-500