Scientific Reports

NUMBER 28 MARCH 1982

C.A. O'MUIRCHEARTAIGH

Methodology of the Response Errors Project

INTERNATIONAL STATISTICAL INSTITUTE
Permanent Office. Director: E. Lunenberg
428 Prinses Beatrixlaan, PO Box 950
Voorburg
Netherlands

WORLD FERTILITY SURVEY
Project Director:
Dr Dirk J. van de Kaa
35–37 Grosvenor Gardens
London SW1W 0BS, UK

The World Fertility Survey is an international research programme whose purpose is to assess the current state of human fertility throughout the world. This is being done principally through promoting and supporting nationally representative, internationally comparable, and scientifically designed and conducted sample surveys of fertility behaviour in as many countries as possible.

The WFS is being undertaken, with the collaboration of the United Nations, by the International Statistical Institute in cooperation with the International Union for the Scientific Study of Population. Financial support is provided principally by the United Nations Fund for Population Activities and the United States Agency for International Development.

This publication is part of the WFS Publications Programme which includes the WFS Basic Documentation, Occasional Papers and auxiliary publications. For further information on the WFS, write to the Information Office, International Statistical Institute, 428 Prinses Beatrixlaan, Voorburg, The Hague, Netherlands.

L'Enquête Mondiale sur la Fécondité (EMF) est un programme international de recherche dont le but est d'évaluer l'état actuel de la fécondité humaine dans le monde. Afin d'atteindre cet objectif, des enquêtes par sondage sur la fécondité sont mises en oeuvre et financées dans le plus grand nombre de pays possible, Ces études, élaborées et réalisées de façon scientifique, fournissent des données représentatives au niveau national et comparables au niveau international.

L'Institut International de Statistique avec l'appui des Nations Unies, a été chargé de la réalisation de ce projet en collaboration avec l'Union Internationale pour l'Etude Scientifique de la Population. Le financement est principalement assuré par le Fonds des Nations Unies pour les Activités en matière de Population et l'Agence pour le Développement International des Etats-Unis.

Cette publication fait partie du programme de publications de l'EMF, qui comprend la Documentation de base, les Documents Non-Périodiques et des publications auxiliaires. Pour tout renseignement complémentaire, s'adresser au Bureau d'Information, Institut International de Statistique, 428 Prinses Beatrixlaan, Voorburg, La Haye, Pays-Bas.

La Encuesta Mundial de Fecundidad (EMF) es un programa internacional de investigación cuyo propósito es determinar el estado actual de la fecundidad humana en el mundo. Para lograr este objetivo, se están promoviendo y financiando encuestas de fecundidad por muestreo en el mayor número posible de países. Estas encuestas son diseñadas y realizadas científicamente, nacionalmente representativas y comparables a nivel internacional.

El proyecto está a cargo del Instituto Internacional de Estadística en cooperación con la Unión Internacional para el Estudio Científico de la Población y con la colaboración de las Naciones Unidas. Es financiado principalmente por el Fondo de las Naciones Unidas para Actividades de Población y por la Agencia para el Desarrollo Internacional de los Estados Unidos.

Esta publicación ha sido editada por el Programa de Publicaciones de la EMF, el que incluye Documentación Básica, Publicaciones Ocasionales y publicaciones auxiliares. Puede obtenerse mayor información sobre la EMF escribiendo a la Oficina de Información, Instituto Internacional de Estadística, 428 Prinses Beatrixlaan, Voorburg-La Haya, Países Bajos.

Scientific Reports

Methodology of the Response Errors Project

C.A. O'MUIRCHEARTAIGH
London School of Economics and Political Science

Contents

Preface

Right from the start, the WFS programme has placed emphasis on the need for assessing the magnitude and impact of the two commonly known kinds of errors – sampling and non-sampling – in survey data. The response errors project, carried out by WFS with financial support from the International Development Research Centre, Canada, forms a major component of the effort of WFS in this area. The main objectives of the project were to investigate certain types of response errors in the data collected in WFS surveys, to estimate the magnitude of these errors and to examine their implications for analysis as well as for future surveys.

The project comprised studies in four countries – Dominican Republic, Lesotho, Peru and Turkey – carried out along with the national fertility surveys. This report, the first of a set of six publications planned on this project, describes the methodology of the project which is common to all the four country studies. The next four reports will present the results from each of the country studies and the final report will attempt a comparative assessment of the results.

We are grateful to Mr Colm O'Muircheartaigh for his efforts and contribution during all stages of the project. I also recognize that the final outcome of a project of this nature is a result of collective effort and many other colleagues in the WFS and in the countries have made important contributions at different stages. In particular, I wish to acknowledge the contribution of Mr V.C. Chidambaram who, as the co-ordinator, played a major role in the planning and execution of the project as a whole.

Finally, I wish to express on behalf of WFS our thanks to the IDRC of Canada for their assistance and co-operation.

DIRK J. VAN DE KAA
Project Director

1 Introduction

This report describes the methodology of a project carried out by the WFS with the financial support of the International Development Research Centre, Canada. The objectives of the project are to investigate certain types of response error in the data collected in WFS surveys; to estimate the magnitude of these errors; and to examine their implications for the analysis of the data. The project is a component of the overall evaluation of data quality being conducted as part of the WFS analysis programme.

The major objective of the WFS has been to generate substantive results. The surveys co-ordinated by the WFS have had as their primary objective the provision of high quality data at the national level, while the WFS has attempted to achieve a degree of standardization in the collection and reporting of data relating to fertility by different countries.

From the outset the WFS has attempted to reduce to a minimum the delay between fieldwork and dissemination of results by encouraging countries to issue preliminary and largely descriptive Country Reports, based on a detailed standardized set of cross-tabulations. By March 1982, 29 such reports had been published. The completion of country reports marks the beginning of the analysis phase, in which detailed investigation of particular topics is carried out, using more refined demographic and statistical techniques than attempted in the First Country Reports. An essential component of this large-scale analysis programme is a critical evaluation of the quality of the data. The nature of the data collection process implies that the usefulness of the data depends critically on the magnitude and impact of two kinds of error, which are different both in their implications and in their source.

All of the survey data are based on information collected from a sample of the population of interest. Thus all of the estimates obtained from the surveys are subject to sampling variability. In each participating country the study consists of a single-round survey based on a probability sample of households. Though the sample in each case is designed individually to suit the country's situation, all of the samples were designed to be *measurable*, ie the design permits the estimation of sampling errors from the survey data themselves. As a matter of policy, estimates of sampling errors have been computed as part of the first stage of analysis. A full discussion of WFS sample characteristics and an analysis of sampling errors may be found in Verma, Scott and O'Muircheartaigh (1980).

In the absence of misreporting, the detailed fertility and marriage histories obtained in WFS surveys would make it possible to estimate levels and trends in age at marriage, fertility rates and infant and child mortality rates. These estimates would, of course, be subject to sampling variability, but even if the population were to be enumerated completely, the data and the conclusions reached could be subject to serious errors due to faults in the method of measurement or observation. These *response errors* may arise from the respondent, from the questionnaire, from the execution of the fieldwork or from the nature of the data collection process; the form, extent, sources and effects of these errors are the concern not only of survey design but also of survey analysis. Past experience has indicated that restrospective survey data of the WFS type are often particularly prone to such errors. The high standards set by WFS for the data collection operation are expected to result in better quality data than typically obtained in the past, but this expectation in no way obviates the need for a detailed assessment of the quality of the data. Such an evaluation will not only alert analysts by identifying any defects in the data, but may also throw light on the shortcomings of the WFS approach which can be taken into account in the design of future fertility surveys.

No data are free from error and the identification and examination of these errors can be undertaken for a variety of purposes and with considerable variation in applicability. Once a survey (or census or any data collection operation) has been carried out, the data set, as it stands, is the raw material with which the analyst has to work, and any modification of it must be justified by some explicit knowledge, belief or assumption which is sufficiently strong to outweigh the recorded observation. The extent to which response errors are present in the data may be investigated by direct scrutiny, combined with a knowledge of, or assumptions about, the form and content which the data ought to reflect.

Data assessment at its most powerful may involve the 'correction' of errors in the data for particular individuals. This usually occurs at a stage which precedes the formal analysis and includes editing, pre-processing and imputation. To the extent that these operations take place before the data tape is produced, they must be treated as part of the data collection process.

At the next level, the assessment and evaluation of data constitutes a hurdle which must be negotiated by the data before certain types of analysis are sanctioned as permissible or justified. Much of the demographic data evaluation falls in this category and includes both checking the plausibility of the data against demographic axioms and substantive models and comparing the survey results with external sources of data. The WFS has initiated a continuous programme for evaluating the data from each country survey as soon as possible after the publication of the First Country Report. Six of the detailed country-specific evaluations have already been published: Nepal (Goldman, Coale and Weinstein 1979), Fiji (Potter 1975), Dominican Republic (Guzmán 1980), Sri Lanka (Alam and Cleland 1981) and two studies of Colombia (Hobcraft 1980; Flórez and Goldman 1980). A study of Bangladesh (Brass) is forthcoming and one of Pakistan by Booth and Alam is available in draft form. A study of Jamaica (Singh) is in press. In addition, several substantive analyses have incorporated sections on data quality (eg Somoza 1980; and Trussell 1980). A general assessment of the quality of WFS demographic data is presented in Chidambaram, Cleland and Verma (1980), based largely on the analyses listed above. The data from the household surveys and the detailed individual interview are compared in order to assess the two different methodologies in terms of their apparent data quality. The data from the individual surveys are evaluated both by examining their internal consistency and, where possible, by comparison with external sources.

The review concludes that in most cases the data are of surprisingly high quality. Although for some countries there is evidence of omission of births and displacement of dates of birth and dates of marriage, these errors appear to be restricted to the older women. In addition, there seems to be little evidence of substantial omission of infant and child deaths in the maternity histories. Hence, for most countries, errors of omission and displacement of vital events do not seriously distort the data, and permit estimates of levels and trends in age at marriage, fertility, and infant mortality for most of the sample.

The methods of data assessment described above do not provide an exhaustive evaluation of the data. Even if the data satisfy the internal and external checks carried out by demographers, measurement errors may have serious implications for further analysis. The potential impact of these errors depends on two factors: (1) the nature of the errors, and (2) the type of analysis being carried out. In order to examine the quality of the data in more detail, however, modifications or additions to the data collection process must be introduced. It is also necessary to formulate an explicit model of the response process and to express the individual response in terms of its component parts. This analysis of data quality is essentially statistical rather than demographic, but it has practical and substantive relevance to demographic analysis.

2 Approaches to Response Error Estimation in WFS

In defining the concept of error it is necessary to define a 'true value' for each individual in the population. This true value must be independent of the conditions under which the survey takes place, which can effect the individual's response. The concept of *individual true value* of a variable for a population element was developed by Hansen, Hurwitz and Madow (1953) as follows:

1 The true value must be *uniquely* defined.
2 The true value must be defined in such a manner that the purposes of the survey are met.
3 Where it is possible to do so consistently with the first two criteria, the true value should be defined in terms of operations which can actually be carried through (even though it might be difficult or expensive to perform the operations).

It is possible to define the true value in such a way that there are no response errors. A respondent's age could be defined as the answer given to the question 'What is your age?'. Similarly a respondent's attitude to the reintroduction of the death penalty could be defined as the answer he gives to the question 'Do you think the death penalty should be reintroduced?'. Both definitions satisfy the first and third criteria: the response is unique and it is defined in terms of operations which can be carried through. However, it is probable that they do not satisfy the survey objectives in such a way that such 'true values', dependent as they are on the specific conditions obtaining at the interview, would be acceptable as an ideal, although they might be acceptable as approximations to the true value. The individual true value should be seen as a characteristic which is independent of the survey conditions which affect the individual response. Age, for example, is defined as a time interval between two events, and this definition is independent of the method by which, and the conditions under which, we determine or observe the individual's age. For some other variables, such as income, the true value may be easy to define but difficult to obtain. For attitudinal items even the definition of the true value may be obscure. In all cases however the individual true value is a useful ideal at which to aim and the consideration of departures from this value is helpful in assessing the methods by which we obtain information.

The *individual response error* is the difference between the true value for the individual and the observation recorded. For example, if for a respondent born on 16 January 1946, age is recorded on 16 January 1981 as 30 years, the individual response error would be five years. The individual response is defined as the value obtained on a particular observation. Under different conditions (with a different interviewer or with a different form of question, for instance), a different individual response might be obtained.

The basic approach to the analysis of the individual response errors depends on an understanding of the survey process and the way in which the conditions under which the survey is carried out may affect the results of the survey. It is useful to distinguish between two components of the response error. The distinction is based on the definition of some of the characteristics of a survey as the *essential survey conditions*: for example, the subject matter, the data collection and recording methods, the timing and sponsorship, the type or class of interviewers and coders to be used in an interview survey, etc, can be considered as essential parts of the survey design. The expected value under these conditions can be defined. The difference between this value and the true value is the *response bias*, either for the individual or for a group of individuals. In addition to this there are 'random' fluctuations about the expected value. The particular interviewers chosen from the designated class of interviewers, the particular coders and transient characteristics of the observation situation are sources of such fluctuations. These variable errors also contribute to the response error, in the form of *response variance*. In order to appreciate the meaning of response variance, it is necessary to postulate that a survey is conceptually repeatable under identical conditions, the essential survey conditions. A survey is then seen as a single observation from a set of possible observations. The response variance is a measure of the variability between these observations.

The response bias and response variance differ also in implications for the survey analyst. First, the bias term is a constant which cannot be measured from within the survey; it is necessary to have data from some other source in order to assess it. On the other hand, the different components of response variance can in principle be estimated from the survey observations themselves. Secondly, the effect of the response bias is fixed regardless of the number of observations taken. By definition, even a complete enumeration would have the same response bias under the same essential survey conditions. However, the effect of response variance can be changed by sampling a larger number of the units involved. By increasing the number of interviewers, for example, the interviewer variance can be reduced. Thirdly, response bias is of particular concern in the estimation of means and totals for the whole sample. For comparisons between means of subclasses and in the calculation of measures of correlation and association the effect may be slight. Response variance will not affect the expected value of estimators of means and totals but will contribute to their imprecision. In addition, response variance will tend to attenuate measures of relationship between variables. This will apply even to the comparison of two subclass means if the classifying characteristic is subject to measurement error.

It is important that the analysis of response errors should not lose relevance to real problems due to the sophistication of the mathematical approach. The basic objective of a survey is to provide data on the basis of which the survey variables can better be understood, described or predicted. The aim in the analysis of response errors should therefore be to maximize the information which can be abstracted from the data. In the context of the WFS, methodological experimentation is by and large excluded by the very nature of the operation. The primary objective has been to assist countries in obtaining the best possible data from a single operation, which necessarily requires the choice of a study design considered a priori to be the most suitable. Thus it is not possible in general to compare different survey procedures in order to ascertain which is superior. Furthermore for the data collected in WFS surveys, there is no source of external validation data available at the level of the individual respondent. Consequently the analysis of response errors must be based on an examination of the internal consistency of the data. The emphasis therefore is on the *reliability* of the data, rather than on its validity; in other words, on response variance rather than on response bias.

A severe constraint is therefore imposed on any investigation of response errors in WFS surveys. It is not possible to interfere with the principles laid down for the conduct of the survey by introducing any new or experimental procedures which might reduce the quality of the data collected. Furthermore it is not possible, given the absence of an external source of validation data, to examine the absolute magnitudes of the individual response errors. There are however two possible approaches which can provide some information on the magnitude and impact of the errors: *re-enumeration* and *interpenetration*.

The first approach involves re-interviewing at least some of the respondents in the main survey. The re-interviews should be carried out soon after the main survey under the same (or similar) essential survey conditions. This would provide two separate observations on each of these respondents.

Certain characteristics of the survey would be constant for the two observations: the subject matter, the questions asked, the field force, the procedures for the supervision and control of the fieldwork, the coding and processing of the questionnaires. Thus the data could provide no information on the effects of these conditions on the survey results. In order to assess the systematic impact of any or all of these factors, either some source of information outside the survey procedure or an experimental design controlling these factors would be necessary.

Some factors would vary between the two surveys, however. The transient situational factors certainly vary, the two interviews being conducted on different occasions in every case. In addition, two different interviewers would be used for each individual and thus a part of the difference between the observations might be due to differences between the interviewers. The same would be true of the coding and processing, although the allocation of schedules to coders might not be conducted as rigorously as the allocation of respondents to interviewers.

In essence, therefore, such data would not, and could not, provide any information on response bias. Without external validation data, no assessment can be made of any *systematic* distortion of the observations produced by the conduct of the survey. What they would provide is an opportunity to examine the *reliability* of the measurements, the extent to which the application of the same essential survey conditions on two occasions would produce different results. Thus, they would afford us an opportunity to partition the variability observed in the survey observations into two components, one due to the inherent variability in the variable being measured, the other introduced into the recorded responses by the observation process itself.

The second approach involves a modification of the survey design. It has been established in other contexts that interviewers may influence in a systematic way the responses they obtain. If this is so for WFS surveys, then the estimates of variability obtained in the usual way for statistics calculated from the sample observations may seriously underestimate the true variance. This component of variance – the *correlated response variance* due to interviewers – will be present in any statistics calculated from the survey data, but the difficulty in practice is that there is usually no way of estimating it. The problem arises because respondents are usually allocated purposively (or haphazardly) to interviewers and any differences between the results obtained by different interviewers may be due to differences between the individuals whom they interview rather than to differences between the influence of the interviewers themselves. It is possible, however, to modify the survey execution in such a way that this component of variance is estimable. The basic feature of the design is that the respondents must be allocated *randomly* to interviewers, so that no systematic difference between the workloads of the interviewers can contaminate the comparison of the results of the interviewers. There will of course be differences between the workloads, but as long as the allocation of respondents to interviewers is random, these differences can be taken into account in the analysis. This procedure of random allocation of workloads is called *interpenetration.*

It is obviously impossible in practice to allocate a random subsample of a national sample to each interviewer. Not only would the cost of such an operation be enormous, but the disruption of the field execution of the survey would make it unacceptable in terms of the WFS objectives. However, the field strategy of the WFS lends itself to a modification of the design which is equally satisfactory. In the field, interviewers work in teams, a team usually consisting of four to six interviewers and two supervisors responsible for organizational supervision and timely scrutiny of interviewers' work. Each team works and travels as a unit. The allocation of work to the interviewers is normally the responsibility of the supervisors. The supervisors have, for each area, a list of the individuals (or in some cases, households) to be interviewed. It would obviously be a straightforward matter to determine the allocation of respondents to interviewers before the fieldwork in such a way that each interviewer is allocated, in effect, a random subsample of the work in that area.

Thus, without any significant interference with the procedures of data collection, it would be possible to modify the execution of the survey so that the contribution of the correlated response variance due to interviewers could be estimated and its impact on the survey results assessed.

The basic approach of this project thus involves two elements:

1 Re-enumeration A subsample of the respondents in the main survey will be re-interviewed under the same (or similar) essential survey conditions. This will permit the partitioning of the observed variability of the responses into two components: the sampling variance and the simple response variance. It will also make it possible to examine in detail the extent to which the same individuals (the respondents) give identical (or different) answers to the same questions on different occasions.

2 Interpenetration By allocating the interviewers' workloads randomly within teams, it will be possible to estimate the extent to which the usual estimates of variance underestimate the true variance, and thus to provide a more valid estimate of the total variance of the survey.

The particular design used in the project combines the two procedures of interpenetration and re-enumeration in a way which permits the estimation of some additional parameters of the response errors. The technical aspects of the design, suggested in a paper by Fellegi (1964), are described briefly in chapter 4. The practical features are discussed in chapter 3.

3 The Structure and Design of the Project

The WFS programme provides a unique opportunity for a cross-national study of response error. The diversity of the WFS participant countries permits the selection of study areas covering a considerable regional, developmental, ethnic and cultural spread, while the standardization of the survey instruments and the centralization of technical control make it possible to achieve meaningful comparison between the national studies. Moreover, the main surveys are already financed and only the marginal costs of the operations and modifications directly attributable to the response errors project need to be found.

The main emphasis of the project is twofold. First, the observed variability of the results is to be partitioned into the components representing *sampling variance* (sampling error) and the *simple response variance*; and secondly, the magnitude of the *correlated response variance* due to the interviewers is to be estimated and its impact assessed (this component is frequently known as *interviewer variance* or *interviewer effect*).

The project also provides an opportunity to gain some insight into the way in which the individual responses vary from one occasion to another for the individual. The cross-tabulation of the responses for the main survey interview and the re-interview will provide a detailed picture of the *stability* of the responses. By introducing a third (reconciliation) interview in cases where the observed discrepancies are large, a measure (defective but not negligible) of the *validity* of the responses and some indication of the sources of the errors may be obtained.

It was decided that the project should be carried out in four countries representing a wide range of developmental levels, each selected from a different broad cultural area. The four countries originally chosen were Lesotho, in Africa; Peru, in South America; Turkey, in the Middle East; and Philippines, in Asia. Due to difficulties in implementation, Burma was substituted for Philippines and when the main Burma survey was subsequently postponed, the opportunity arose to carry out the project in Dominican Republic. The project was thus actually carried out in Lesotho, Peru, Turkey and Dominican Republic.

The main features of the project design were the same for each of the countries, although the detailed implementation differed in each case. The overall design consists of three stages. The first involves a modification of the design of the main survey. The other two stages are separate from and additional to the main survey.

Stage 1 The main survey as planned for the country forms the basis of the first stage. The only modification is that a subsample of the respondents is allocated randomly to each of the interviewers in the team in each location. The randomization is carried out at the survey headquarters before the fieldwork commences in that location. The cost of this stage arises from the additional clerical work involved in designating the subsample to be interviewed by each interviewer; and from the possible increase in the cost of the fieldwork due to the inflexibility of interviewer allocation, since the freedom of decision of the supervisor is restricted by specifying in advance which interviewer should carry out each interview.

Stage 2 The second stage consists of a re-interview with at least a subsample of the respondents from the main survey. The re-interviews should be carried out under the same essential survey conditions as the original interviews. Furthermore the same team should do the original interviews and the re-interviews in each location selected for the second stage. No interviewer should interview the same respondent twice; the re-interview should be carried out by another member of the same team. The questionnaire used in the re-interview should be the same as that used in the original interview in so far as this is possible.

Cost considerations demand that the re-interview survey should be carried out only for a subsample of the main survey. It is desirable that this subsample should be a probability sample of the main survey sample.

For the re-interview survey, the basic problem is that of timing. If the time period between the main survey and the re-interview is short, the responses will not be independent and may indeed be highly correlated. The first interview may affect the second response if, for example, the respondent remembers the answers given in the first interview. This problem may be avoided, or at least reduced, by leaving a long time period between the two interviews, but this would lead to two further problems, namely that genuine changes may occur in the interval, which will exaggerate apparent discrepancies between the two responses, and that the interviewers used in the main survey may no longer be available. On balance the advantage lies with choosing a short interval between the two stages, particularly because of the need to have the same interviewers in both surveys.

Stage 3 This stage is designed to obtain an understanding of the way in which response errors arise for particular questions. Where the observed discrepancies between the first and second responses are large, a third (reconciliation) interview will be carried out to identify the source of the discrepancy and the 'true value', wherever possible. These reconciliation interviews must be carried out by a *female supervisor* or *female field editor*. There are two possible strategies: (i) to have additional supervisory staff travel with the team of interviewers during stage 2 who will carry out the reconciliation interviews as soon as the re-interview responses are compared with the original responses; and (ii) to do the editing and comparison of the two completed interview schedules at headquarters and to mount a separate and smaller field operation to carry out the reconciliation interviews.

Stages 2 and 3 must be handled very carefully. For the re-interview, the interviewer must make it clear that a special study is being undertaken and that the respondents have been chosen at random for another interview which will be shorter than the first. The diplomatic aspect will become even more important during the third (reconciliation) interview, in which inconsistencies are clarified. This is why a supervisor or editor must carry out the third interview. She must stress that no one doubts the respondent but that for the special study it is necessary to find out why different answers were obtained.

Sections 3.1 to 3.4 below describe the procedures used in each of the four countries included in the project.

3.1 PERU

The Peru Fertility Survey, conducted by the National Statistics Office during 1977-8, was based on a three area-stage national probability sample. Districts (of which there are around 1700 in the country) formed the primary sampling units (PSUs). In all, 124 PSUs were selected, 57 self-representing, appearing in the sample with certainty, and 67 non-self-representing, selected with probability proportional to size. In urban areas *blocks*, and in rural areas *localities* constituted the second-stage units (SSUs). Generally, an SSU consisted of 25-100 dwelling units, and a total of 1424 SSUs were selected with probability proportional to size. The third sampling stage involved the systematic selection of dwellings from within the selected SSUs, yielding a self-weighting sample (except that jungle areas were oversampled by a factor of 4). All ever-married women aged 15-49 residing (on a *de facto* basis) in the 8330 sample dwellings were eligible to be interviewed in detail regarding their maternity and marriage histories, knowledge and use of contraception, fertility intentions and preferences and socio-economic background. In all 5640 individual female interviews were successfully completed, representing a response rate of around 90 per cent.

Fieldwork for the main survey was to be conducted by 36 female interviewers divided into six teams, each team working under one supervisor and one field editor. It was necessary to use five different languages or dialects for interviewing: Spanish; Aymara; and three Quechua dialects, Ancash, Ayacucho and Cuzco.

Arrangements were made in the main survey for the interpenetration (randomization) of the interviewer workloads within teams for the secondary sampling units (SSUs) selected for the response errors project, the designated SSUs. For each SSU a folder was prepared containing the basic information about the SSU and listing the selected households. Each team was given a set of these folders before going into the field. For each SSU a decision was taken as to how many interviewers should be sent to the SSU. At least two interviewers were to travel to each SSU and the interviews were to be allocated randomly between them; the maximum number of interviewers in a team was seven.

In urban PSUs, particularly Lima, the allocation of interviewers to households was carried out over the whole designated sample. If, for example, a PSU contained five SSUs and the team contained seven interviewers, the letters A to G were allocated to each successive set of seven households in the PSU as follows:

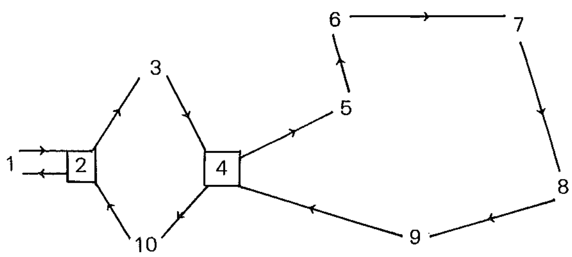
SSU 1	9 households	A B C D E F G A B
SSU 2	10 households	C D E F G A B C D E
SSU 3	7 households	F G A B C D E
SSU 4	15 households	F G A B C D E F G A B C D E F
SSU 5	4 households	G A B C

Each interviewer was allocated randomly one of the letters A to G and the households bearing that letter constituted the interviewer's workload.

In rural areas and where the number of households in a group of designated SSUs was too small, an appropriate subset of the letters A to G was to be used, eg if there were only four households, the letters A, B, C, D, or A, B, A, B would be allocated. Each letter would identify one of the interviewers sent to the SSU.

Approximately one in four of the main survey SSUs (urban blocks and rural localities) were designated for the Response Errors Study (RES). The RES consisted of conducting a *re-interview* (R2) with all respondents in the designated SSUs, using a shortened but otherwise identical version of the original (R1) questionnaire. Following this, the completed questionnaires for the first and the second interviews were compared by the field editors, and in cases where major inconsistencies occurred a third, *reconciliation*, interview was carried out to ascertain the 'true' response and also the cause of the discrepancy.

In Lima the designated SSUs were selected at random and the re-interview involved a separate trip to the selected areas. Outside Lima, owing to more difficult travel, the sample was selected purposively, and fieldwork logistics were planned such that while covering a group of neighbouring SSUs for the original interview the team would pass through the designated cluster(s) twice, with an interval of 1-2 months between the two visits. The diagram below illustrates the principle. Starting from SSU 1 (say, a district centre), a team conducts the first interview in 10 clusters and conducts re-interviews in the purposively selected clusters 2 and 4 during its return trip.



A rotating system of allocating workloads to interviewers for the first and second interviews was devised and is given below. The allocation is presented for the maximum team size of seven interviewers and for each subset of interviewers.

If, for any reason, any interviewer should fail to carry out any part of her workload, or if an interviewer should complete an interview allocated to another, this fact, the reasons for it, and the names and numbers of both interviewers should be recorded by the supervisor and reported to headquarters at the end of the fieldwork.

The questionnaire for the main survey in Peru incorporated the WFS core questionnaire and the fertility regulation module. The questionnaire for the re-interview was

Rotating System of Allocation of Workloads

No of interviewers in the area	2	3	4	5	6	7
Interviewers for main survey	A B	A B C	A B C D	A B C D E	A B C D E F	A B C D E F G
Interviewers for re-interview	B A	C A B	D C B A	E C D B A	F E D C B A	G F D E C B A

shorter but all the questions included had already been asked during the original interview.

In Lima the questionnaires from the main survey were edited and coded in the survey headquarters before the re-interviews were carried out. The completed questionnaires were *not* shown to any of the interviewers before the re-interviews. The completed questionnaires for the two interviews were compared by the editors for all the questions asked in the re-interview. When inconsistencies were found, a reconciliation interview was carried out to ascertain, if possible, the cause of the discrepancy. In rural areas, the completed questionnaires were kept in the custody of the supervisor/editor and the reconciliation interview was carried out before the team left the SSU.

Implementation

The study design required that (1) at least two interviewers travel together to an SSU with interviews within the cluster allocated randomly between the interviewers, and that (2) re-interviewing in the cluster should be done by the same team, following a predetermined random allocation such that no respondent is interviewed twice by the same interviewer. For several reasons the pattern of interview allocation diverged rather substantially from that planned. The primary reason was the disruption of the implementation of the main survey, due to climatic and budgeting problems, resulting in the fieldwork being stretched over a very long period. Consequently the time elapsed between the two interviews also tended to be lengthened: while 60 per cent of the re-interviews were conducted within three months of the original interview, the time elapsed exceeded six months for nearly 30 per cent. This made it difficult in practice to follow the above-mentioned allocation rules. Further, urban and rural areas differed greatly (not unexpectedly) in relation to the elapsed time: nearly 80 per cent of the re-interviews in urban areas but only 30 per cent of those in rural areas were conducted within three months of the original interview; the interval exceeded six months for only 5 per cent in urban areas, but for nearly 70 per cent in rural areas (there being very few re-interviews in rural areas between the fourth and sixth months). This disrupted the plan to conduct re-interviews in rural areas during the return trip. It is noteworthy, nevertheless, that an overall response rate of around 85 per cent was achieved in the re-interview survey.

Another difficulty resulted from the rather small sample take per SSU (an average of around four, not infrequently only one or two interviews per cluster). It was not always possible to send two interviewers to each cluster.

Though an attempt was made to achieve a reasonable geographical spread in purposively designating re-interview areas, the resulting re-interview sample none the less differed significantly in composition from the main survey sample. There was an over-representation in the former of urban areas, as well as of better educated women. Since both these characteristics are likely to be strongly related to response errors, it was necessary to weight the re-interview sample so that its *joint* distribution by city size (four categories) and woman's level of education (five categories) agreed with the main survey sample. The range of weights introduced was around 1–5.

The fact that – due to interruptions and practical difficulties beyond the control of the survey organizers – the pattern of re-interview allocation diverged substantially from that planned has considerable implications in terms of analysis of the data, particularly the study of interviewer effect. Any realistic model to be fitted to the data will now be considerably more complex than originally intended.

3.2 LESOTHO

The Lesotho Fertility Survey, conducted by the Central Bureau of Statistics in 1977-8, was based on a national two-stage probability sample. Census enumeration areas (of which there are 1066 in the country) were the primary sampling units (PSUs). One hundred PSUs were selected with probabilities proportional to size and a sample of households was selected within each selected PSU such that each household in the population had an equal probability of selection. The PSUs were stratified by ecological zone, population density and size before the first stage of selection. All ever-married women residing (on a *de facto* basis) in the selected households were eligible for interview. In all, 3603 individual interviews were successfully completed, giving an overall response rate of around 88 per cent.

Fieldwork for the main survey was carried out by eight teams of interviewers, each consisting of either four or five interviewers, one supervisor and one field editor. In all, 34 interviewers worked on the survey. The language in which the interviews were conducted was Sesotho. The questionnaire itself was also in Sesotho, although the interviewers' instruction manuals were in English.

Arrangements were made for the interpenetration (randomization) of the interviewer workloads within the teams for every PSU in the sample. For each PSU, the selected sample of households was listed, village by village, in the order in which the fieldwork was to be carried out. The numbers 1 to 5 (for teams with five interviewers) or 1 to 4 (for teams of four interviewers) were allocated to each successive set of five or four households on the list. For each of the numbers, a separate list of the households with that number was written out. For each team one of the lists was allocated at random to each interviewer in the team, before the fieldwork began. The supervisor received the master list and the set of interviewer lists for each cluster (PSU) in the team's work allocation, together with a list giving the allocation of workloads to interviewers. The supervisor was given the responsibility of ensuring that each interviewer carried out all her own workload.

In deciding on the subsample to be selected for the re-interview survey, two alternative strategies were considered. The first strategy was to use all eight teams in the re-interview survey and to have each team complete a part of its previous work allocation. The difficulty with this approach, however, lay in the fact that each team required a vehicle to carry out its fieldwork and vehicles were available for all teams only for the period of the main fieldwork (August-October 1977). Only three vehicles could be obtained for the period October-December 1977. Thus the second strategy was adopted; three teams were chosen for the re-interview survey and each of these was allocated two-thirds of the PSUs in which it had worked in the main survey. Each team was assigned an additional female field editor for the re-interview survey. These field editors were chosen from those who had worked in other teams in the main survey.

The system of allocation of workloads to interviewers in the re-interview survey is given below. The allocation is given for teams of four and five interviewers.

No of interviewers in the team	4	5
Interviewers for main survey	1 2 3 4	1 2 3 4 5
Interviewers for re-interview	4 3 2 1	5 4 2 3 1

Where interviews were, for any reason, re-assigned for the original interview, the allocation for the re-interview was re-assigned accordingly.

The questionnaire for the main survey in Lesotho incorporated the WFS core questionnaire and two modules. The questionnaire for the re-interview was shorter, consisting of sections 1, 2, 3 and 5 of the core questionnaire. All the questions asked in the re-interview had already been asked, in exactly the same form, in the original interviews.

The questionnaires from the original interviews were edited and coded in advance of the re-interview survey. The re-interviews were edited and coded in the field by the supervisory staff and were compared by the editors for a subset of the questions. Where discrepancies were found between the answers to these questions, a third, reconciliation interview was to be carried out by one of the female supervisory staff. These reconciliation interviews were to be carried out before the team left the area. The editors were instructed to use a clean questionnaire, fill in the identification information, and mark the questions to be reconciled (ie the questions where inconsistencies were found). A special summary form was prepared for each reconciliation interview.

Implementation

In Lesotho the execution of the project design in the field conformed closely to that described above. The fieldwork for the main survey lasted from August to early October 1977. The re-interview survey commenced in late October and was completed in December 1977. One of the interviewers left the field staff between the two field operations and was replaced by an interviewer from one of the teams not involved in the re-interviews. The time interval between the two interviews varied between one month and four months.

Twenty-five PSUs were included in the re-interview survey and a total of 724 interviews were obtained from the 867 individuals, a response rate of 84 per cent. The system of allocation of workloads to interviewers in the two field operations was implemented satisfactorily.

One additional benefit obtained from the response errors project may be noted here. On examining the field records for the main survey, there appeared to be too many cases where the code 'dwelling vacant' had been obtained as the final response category. Since it seemed possible that this code had been misunderstood by the interviewers, it was decided to check the dwellings with this code in a number of PSUs during the re-interview survey fieldwork. Of a total of 62 such cases in the 15 PSUs which were checked, 26 (or 42 per cent) produced completed interviews. These cases will provide both an opportunity to improve the data from the main survey and an indication of the possible impact of such non-response on the results of the main survey.

3.3 TURKEY

The Turkish Fertility Study was undertaken by the Institute of Population Studies, Hacettepe University, and the fieldwork was carried out between September and December 1978. The sample was a national probability sample based on a stratified multi-stage area design. The primary sampling units were stratified by population size and by geographical regions defined in terms of a number of spatial and socio-economic variables. In the urban part of the sample, localities, then wards, and in the third stage blocks were selected systematically with probability proportional to size (PPS); within sample blocks, small segments of five dwelling units each were selected so as to yield a self-weighting sample. In the rural part, villages were selected systematically with constant probability within each size stratum; this was followed by equal probability selection of segments of households to provide, again, a self-weighting sample.

Within sample segments, all households were enumerated using a household schedule in which usual residents were listed and data on a number of demographic and socio-economic items obtained. This was followed by the detailed individual interview of ever-married women aged 15-49 on a *de jure* basis.

The sample was originally designed for the Turkish Demographic Survey of the State Institute of Statistics, Turkey, in May 1976. Since data from the 1975 Population Census were not available at the time, the 1970 Census data formed the basis for the selection of primary sampling units. The frame within the sample PSUs was updated in December 1976. The TFS is based on the same clusters as the Turkish Demographic Surveys but on different segments of households.

About 4500 individual interviews were successfully completed, representing an overall response rate of about 85 per cent.

Fieldwork for the main survey was conducted by eleven teams, each consisting of either four or five interviewers, one supervisor and one field editor. It was considered impractical to randomize the allocation of respondents to interviewers for the whole of the sample for the main survey. The sample areas had already been divided into 33 groups of areas for reasons associated with fieldwork organization. Each team was allocated three groups of areas. One area from each group was chosen for the response errors project. Half of the areas chosen were urban, and half were rural. Each field trip for each of the eleven teams was scheduled to start from an area designated for the project. Four segments, each containing five to six households, were selected from each area.

A pair of interviewers was allocated to each segment in the designated area, as follows:

Main survey

Segment	1	2	3	4
Interviewers	A B	C D	A B	C D

Within each segment the workload was split randomly between the two interviewers. Sample lists and the allocation of respondents to interviewers were determined in advance of the fieldwork for the designated areas. In the cases where there were more than four interviewers in a team, four were chosen at random for this part of the work.

At the end of each field trip, the team returned to the designated area from which the trip had commenced. The re-interviews were carried out at this stage, with the interviewers in each pair exchanging the workloads from the main survey. Thus the pattern of allocation was:

Re-interview survey

Segment	1	2	3	4
Interviewers	B A	D C	B A	D C

Since the field procedures differed substantially for the designated areas from those used in the rest of the sample, the field directors were concerned that both the areas to be used in the re-interview survey and the interviewers who would be employed would be known in advance to the field staff and that this knowledge might influence the interviewers' behaviour and consequently distort the results of the project. Consequently an additional area in each of the 33 groups of areas was selected and treated in a similar manner in terms of field procedure. Where the team consisted of four interviewers, the allocation was identical to that given above. If, however, there were five interviewers in the team, a different allocation was used. Thus for a team of five interviewers (A, B, C, D, E), if the allocation above was used for the designated areas, the allocation below might be used for the additional areas.

Main survey

Segment	1	2	3	4
Interviewers	A C	D E	A C	D E

If for any reason, one of the interviewers were to drop out, she was to be replaced for the re-interviews from within the team if possible, and if this were not possible, then one of the reserve interviewers was to be used. If, for example, interviewer A were to drop out after the main survey interviews in an area, and the replacement interviewer is denoted by X, then X should carry out the re-interviews allocated to interviewer A as follows:

Segment	1	2	3	4
Main survey	A B	C D	A B	C D
Re-interview survey	B X	D C	B X	D C

The questionnaire for the main survey consisted of the WFS core questionnaire and three additional modules. For the re-interviews, a shortened version of the main survey questionnaire was prepared which consisted mainly of the questions from sections 1, 2, 3 and 5 of the original questionnaire; the *question numbers and the questions* on the re-interview questionnaire were identical to those on the original questionnaire.

The completed questionnaires from the main survey were coded in the headquarters office and the coded forms were brought to the field by the Institute's headquarters staff who then supervised the re-interview survey and the reconciliation interviews. The period between the first and second interviews was expected to be between 14 and 21 days for each area.

In deciding whether a reconciliation interview was necessary, a set of criteria was set up in consultation with the national staff. A special form, on which three sets of responses could be coded, was designed with spaces for all the questions on the re-interview questionnaire. The responses from the first interview were coded on to this form. The re-interview responses were coded on to the same form by the supervisor or field editor immediately after the re-interview. If a reconciliation interview was necessary, the remainder of the form was completed during that interview. The reconciliation interviews were conducted by the female field editor attached to each team.

Implementation

It had originally been planned to conduct re-interviews in a total of 600 households in the 132 designated segments (four segments in each of the 33 groups of areas). However, due to the low overall response rate for the main survey, this strategy would have produced only 380 interviews. Consequently the Project Director and his staff decided to attempt re-interviews in all the households in each segment of the 33 designated areas. A total of 657 respondents from the main survey were selected for re-interview. Of these 560 were successfully re-interviewed and the reconciliation interview was completed in all the 194 cases for whom it was required. The response rate for the re-interviews was 85 per cent.

It should be noted that the selection of the designated areas was *purposive* and not random; although every effort was made to make the areas as representative as possible, travel considerations were an important constraint and the final choice of areas was decided by the Turkish National Staff. It may be necessary to introduce weights into the analysis of response errors if the results are to be representative of the sample as a whole.

3.4 DOMINICAN REPUBLIC

The Dominican Republic was the first Latin-American country to participate in the programme of the World Fertility Survey. The survey was carried out between 1974 and 1976, the fieldwork taking place in 1975. In 1979, the Government of the Republic decided to embark on a second-round fertility survey, to be carried out in 1980. This survey, like that of 1975, was conducted by the National Council for Population and the Family (CONOPOFA), with the collaboration of the WFS. It was agreed that an investigation of response errors would be incorporated in the design of the second-round survey.

The sample design for the main survey (1980) was a multi-stage area probability sample. The population was divided into three zones:

Zone 1 rural areas, comprising about 50 per cent of the total population;
Zone 2 small cities, with 1970 population < 20 000, comprising about 10 per cent of the population;
Zone 3 larger cities, with 1970 population > 20 000.

Within zone 1, the 598 rural *secciones* were stratified explicitly by region and implicitly by province and *municipio*. Sixty-two *secciones* were selected with probability proportional to size. Villages and then *chunks* were subsequently selected within a *seccion* to give an equal probability sample of women.

In zone 2, 11 of the 83 cities were selected systematically with probability proportional to size, having ordered the list by region and province. *Blocks* were selected within cities to give an equal probability sample of women.

In zone 3, all 15 large cities were selected with certainty. Within each city, blocks were selected systematically with equal and constant probability.

Of the 5569 women identified for individual interview, 5137 interviews were successfully completed, a response rate at this stage of 92 per cent. In the urban domain, 2699 of 2921 intended interviews were obtained; in the rural domain, 2438 responses were obtained from a sample of 2648.

Each team consisted of four interviewers together with a supervisor and field editor. It was decided that the Ultimate Area Units (UAUs) were too small to permit randomization of the workload between all four members of the team. Consequently two interviewers were used in each UAU and the workload was divided randomly between them. It was considered desirable that the pairs of interviewers should not be kept unchanged throughout the period of the fieldwork, but that each interviewer should work with each of the other interviewers in the team at some stage. The strategy described below was devised to satisfy this requirement.

The total work assigned to a team was divided into a set of workloads, each accounting for about one week's fieldwork. For each of these workloads, each interviewer was assigned one of the letters A, B, C or D. During the execution of the workload, interviewers A and B worked together as did interviewers C and D. The rotating system of allocation is given below for one of the teams.

Workload no.	Interviewer's name			
	Marta	Anna	Carmen	Altagracia
1	A	B	C	D
2	A	C	B	D
3	A	D	C	B

The pattern is then repeated for workloads 4-6, etc.

For workload 1 Marta and Anna work together, as do Carmen and Altagracia. For workload 2 the pairings are Marta with Carmen and Anna with Altagracia. For workload 3 Marta and Altagracia share the work in a set of UAUs while Anna and Carmen do likewise. Thus each of the interviewers works with each of the other interviewers in a subset of the UAUs allocated to the team.

The re-interview survey was confined to the population of ever-married women. One in four of the UAUs was selected with equal probability from the sample for the main survey and re-interviews were attempted with all the eligible women in the subsample. All six teams were used in the re-interview survey. The re-interviews were assigned according to the general rule that interviewer A completed the re-interviews with the women who had been interviewed by interviewer B in the main survey; similarly interviewers C and D exchanged workloads for the re-interview survey. If any of the interviewers left the field staff during the course of the fieldwork, her respondents were to be divided equally between the other three interviewers in the team.

The fieldwork for the re-interview survey was a separate field operation from that for the main survey. The main fieldwork was carried out between February and April 1980; during May 1980 the documentation for the re-interview survey was prepared at the survey headquarters; and the fieldwork was executed in June 1980.

The questionnaire for the main survey consisted largely of the WFS core questionnaire, together with some additional questions. The questionnaire for the re-interviews was a shortened, but otherwise identical, version of the main questionnaire.

As the fieldwork for the main survey progressed, the completed questionnaires for the areas (UAUs) selected for the re-interview survey were identified and separated from the rest of the completed questionnaires. A form had been designed on which the responses to the questions on the re-interview schedule could be coded. The numbers on the columns of the form corresponded to the column numbers on the original questionnaire. All the questionnaires for the designated areas were coded on to these forms, one form being used for each questionnaire. The form also contained space for coding the responses to the re-interview and to the reconciliation interview, if required.

Blank questionnaires were prepared for the re-interview survey, each one bearing the necessary identification information to enable the interviewer to identify the correct respondent. The completed coding form from the original interview was kept in the custody of the supervisor and was *not* shown to the interviewer at any time. Upon completing the re-interview, the interviewer returned the completed questionnaire to the supervisor, who then coded the responses on to the appropriate part of the special coding form.

The supervisor/field editor then compared the responses to the two interviews and decided on the basis of the criteria provided whether a reconciliation interview was necessary. If it was considered necessary, the reconciliation interview was carried out by the supervisor and the results were coded directly on to the form.

Implementation

The execution of the project design in the field conformed closely to that described above. The fieldwork was completed on time in June 1980. The time interval between the two interviews for each individual varied between one month and four months.

A total of 936 re-interviews were completed successfully, 447 in the urban domain and 489 in the rural domain, giving a response rate of 80 per cent. The system of allocating workloads to interviewers in the two field operations was implemented satisfactorily. A detailed description of the project implementation will be given in the report on the analysis of the results.

4 Measures of Response Variability

In investigating the reliability of the data from WFS surveys, our attention can be focussed on two different but interrelated aspects of the data. For each individual j we have for each variable y the results of two separate observations $-y_{j1}$ and y_{j2}. The differences within and between the pairs of observations provide the raw material for the investigation; the interpenetrated subsamples allocated to the interviewers provide the basis for further analysis. There are two issues with which we will be concerned. The first objective of the analysis of variability is to examine the effect of these response deviations on the results derived from the survey, or, more importantly, on the *substantive conclusions* we would reach on the basis of the survey results. Secondly, the analysis is concerned with the identification, and assessment of the magnitude, of the deviations between the individual responses recorded in the two interviews.

The simplest illustration of the impact of the response deviations on the results could be provided by comparing the summary measures of the distributions obtained in the first and second interviews, treating each as a separate survey. Since the same individuals are observed in the two interviews, inconsistencies between the two distributions are the result of the response deviations, at least in so far as the essential survey conditions remained constant. However, there are two drawbacks to this approach. First, it would be impractical to carry out a full replication of all our analysis on the two data sets separately. Secondly, a direct comparison of the statistics computed from the two data sets would by itself provide estimates of precision based on only 1 degree of freedom.

Much more information on the underlying variability in the results can be obtained by looking at the individual response deviations. In comparing the responses obtained for a particular variable or individual, our approach will be constrained by two considerations.

First, the measures available to us will be different for different levels of measurement. For nominal scale variables, we will be restricted to frequency measures of agreement and disagreement; for ordinal scale variables, these measures can be modified to take into account partial disagreement, depending on the difference between the scale values; whereas for interval and ratio scale variables, and for binary variables, there is considerably more flexibility in describing the response deviations – in particular, we can partition the total variance into the components due to sampling and non-sampling error.

Secondly, the impact of different levels and kinds of response errors will vary depending on the analysis being undertaken. In particular, we will consider separately the effect of the response errors on univariate means and on measures of relationship between two or more variables.

A review of the sources of response errors and of models used to describe them may be found in O'Muircheartaigh (1977).

In examining the responses obtained on the two occasions for a particular variable, the data can be represented by the cross-classification of the two sets of responses. For interval level data, the number of possible values for some variables is too large to permit useful evaluation of this matrix although measures of crude agreement, and measures of agreement within k units (partial disagreement), obviously provide useful summary information.

The traditional approach in measuring reliability is to calculate the correlation (product-moment) between the two sets of observations, which provides a crude index of error. The design of this project, however, provides a great deal more information about the nature of the data and of the errors.

There are three conceptually distinct sources of variation in the results from a survey.

1 The variation among the true values for different individuals. These true values are the quantities of interest in the survey itself. The true value for each individual is fixed. The only variability to which the results would be subject if the true values were observed directly would arise from the fact that typically only a sample from the population is observed. The sampling variance of the estimator is the variance of the sampling distribution of the estimator and depends only on the sample design.

2 The value of an observation is determined not only by the true value for the individual but also by errors of measurement. The sources of these errors are many and their impact will vary considerably from one variable to another. We can specify a general form for the distribution of these errors even though we do not know all the particular influences which generate them. In order to specify such a distribution, we need to assume that a survey is conceptually repeatable. The distribution of the response errors is identifiable only from replications of the survey. If we can, however, identify particular potential sources of response errors it is possible to obtain a measure of their impact by appropriate design of the survey.

3 The third potential source of variation in the observations arises from possible interaction between the observation process and the true values of the individuals in the sample. Since in the theoretical treatment we assume that we are dealing with a simple random sample of individuals from the population, the only way in which this may come about is that the allocation of the sample elements among interviewers influences the response errors generated by the interviewer. Thus it is conceivable that the particular subsample allocated to one interviewer may affect the response errors within that interviewer's workload. The probability distribution of these components of error

is therefore over different samples and different allocations of the sample.

4.1 A GENERAL MODEL FOR RESPONSE ERRORS

The basic mathematical model for response errors is given below. The model is applicable to all variables which are measured on an interval scale, and also to binary variables. For simplicity the discussion is restricted to the estimation of the population mean.

A particular survey is regarded as a single trial, ie the survey is regarded as conceptually repeatable. An observation for the jth element in the survey for trial t is denoted by $y_{jt(s)}$ where j denotes the individual, t denotes the trial and s denotes the sample and its allocation.

The observation $y_{jt(s)}$ can be partitioned as follows:

$$\begin{aligned} y_{jt(s)} &= y_j + (y_{j(s)} - y_j) + \epsilon_{jt(s)} \\ &= y_j + \beta_{j(s)} + \epsilon_{jt(s)} \end{aligned} \quad (4.1)$$

where $\beta_{j(s)}$ is the *fixed response error* obtained for unit j in sample s

and $\epsilon_{jt(s)}$ is the *variable response error* obtained for unit j in sample s at trial t.

The model specified in (4.1) recognizes explicitly that the response error for unit j in sample s at trial t can be partitioned into two components. The first term ($\beta_{j(s)}$) is essentially a bias term since it does not vary from one trial to another: it is the expected value of the response error over all possible measurements of individual j when the sample s containing j is selected. In the most general formulation this fixed effect is allowed to depend on the sample s and its allocation. The second term ($\epsilon_{jt(s)}$) represents the deviation of the observation $y_{jt(s)}$ from the expected value $y_j + \beta_{j(s)}$. The model is completely specified once the probability distribution for the $\{\epsilon_{jt(s)}\}$ is specified. The value of $\epsilon_{jt(s)}$ may depend on j, the trial t and the sample s. The distribution of the $\{\epsilon_{jt(s)}\}$ is called the η-distribution. A very general specification of the η-distribution is given in (4.2)

$$\left.\begin{aligned} &E_\eta(\epsilon_{jt(s)}) = 0 \\ &V_\eta(\epsilon_{jt(s)}) = \sigma^2_{j(s)} \\ &\mathrm{Cov}_\eta(\epsilon_{jt(s)}, \epsilon_{j't(s)}) = \rho_{jj'}\,\sigma_{j(s)}\,\sigma_{j'(s)} \end{aligned}\right\} \quad (4.2)$$

The parameters of the model are the $\{\beta_{j(s)}\}$, $\{\rho_{jj'(s)}\}$ and $\{\sigma^2_{j(s)}\}$. Although it will obviously be impossible to estimate all these parameters, it is useful to present the results for the general case.

The objective of the survey is to estimate the population mean

$$\bar{y} = \sum_{j=1}^{N} y_j$$

In assessing the impact of response errors on the estimator, it would be possible in principle to study any sample design. The presentation below is confined to simple random sampling. The sample mean of the observations is

$$\bar{y}_{.t(s)} = \frac{1}{n} \Sigma_{j\epsilon s}\, y_{jt(s)}$$

Define $\Delta_j = y_j - \bar{y}$

The observation $y_{jt(s)}$ may now be written as

$$y_{jt(s)} = \bar{y} + \Delta_j + \beta_{j(s)} + \epsilon_{jt(s)}$$

Consequently

$$\bar{y}_{.t(s)} = \bar{y} + \bar{\Delta}_{.(s)} + \bar{\beta}_{.(s)} + \bar{\epsilon}_{.t(s)} \quad (4.3)$$

where

$$\bar{\Delta}_{.(s)} = \frac{1}{n} \Sigma_{j\epsilon s}\, \Delta_j$$

$$\bar{\beta}_{.(s)} = \frac{1}{n} \Sigma_{j\epsilon s}\, \beta_{j(s)}$$

$$\bar{\epsilon}_{.t(s)} = \frac{1}{n} \Sigma_{j\epsilon s}\, \epsilon_{jt(s)}$$

The difference (net deviation) between $\bar{y}_{.t(s)}$ and $\bar{y}$ is made up of three components: the sampling deviation, the fixed response deviation and the average variable response deviation.

The overall bias of the strategy is given by

$$\begin{aligned} E(\bar{y}_{.t(s)} - \bar{y}) &= E_p\, E_\eta(\bar{y}_{.t(s)} - \bar{y}) \\ &= E_p(\bar{\Delta}_{.(s)} + \bar{\beta}_{.(s)}) \\ &= E_p(\bar{\beta}_{.(s)}) \end{aligned}$$

Thus the overall bias is due only to the fixed response effects. This *response bias* is not estimable from the survey data themselves.

The mean squared error of $\bar{y}_{.t(s)}$ can be obtained from (4.3) and is

$$\begin{aligned} E_p\, E_\eta(\bar{y}_{.t(s)} - \bar{y})^2 &= V_p(\bar{\Delta}_{.(s)}) + V_p(\bar{\beta}_{.(s)}) \\ &+ E_p E_\eta(\bar{\epsilon}^2_{.t(s)}) \\ &+ 2\mathrm{cov}_{p\eta}(\bar{\Delta}_{.(s)}, \bar{\epsilon}_{.t(s)}) \\ &+ 2\mathrm{cov}_p(\bar{\Delta}_{.(s)}, \bar{\beta}_{.(s)}) + [E_p(\bar{\beta}_{.(s)})]^2 \end{aligned}$$

The total variance of $\bar{y}_{t(s)}$ is given by

$$\begin{aligned} E_p\, E_\eta\,[\bar{y}_{.t(s)} - E_p\, E_\eta(\bar{y}_{.t(s)})]^2 &= V_p(\bar{\Delta}_{.(s)}) \\ &+ V_p(\bar{\beta}_{.(s)}) + E_p\, E_\eta(\bar{\epsilon}^2_{.t(s)}) + 2\mathrm{cov}_{p\eta}(\bar{\Delta}_{.(s)}, \bar{\epsilon}_{.t(s)}) \\ &+ 2\mathrm{cov}_p(\bar{\Delta}_{.(s)}, \bar{\beta}_{.(s)}) \end{aligned} \quad (4.4)$$

The overall variance of the estimator can therefore be partitioned into four components, ie

overall variance

= sampling variance (of true values)

+ variability due to fixed response errors

+ variability due to variable response errors

+ covariance between sampling deviations and response effects.

The model whose total variance is given by (4.4) is the general model for the system. In the subsections below some special cases are considered.

Case I Simple Response Variance

The simplest situation is that in which the only distortion of the true values is a random disturbance term which is uncorrelated with the true values. This can be expressed by modifying the model (4.1) and (4.2), using the following simplifying assumptions:

$$\left.\begin{aligned} &\beta_{j(s)} = \beta, \text{ all j} \\ &V_\eta(\epsilon_{jt(s)}) = \sigma_j^2 = \sigma_\epsilon^2, \text{ all j} \\ &\text{Cov}_\eta(\epsilon_{jt(s)}\,\epsilon_{j't(s)}) = \rho_{jj'}\sigma_{j(s)}\sigma_{j'(s)} = 0 \text{ all j, j}' \end{aligned}\right\} \quad (4.5)$$

The model (4.1) now becomes

$$y_{jt} = y_j + \beta + \epsilon_{jt} \quad (4.6)$$

and the total variance (4.11) becomes

$$E_p E_\eta [\bar{y}_{.t} - E_p E_\eta(\bar{y}_{.t})]^2 = V_p(\bar{\Delta}_.) + E_p E_\eta(\bar{\epsilon}_{.t}^2)$$

$$= \left(1 - \frac{n-1}{N-1}\right)\frac{\sigma_y^2}{n} + \frac{\sigma_\epsilon^2}{n}$$

If the finite population correction is ignored, this gives

$$V_{p\eta}(\bar{y}_{.t}) = \frac{\sigma_y^2}{n} + \frac{\sigma_\epsilon^2}{n} \quad (4.7)$$

The first term in (4.7) is the *sampling variance* of the estimator; the second term is the *simple response variance.* We define the *index of inconsistency*

$$I = \frac{\sigma_\epsilon^2}{\sigma_y^2 + \sigma_\epsilon^2} \quad (4.8)$$

which measures the proportion of the total element variance which is due to response variability.

It is interesting to note that under this model, the index of inconsistency I is closely related to the correlation between successive observations on the same individual. Indeed

$$\underset{j}{\text{Corr}}(y_{j1}, y_{j2}) = \frac{\sqrt{\underset{j}{E}(y_{j1}\,y_{j2}) - \underset{j}{E}(y_{j1})\,E(y_{j2})}}{\sqrt{\underset{j}{E}(y_{j1} - \bar{y})^2\,\underset{j}{E}(y_{j2} - \bar{y})^2}}$$

$$= \frac{\sigma_y^2}{\sigma_y^2 + \sigma_\epsilon^2}$$

$$= 1 - I \quad (4.9)$$

In order to estimate σ_ϵ^2 we need to have at least two observations on each individual in the sample. We do not have any direct information on the values of the $\{\epsilon_{jt}\}$. However the set of differences $\{y_{j1} - y_{j2}\}$ provide for us the values of $\{\epsilon_{j1} - \epsilon_{j2}\}$. The variance of $(\epsilon_{j1} - \epsilon_{j2})$ can be estimated simply and is

$$\sigma_{\epsilon_{1,2}}^2 = \sigma_{\epsilon_1}^2 + \sigma_{\epsilon_2}^2 - 2\sigma_{\epsilon_1\epsilon_2}$$

If we assume, not unreasonably, that $\sigma_{\epsilon_1}^2 = \sigma_{\epsilon_2}^2 = \sigma_\epsilon^2$, this gives

$$\sigma_{\epsilon_{1,2}}^2 = 2\sigma_\epsilon^2(1 - \rho_{\epsilon_1\epsilon_2}) \quad (4.10)$$

We estimate σ_ϵ^2 by

$$\hat{\sigma}_\epsilon^2 = \tfrac{1}{2}\hat{\sigma}_{\epsilon_{1,2}}^2$$

The critical problem with this estimator is that there may be a correlation (usually positive in practice) between the response errors of the same individual on the two occasions; the respondent may for example remember some of the responses from the first interview, and tend to report the same answers in the re-interview. If the correlation is positive, $\hat{\sigma}_\epsilon^2$ underestimates the simple response variance in the survey by a factor of $(1 - \rho_{\epsilon_1\epsilon_2})$. The data may be used to investigate whether such a positive correlation is present by comparing the variance of the response deviations for different time intervals between the interviews.

Case II Correlated Interviewer Variance

The assumptions in case I are unrealistic in so far as the variable response errors $\{\epsilon_{jt(s)}\}$ are assumed to be independent of one another. There are various factors which make this assumption unlikely to be appropriate. In particular, each interviewer carries out a number of interviews and it may be expected that the responses obtained from each individual in the interviewer's workload may be influenced in a similar way by the interviewer, and that, in consequence, the response errors for these individuals may be correlated. The model can be specified in such a way that this can be taken into account. Each observation is now denoted by $y_{ijt(s)}$ where i denotes the interviewer. The assumptions are given below:

$$\left.\begin{aligned} &\beta_{ij(s)} = \beta \quad \text{all i, j} \\ &V_\eta(\epsilon_{ijt(s)}) = \sigma_\epsilon^2 \quad \text{all i, j} \\ &\text{Cov}_\eta(\epsilon_{ijt(s)}, \epsilon_{i'j't(s)}) = \begin{cases}\rho_1 & \text{if } i = i' \\ \rho_2 & \text{if } i \neq i'\end{cases} \end{aligned}\right\} \quad (4.11)$$

The total variance of $\bar{y}_{.t}$ under this model is

$$V_{p\eta}(\bar{y}_{.t}) = \frac{\sigma_y^2}{n} + \frac{\sigma_\epsilon^2}{n}[1 + \rho_1(m-1) + \rho_2 m(k-1)] \quad (4.12)$$

where m is the size of each interviewer's workload.

In order to estimate the correlated response variance due to the interviewers the survey design must be modified. The basic principles can most easily be illustrated in the case of a simple random sample. A simple random sample s of size $n = km$ is selected from the population; the sample is partitioned into k equal subsamples of size $m - s_1, s_2, \ldots, s_i, \ldots, s_k$. Each subsample is allocated to a single interviewer. The label (i,j) is used to indicate that individual j belongs to the workload of interviewer i.

The usual estimator of the variance of the sample mean is

$$\hat{V}(\bar{y}_{..t(s)}) = (1-f)\frac{1}{n(n-1)}\sum_{i=1}^{k}\sum_{j\epsilon s_i}(y_{ijt(s)} - \bar{y}_{..t(s)})^2$$

It can be shown that the expected value of this estimator is

$$E_p E_\eta(\hat{V}(\bar{y}_{..t(s)})) = (1-f)\left[\frac{\sigma_y^2}{n}\frac{N}{N-1} + \frac{1}{n(n-1)}\sigma_\epsilon^2\left\{(n-1) - \rho_1(m-1) - \rho_2(m)(k-1)\right\}\right]$$

If the sampling fraction is negligible, the bias of this estimator is given by

$$E_p E_\eta \hat{V}(\bar{y}_{..t(s)}) - V(\bar{y}_{..t(s)})$$
$$= \frac{1}{n-1}[-\rho_1(m-1)\sigma_\epsilon^2 - \rho_2 m(k-1)\sigma_\epsilon^2]$$

From the data we can calculate two linearly independent sums of squares:

1 the between-interviewers sum of squares, and
2 the within-interviewer sum of squares.

If we denote the mean between-interviewers sum of squares by C and the mean within-interviewer sum of squares by F, we can show that

$$\left.\begin{aligned} E_p E_\eta [C] &= \sigma_y^2 + \sigma_\epsilon^2[1 + \rho_1(m-1) - \rho_2 m] \\ \text{and} \\ E_p E_\eta [F] &= \sigma_y^2 + \sigma_\epsilon^2[1 - \rho_1] \end{aligned}\right\} \quad (4.13)$$

Hence $\frac{1}{m}$ [C − F] provides a possible estimator of $\rho_1\sigma_\epsilon^2$. In fact, under this model,

$$E[\tfrac{1}{m}[C-F]] = \sigma_\epsilon^2[\rho_1 - \rho_2]$$

It is usually recommended as an estimator of $\rho_1\sigma_\epsilon^2$ since ρ_2 can generally be assumed to be small. See, for example, Hansen, Hurwitz and Bershad (1961), Fellegi (1964) and Kish (1962).

It is worth noting that an almost unbiased estimator of the total simple variance $[(\sigma_y^2 + \sigma_\epsilon^2)]$ is given by $\frac{1}{m}$ (C − F) + F. In fact,

$$E_p E_\eta [\tfrac{1}{m}(C-F) + F] = \sigma_y^2 + \sigma_\epsilon^2(1-\rho_2)$$

Case III Simple and Correlated Response Variance

It is clear from cases I and II above that both re-enumeration and randomized allocation of respondents to interviewers are necessary if both the simple and correlated response variances are to be estimated. A more elaborate survey design is necessary in this case. The design used in this project is based on that described by Fellegi (1964).

1 A *simple random sample* of n = km units denoted by s, is selected *without replacement* from a population of N units.
2 The sample is partitioned, again at *random*, into k subsamples of m units each, denoted by $s_1, s_2, \ldots, s_k$.
3 Each subset is paired at *random* with another (different) subset so that if $(s_1, s_{q_1}), (s_2, s_{q_2}), \ldots, (s_k, s_{q_k})$ are the pairs then $q_1, \ldots, q_k$ exhaust the integers $1, \ldots, k$.
4 There are k interviewers, numbered from 1 to k and the k pairs of subsets are allocated at *random* to these. Denote by $(s_{i(1)}, s_{i(2)})$ the pair allocated to the i-th interviewer; $s_{i(1)}$ and $s_{i(2)}$ are respectively the first and second assignments of the i-th interviewer.

Steps (1), (2) and (3) define the experimental design. The randomization is artificial and the probability distribution associated with it can, in principle, be calculated without too much difficulty. Let Q denote the outcome obtained from steps (1), (2), (3) and (4). Let p(Q) denote the probability of obtaining the outcome Q.

5 Each interviewer interviews his first assignment and this constitutes the original survey.
6 Each interviewer interviews his second assignment and this constitutes the repeat survey.

Steps (5) and (6) define the measurement process. The η-distribution is introduced here.

Let $y_{ijt(Q)}$ be the observed value of unit j at trial t given that the outcome Q was obtained. If, given Q, $j \notin s_{i(t)}$ (t = 1,2,), $y_{ijt(Q)}$ need not be defined at all. The model is analogous to the model (4.11). Thus we write

$$y_{ijt(Q)} = y_j + \beta_{ijt(Q)} + \epsilon_{ijt(Q)} \quad (4.14)$$

The implications of the model (4.14) depend on the specification of the $\{\beta_{ijt(Q)}\}$ and the probability distribution of the $\{\epsilon_{ijt(Q)}\}$.

Once again we assume that the $\beta_{ijt(Q)}$ are constant. The probability distribution of the $\{\epsilon_{ijt(Q)}\}$ must also be specified. We make two simplifying assumptions:

$$\left.\begin{array}{l} E_\eta(\epsilon_{ijt(Q)}) = 0 \\ V_\eta(\epsilon_{ijt(Q)}) = \sigma^2_{\epsilon_t} \end{array}\right\} \qquad (4.15)$$

We must also define all the other second-order moments of the $\{\epsilon_{ijt(Q)}\}$. In general, using (4.14)

$$\mathrm{Cov}_\eta(\epsilon_{ijt(Q)}\,\epsilon_{i'j't'(Q)}) = \rho_{ijt,i'j't'}\,\sigma_{\epsilon_t}\,\sigma_{\epsilon_{t'}}$$

There are many ways in which the $\{\rho_{ijt,i'j't'}\}$ could be specified. The survey design however permits us to define easily the following correlation coefficients. Let i° be the interviewer who carries out the i-th interviewer's first workload in the re-interview survey. Then, $s_{i(1)} = s_{i^\circ(2)}$.

$$\rho_{ijt,i'j't'(Q)} = \begin{cases} \rho_{1t} & i' = i, j' \neq j \in s_{i(t)} \\ & \text{(same interviewer, different units, same trial)} \\ \rho_{2t} & i' \neq i, j \in s_{i(t)}, j' \in s_{i'(t)} \\ & \text{(different interviewers, different units, same trial)} \\ \rho_3 & i' = i^\circ, j' = j, t = 1\ t' = 2 \\ & \text{(same unit, different interviewers, different trials)} \\ \rho_4 & i' = i, j \in s_{i(1)}, j' \in_s s_{i'(2)}\ t = 1, t' = 2 \\ & \text{(same interviewer, different units, different trials)} \\ \rho_5 & i' = i^\circ, j' \neq j \in s_{i(1)}, t = 1, t' = 2 \\ & \text{(same subsample, different units, different trials)} \\ \rho_6 & i' \neq i, i^\circ, j \in s_{i(1)}, j \in s_{i'(2)}, t = 1, t' = 2 \\ & \text{(different interviewers, different units, different subsamples, different trials).} \end{cases} \qquad (4.16)$$

Note that ρ_{1t} and ρ_{2t} correspond to ρ_1 and ρ_2 in case II; and ρ_3 is equivalent to $\rho_{\epsilon 1,\epsilon 2}$ in case I.

There are eight correlation coefficients defined in (4.16). In addition to these we also wish to estimate $\sigma^2_{\epsilon 1}$, $\sigma^2_{\epsilon 2}$ and σ^2_y. Thus we have eleven parameters in all. From the data we can obtain seven linearly independent sums of squares. These are

1 For each of the two surveys
   (a) Between interviewers
   (b) Within interviewers

2 Between the two surveys
   (a) within sampling units
   (b) within subsamples, between interviewers
   (c) within interviewers, between subsamples (this sum of squares is linearly independent of the previous ones if and only if $k > 2$).

The corresponding mean squares are listed as follows:

$$C_t = \frac{m}{k-1}\sum_{i=1}^{k}(\bar{y}_{i.t(Q)} - \bar{y}_{..t(Q)})^2, \; t = 1,2$$

$$F_t = \frac{1}{k(m-1)}\sum_{i=1}^{k}\sum_{j\in s_{i(t)}}(y_{ijt(Q)} - y_{i.t(Q)})^2, \; t = 1,2$$

$$L = \frac{1}{2(km-1)}\sum_{i=1}^{k}\sum_{j\in s_{i(1)}}(y_{ij1(Q)} - y_{i^\circ j2(Q)} - \bar{y}_{..1(Q)} + \bar{y}_{..2(Q)})^2$$

$$M = \frac{m}{2(k-1)}\sum_{i=1}^{k}(\bar{y}_{i.1(Q)} - \bar{y}_{i^\circ.2(Q)} - \bar{y}_{..1(Q)} + \bar{y}_{..2(Q)})^2$$

$$P = \frac{m}{2(k-1)}\sum_{i=1}^{k}(\bar{y}_{i.1(Q)} - \bar{y}_{i.2(Q)} - \bar{y}_{..1(Q)} + \bar{y}_{..2(Q)})^2$$

If $k = 2$ the seven mean squares above are not linearly independent. In fact, in this case, $M + P = C_1 + C_2$

Fellegi (1964) suggests that, instead of using M, it may be more convenient to use

$$R = \frac{1}{2k(m-1)} \sum_{i=1}^{k} \sum_{j \epsilon s_i(1)} (y_{ij1} - y_{i^\circ j2} - \bar{y}_{i.1} + \bar{y}_{i^\circ.2})^2 = \frac{km-1}{k(m-1)} L - \frac{k-1}{k(m-1)} M$$

The design has provided us with seven linearly independent sums of squares to estimate eleven parameters. It is not therefore possible to obtain unbiased estimators for all of the parameters. However, for the surveys with which we are concerned here, some further simplifying assumptions can be made:

$$\sigma_{\epsilon_1}^2 = \sigma_{\epsilon_2}^2 = \sigma_\epsilon^2 \qquad \rho_{11} = \rho_{12} = \rho_1 \qquad \rho_{21} = \rho_{22} = \rho_2 \tag{4.17}$$

These assumptions correspond to an assertion that the essential survey conditions over the two surveys are the same. If (4.17) holds, then the expected values of the mean squares listed above are:

$$E_p E_\eta(C_1) = E_p E_\eta(C_2) = \sigma_y^2 + \sigma_\epsilon^2[1 + (m-1)\rho_1 - m(k-1)\rho_2]$$

$$E_p E_\eta(F_1) = E_p E_\eta(F_2) = \sigma_y^2 + \sigma_\epsilon^2[1 - \rho_1]$$

$$E_p E_\eta(L) = \sigma_\epsilon^2 \left[1 - \frac{m-1}{km-1}\rho_1 - \frac{m(k-1)}{km-1}\rho_2 - \rho_3 + \frac{m}{km-1}\rho_4 + \frac{m-1}{km-1}\rho_5 + \frac{m(k-2)}{km-1}\rho_6\right]$$

$$E_p E_\eta(M) = \sigma_\epsilon^2 \left[1 + (m-1)\rho_1 - m\rho_2 - \rho_3 + \frac{m}{k-1}\rho_4 - \frac{m-1}{k-1}\rho_5 + \frac{m(k-2)}{k-1}\rho_6\right]$$

$$E_p E_\eta(P) = \frac{k}{k-1}\sigma_y^2 + \sigma_\epsilon^2 \left[1 + (m-1)\rho_1 - m\rho_2 + \frac{1}{k-1}\rho_3 - m\rho_4 + \frac{m-1}{k-1}\rho_5 + \frac{m(k-2)}{k-1}\rho_6\right]$$

$$E_p E_\eta(R) = \sigma_\epsilon^2[1 - \rho_1 - \rho_3 + \rho_5]$$

In our analysis we use

$$C = \tfrac{1}{2}(C_1 + C_2) \text{ and } F \doteq \tfrac{1}{2}(F_1 + F_2).$$

We now have five linearly independent sums of squares to estimate eight parameters. The system of equations we have to solve may be written as:

$$E_p E_\eta \begin{bmatrix} C \\ F \\ L \\ R \\ P \end{bmatrix} = \begin{bmatrix} 1 & 1 & m-1 & -m & 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -\frac{m-1}{km-1} & -\frac{m(k-1)}{km-1} & -1 & \frac{m}{km-1} & \frac{m-1}{km-1} & \frac{m(k-2)}{km-1} \\ 0 & 1 & -1 & 0 & -1 & 0 & 1 & 0 \\ \frac{k}{k-1} & 1 & m-1 & -m & \frac{1}{k-1} & -m & \frac{m-1}{k-1} & \frac{m(k-2)}{k-1} \end{bmatrix} \begin{bmatrix} \sigma_y^2 \\ \sigma_\epsilon^2 \\ \rho_1\sigma_\epsilon^2 \\ \rho_2\sigma_\epsilon^2 \\ \rho_3\sigma_\epsilon^2 \\ \rho_4\sigma_\epsilon^2 \\ \rho_5\sigma_\epsilon^2 \\ \rho_6\sigma_\epsilon^2 \end{bmatrix}$$

It should be possible to find estimators of five of the eight parameters so that the biases in the estimators are in terms of the remaining three parameters. The most important parameters are σ_y^2, σ_ϵ^2 and $\rho_1\sigma_\epsilon^2$. If we choose to estimate these, however, this places constraints on the other parameters we can estimate. In particular, due to the structure of the coefficient matrix, the biases in the estimators of σ_y^2, σ_ϵ^2 and $\rho_1\sigma_\epsilon^2$ must be in terms of $\rho_2\sigma_\epsilon^2$ and $\rho_3\sigma_\epsilon^2$. This can be seen immediately from the first five columns of the transformed coefficient matrix given below. Consequently we must choose to estimate two of the remaining three parameters. Of these, $\rho_4\sigma_\epsilon^2$ was chosen because it may be expected to be large; $\rho_6\sigma_\epsilon^2$ was chosen on the grounds of convenience.

A sequence of elementary operations was carried out on the coefficient matrix to facilitate the estimation of the five chosen parameters. The system of equations can be rewritten as follows:

$$[T]\, E_p E_\eta \begin{bmatrix} C \\ F \\ L \\ P \\ R \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 1 & 0 & -1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \sigma_y^2 \\ \sigma_\epsilon^2 \\ \rho_1\sigma_\epsilon^2 \\ \rho_2\sigma_\epsilon^2 \\ \rho_3\sigma_\epsilon^2 \\ \rho_4\sigma_\epsilon^2 \\ \rho_5\sigma_\epsilon^2 \\ \rho_6\sigma_\epsilon^2 \end{bmatrix}$$

where

$$[T] = \begin{bmatrix} \frac{1}{m} & -\frac{1}{m} & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ -\frac{1}{m}\left(\frac{k-1}{k-2}\right) & \frac{1}{m} & \frac{km-1}{km}\left(\frac{k-1}{k-2}\right) & -\frac{1}{m}\ \frac{km-m-1}{k-2} & \frac{1}{km}\left(\frac{k-1}{k-2}\right) \\ \frac{1}{m} & -\frac{1}{m} & 0 & 1 & 0 \\ 0 & -\frac{1}{m} & -\frac{km-1}{km} & 1 & \frac{k-1}{km} \end{bmatrix}$$

Consequently the following five linearly independent combinations of parameters have unique unbiased estimators:

1 $(\rho_1 - \rho_2)\sigma_\epsilon^2$

2 $\sigma_y^2 + \rho_3\sigma_\epsilon^2 - \rho_5\sigma_\epsilon^2$

3 $-\rho_5\sigma_\epsilon^2 + \rho_6\sigma_\epsilon^2$

4 $\sigma_\epsilon^2\,[1 - \rho_2 - \rho_3 + \rho_5]$

5 $-\rho_4\sigma_\epsilon^2 + \rho_5\sigma_\epsilon^2$

The estimators for 1 to 5 are:

$$E_1 = \frac{1}{m}[C - F]$$

$$E_2 = F - R$$

$$E_3 = -\frac{1}{m}\left(\frac{k-1}{k-2}\right)C + \frac{1}{m}F + \frac{km-1}{km}\left(\frac{k-1}{k-2}\right)L - \frac{1}{m}\left(\frac{km-m-1}{k-2}\right)R + \frac{k-1}{km(k-2)}P$$

$$E_4 = \frac{1}{m}[C - F] + R$$

$$E_5 = -\frac{1}{m}F - \frac{km-1}{km}L + R + \frac{k-1}{km}P$$

On the assumption that ρ_2 and ρ_5 are negligibly small, and disregarding ρ_3 for the moment, we obtain the following estimators for the principal parameters of the system:

$$\hat{\sigma}_y^2 = F - R \qquad (4.18)$$

$$\hat{\sigma}_\epsilon^2 = R + \frac{1}{m}[C - F] \qquad (4.19)$$

$$\widehat{\rho_1\sigma_\epsilon^2} = \frac{1}{m}[C - F] \tag{4.20}$$

$$\widehat{\rho_4\sigma_\epsilon^2} = \frac{1}{m}F + \frac{km-1}{km}L + R + \frac{k-1}{km}P \tag{4.21}$$

$$\widehat{\sigma_y^2 + \sigma_\epsilon^2} = \frac{1}{m}[C - F] + F \tag{4.22}$$

and hence,

$$\hat{\rho}_1 = \frac{\frac{1}{m}[C-F]}{R + \frac{1}{m}[C-F]} \tag{4.23}$$

$$\hat{I} = \frac{\hat{\sigma}_\epsilon^2}{\hat{\sigma}_\epsilon^2 + \hat{\sigma}_y^2} = \frac{R + \frac{1}{m}[C-F]}{\frac{1}{m}[C-F] + F} \tag{4.24}$$

$$\text{and} \quad \hat{\rho}_1 I = \frac{\frac{1}{m}[C-F]}{\frac{1}{m}[C-F] + F} \tag{4.25}$$

Apart from the problem of ρ_3, the biases in the estimators derived from the data include only terms in ρ_2 and ρ_5, which may reasonably be assumed to be small. However in social surveys ρ_3 is not generally negligible, since it is a measure of the recall effect, and it is usually positive. Hence, ignoring terms in ρ_2 and ρ_5, the estimate of σ_y^2 has an expected value

$$E(\hat{\sigma}_y^2) = \sigma_y^2 + \rho_3\sigma_\epsilon^2$$

Similarly, the estimate of σ_ϵ^2 has an expected value

$$E(\hat{\sigma}_\epsilon^2) = \sigma_\epsilon^2 - \rho_3\sigma_\epsilon^2$$

Thus $\hat{\sigma}_y^2$ overestimates σ_y^2 and $\hat{\sigma}_\epsilon^2$ underestimates σ_ϵ^2. However the total simple variance $(\sigma_y^2 + \sigma_\epsilon^2)$ can be estimated with negligible bias using $\frac{1}{m}[C - F] + F$(4.22). It is only by introducing information external to the model presented here – for example, the time interval between two interviews – that any assessment of ρ_3 can be obtained.

Case IV Interaction between Sampling and Response Deviations

It may be postulated that an interviewer's workload will influence the response deviations obtained by the interviewer. If this is so, then the sampling deviations $\{\Delta_j\}$ may be correlated with the response deviations (the $\{\epsilon_{ijt}\}$). The structure of the model will thus be more complex and the expected values of the seven mean squares will be as follows:

$$E_{p\eta}\begin{bmatrix} C \\ F \\ L \\ R \\ P \end{bmatrix} = \begin{bmatrix} 1 & 1 & (m-1) & (m-1) & -m & 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -\frac{m}{km-1} & -\frac{m(k-1)}{km-1} & -1 & \frac{m}{km-1} & \frac{m-1}{km-1} & \frac{m(k-2)}{km-1} \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ \frac{k}{k-1} & 1 & \frac{k(m-1)}{k-1} & m-1 & -m & \frac{1}{k-1} & -m & \frac{m-1}{k-1} & \frac{m(k-2)}{k-1} \end{bmatrix} \begin{bmatrix} \sigma_y^2 \\ \sigma_\epsilon^2 \\ 2\alpha\sigma_y\sigma_\epsilon \\ \rho_1\sigma_\epsilon^2 \\ \rho_2\sigma_\epsilon^2 \\ \rho_3\sigma_\epsilon^2 \\ \rho_4\sigma_\epsilon^2 \\ \rho_5\sigma_\epsilon^2 \\ \rho_6\sigma_\epsilon^2 \end{bmatrix}$$

where $\alpha \sigma_y \sigma_\epsilon = E_{p\eta} [\Delta_j e_{ijt}]$

The set of transformations previously applied to the coefficients matrix in case III now gives:

$$[T]\ E_{p\eta} \begin{bmatrix} C \\ F \\ L \\ R \\ P \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 1 & 1 & 0 & -1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \sigma_y^2 \\ \sigma_\epsilon^2 \\ 2\alpha\sigma_y\sigma_\epsilon \\ \rho_1\sigma_\epsilon^2 \\ \rho_2\sigma_\epsilon^2 \\ \rho_3\sigma_\epsilon^2 \\ \rho_4\sigma_\epsilon^2 \\ \rho_5\sigma_\epsilon^2 \\ \rho_6\sigma_\epsilon^2 \end{bmatrix}$$

Hence the estimators E_1 to E_5 in case III now have the following expectations:

$$E(E_1) = \rho_1\sigma_\epsilon^2 + 2\alpha\sigma_y\sigma_\epsilon - \rho_2\sigma_\epsilon^2$$

$$E(E_2) = [\sigma_y^2 + \rho_3\sigma_\epsilon^2] - 2\alpha\sigma_y\sigma_\epsilon$$

$$E(E_3) = -2\alpha\sigma_y\sigma_\epsilon - \rho_5\sigma_\epsilon^2 + \rho_6\sigma_\epsilon^2$$

$$E(E_4) = [\sigma_\epsilon^2 - \rho_3\sigma_\epsilon^2] + 2\alpha\sigma_y\sigma_\epsilon - \rho_2\sigma_\epsilon^2 + \rho_5\sigma_\epsilon^2$$

$$E(E_5) = 2\alpha\sigma_y\sigma_\epsilon - \rho_4\sigma_\epsilon^2 + \rho_5\sigma_\epsilon^2$$

Thus we can construct estimators of the most important parameters in the model with biases which involve only ρ_2, ρ_5 and ρ_6, all of which may reasonably be assumed to be negligible.

Parameter	Estimator	Bias
$\sigma_y^2 + \sigma_\epsilon^2$	$E_2 + E_4$	$-\rho_2\sigma_\epsilon^2$
$\sigma_y^2 + \rho_3\sigma_\epsilon^2$	$E_2 - E_3$	$-\rho_6\sigma_\epsilon^2$
$\sigma_\epsilon^2 - \rho_3\sigma_\epsilon^2$	$E_4 + E_3$	$-\rho\ \sigma_\epsilon^2 + \rho\ \sigma_\epsilon^2$
$2\alpha\sigma_y\sigma_\epsilon$	$-E_3$	$\rho_5\sigma_\epsilon^2 - \rho_6\sigma_\epsilon^2$
$\rho_1\sigma_\epsilon^2$	$E_1 + E_3$	$-\rho_2\sigma_\epsilon^2$
$\rho_4\sigma_\epsilon^2$	$-[E_5 + E_3]$	$-\rho_6\sigma_\epsilon^2$

The problem still remains with the estimation of σ_y^2 and σ_ϵ^2 since there is no way of separating the impact of ρ_3. However, $\sigma_y^2 + \rho_3\sigma_\epsilon^2$ and $\sigma_\epsilon^2 - \rho_3\sigma_\epsilon^2$ can be estimated with negligible bias.

4.2 CATEGORICAL AND ORDINAL DATA

The proportion of the sample in any single category for a categorical or ordinal variable can be treated in terms of the general model described in section 4.1. The relative simplicity of ordinal and categorical data, however, also provides an opportunity both to use simpler measures and to obtain simple forms of some of the measures previously described.

For a categorical variable the responses obtained from the two interviews may be represented by the square matrix $\{p_{ij}\}$ where p_{ij} is the proportion of the elements classified in category i according to the first interview and in category j according to the second interview. The diagonal of this square matrix, with entries p_{ii}, contains the cases of exact agreement. The simplest measure of reliability (bivariate agreement) is the *index of crude agreement.*

$$A = \Sigma p_{ii} \tag{4.26}$$

which is the proportion of the cases classified identically by the two observations. This index has considerable descriptive value. In the tables we present its complement, the index of crude disagreement

$$D = 1 - A. \tag{4.27}$$

This crude index has a fairly serious drawback, however: it does not take into account the fact that some agreement will occur by chance even if the measurement is completely unreliable (random). The extent of chance agreement depends upon the two marginal distributions $\{p_{i.} (= \Sigma_j p_{ij})\}$ and $\{p_{.j} (= \Sigma_i p_{ij})\}$. One approach, due to Cohen (1960), is to define an index of consistency, kappa, of the form

$$\kappa = 1 - \frac{\text{observed disagreement}}{\text{expected disagreement}}$$
$$= 1 - \frac{1 - p_o}{1 - p_e} = \frac{p_o - p_e}{1 - p_e} \tag{4.28}$$

Under the baseline constraint of independence between the two observations, we have

$$p_e = (\sum_i p_{ii})_e = \Sigma p_{i.} \, p_{.i}$$

giving

$$\kappa = \sum_i (p_{ii} - p_{i.} \, p_{.i})/(1 - \sum_i p_{i.} \, p_{.i}) \tag{4.29}$$

While (4.29) is a more appropriate measure of reliability, particularly in the presence of skewness in the distribution across categories, it can be misleading in situations where a single category dominates the marginal distributions: the value of κ will in this case tend to suggest a low level of consistency if any elements occur off the diagonal. Another point to note in relation to (4.29) is that it would be inappropriate to use κ on its own to describe the level of agreement since it conditions on the observed marginals. The degree of agreement between the marginals is in itself an important component of the observation process. One of a number of possible measures of the disagreement between marginal distributions themselves is

$$B = \frac{2}{\pi} \cos^{-1}\left[\sum_i (p_{i.} \, p_{.i})^{\frac{1}{2}}\right] \tag{4.30}$$

with value '1' indicating complete disagreement, and '0' complete agreement between the two marginal distributions.

The measures (4.26)–(4.29) above are the traditional measures used in cases of multiple observation of the same individual. It is possible also to define simply two other measures.

The index of inconsistency (I), already defined in (4.8), has a particularly simple form for a binary variable. The results of the two interviews can be summarized as follows:

First interview	Re-interview Yes	No	Total
Yes	a	b	$a + b = p_1$
No	c	d	$c + d = q_1$
Total	$a + c = p_2$	$b + d = q_2$	1

It follows that:

$$E(p_1 q_1) = E(p_2 q_2) = \sigma_y^2 + \sigma_\epsilon^2$$
$$E(b + c) = 2\sigma_\epsilon^2(1 - \rho_{\epsilon_1 \epsilon_2})$$

Given that $\rho_{\epsilon_1 \epsilon_2}$ cannot be estimated, a consistent estimator of I is provided by

$$\hat{I} = \frac{b + c}{p_1 q_1 + p_2 q_2} \tag{4.31}$$

An alternative measure, which takes into account the magnitude of the proportion p is given by

$$R = \frac{b + c}{\frac{1}{2}(p_1 + p_2)^2} \tag{4.32}$$

which is a measure of the absolute increase of the relvariance of p due to the simple response variance.

The measures (4.26)–(4.32) described above apply to any level of measurement of the classification variable: categorical (nominal), ordered or metric. When the scales are categorical, any deviation from the diagonal constitutes disagreement. When the scales are ordinal, interval or ratio, any measure of agreement should take into account the *degree* of disagreement, which is a function of the difference between scale values. We can modify (4.26) by redefining 'agreement' to mean that the two interviews obtain values within some acceptable distance (k units) of each other

$$A_k = \sum_{|i-j| \leq k} p_{ij} = 1 - D_k \tag{4.33}$$

Cohen (1968) introduced a modified form of κ which allows for scaled disagreement or partial credit in terms of weights w_{ij} which reflect the contribution of each cell in the table to the degree of disagreement:

$$\kappa_w = \frac{p_o^* - p_e^*}{1 - p_e^*} \tag{4.34}$$

where

$$p_o^* = \sum_{i,j} w_{ij}\, p_{ij}\,;\ p_e^* = \sum_{i,j} (w_{ij}\, p_{i.}\, p_{.j})$$

Any monotonically decreasing function of the differences between the scale values of i and j can be used as weights.

Cicchetti (1972, 1973) suggests the use of the following weights:

$$\text{for ordinal data, } w_{ij} = 1 - |i - j|/(L - 1) \tag{4.35}$$

where L is the number of categories, and

$$\text{for metric data, } w_{ij} = 1 - (i - j)^2 \tag{4.36}$$

Under observed marginal symmetry, κ_w with weights (4.36) is precisely equal to the product-moment correlation coefficient for the integer-valued categories. Furthermore, under the assumption of the random effects model, the estimate of the intra-class correlation coefficient is asymptotically equal to κ_w (Cohen (1968); Fleiss and Cohen (1973)). These measures are discussed in more detail in Landis and Koch (1976).

5 Conclusion

The aim of the analysis in this project is to assess the reliability or precision of the results obtained from WFS surveys. There are two essentially different aspects to this precision: (i) sampling variance, and (ii) response variances.

The sampling variability arises from the fact that only a sample of the population has been included in the survey. Consequently the results depend on which individuals in the population have actually been selected into the sample. The sample designs used for the surveys produce measurable probability samples and hence the sampling errors – or the extent to which the sample results may be expected to diverge from the population results – can be estimated readily from the survey itself. Such computations for a large number of WFS surveys can be found in Verma, Scott and O'Muircheartaigh (1980).

In addition to the sampling variability, however, there is another important source of imprecision in the data. The responses obtained during the interview may be different from the *true values* for the individuals being interviewed. These differences may be systematic or haphazard.

In so far as the discrepancies are systematic, ie are of the same magnitude and direction for all the individuals in the population (sample), statistical methodology will not be of any use in isolating or measuring them. Such errors will make up the *response bias* in the results, and this project makes no contribution to their illumination. However, if the discrepancies are haphazard (or random), then the design of the survey may be modified to permit the estimation of the contribution of these discrepancies to the imprecision of the survey results. Two principal classes of errors are dealt with in this project.

First, the uncorrelated response errors, errors which arise from transient factors associated with the data collection process or from other unmeasured sources. These uncorrelated response errors contribute to the simple response variance (σ_ϵ^2) and their impact on the precision of the sample mean is expressed by I; these measures are described in case I in section 4.1. The simple response variance also affects estimates of association and correlation based on the data by attenuating all such estimates.

The simple response variance can only be estimated if more than one interview is conducted with a set of respondents. In this project, a subsample of the main survey sample was re-interviewed in each country and the discrepancies or deviations between the responses on the two occasions provide the information necessary to estimate the simple response variance.

The second type of response errors dealt with in this project are the correlated response deviations. These arise from the fact that sets of respondents are subjected to the same influences in the data collection process and that these influences may bring about systematic distortion of the responses within these sets of respondents. This could have the effect of increasing substantially the variability of the survey estimates. The technical details are described in case II in section 4.1.

The most important source of correlated response deviations is the interviewer. In order to estimate this 'interviewer effect', the design of the fieldwork execution must be modified. Basically, within each area covered by the sample, respondents must be allocated randomly to interviewers. By ensuring that the differences between interviewer workloads are due only to random variation in the allocation, the observed differences in results obtained by the interviewers provide a measure of the systematic distortion of responses by the interviewers. The estimator of the correlated interviewer variance is given in case II of section 4.1.

Thus, in order to estimate the simple and correlated response variance, it is necessary both to re-interview at least a subsample of the respondents and to randomize the allocation of respondents to interviewers. The design used in this project incorporates both these features. The details of the estimation procedure are given in case III in section 4.1. There are however, two other aspects of the estimators which require consideration.

First, the estimation of simple response variance depends on the assumption that the response deviation obtained in the re-interview is not correlated with the response deviation in the first interview for the same respondent. This assumption is unlikely to hold for some variables since the respondent may be affected by the recollection of the response on the first occasion. Consequently the estimate of the simple response variance may be biased downwards. By using the information available on the elapsed time between the two interviews it is possible to estimate the magnitude of this bias.

Secondly, it is possible that there may be an interaction between the observation process and the true values of the individuals in the sample. If such an interaction is present, it is necessary to complicate the design still further by randomizing the allocation of respondents to groups of interviewers in such a way that both interviews for each of a set of respondents are carried out by interviewers from that group. The design and the estimation procedure are described in case IV of section 4.1. If this interaction is present, the estimator of the interviewer effect will be modified.

Chapter 3 of this paper describes briefly the implementation of this design in the four countries in which this project was carried out. The application of the estimation procedures described in chapter 4 will be presented in later substantive reports in this series. The results from the four countries will be collated and contrasted and should provide a valuable assessment of the effects of variable response errors on WFS data. The project has three principal objectives:

(i) to provide guidance to users of the substantive data;
(ii) to provide the basis of improved survey design;
(iii) to satisfy methodological reasearch objectives.

The final report in the series will evaluate the success of the project in fulfilling these aims.

References

Alam, I. and J. Cleland (1981). Illustrative Analysis: Recent Fertility Trends in Sri Lanka. *WFS Scientific Reports* no 25.

Booth, H. and I. Alam (forthcoming). The Pakistan Fertility Survey: the Quality of the Data. *WFS Monograph on Pakistan.*

Brass, W. (forthcoming). Screening Procedures for Detecting Errors in Maternity History Data.

Chidambaram, V.C., J.G. Cleland and V.K. Verma (1980). Some Issues of Survey Methodology and Data Quality: the WFS Experience. Paper prepared for the PAA Meeting, Denver.

Cicchetti, D.V. (1972). A New Measure of Agreement between Rank Ordered Variables. *Proceedings, 80th Annual Convention, APA*: 17-18.

Cicchetti, D.V. and T. Allison (1973). Assessing the Reliability of Scoring EEG Sleep Records: an Improved Method. *Proceedings and Journal of the Electro-Physiological Technologists' Association 20*: 92-102.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educ. and Psych. Meas. 20*: 37-46.

Cohen, J. (1968). Weighted Kappa: Nominal Scale Agreement with Provision for Scaled Disagreement or Partial Credit. *Psych. Bull. 70*: 213-22.

Fellegi, I.P. (1964). Response Variance and its Estimation. *J. Amer. Statist. Assoc. 59*: 1016-41.

Fleiss, J.L. and J. Cohen (1973). The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. *Educ. and Psych. Meas. 33*: 613-19.

Flórez, C.E. and N. Goldman (1980). An Analysis of Nuptiality Data in the Colombia National Fertility Survey. *WFS Scientific Reports* no 11.

Goldman, N., A.J. Coale and M. Weinstein (1979). The Quality of Data in the Nepal Fertility Survey. *WFS Scientific Reports* no 6.

Guzmán, J.M. (forthcoming). Evaluation of the Dominican Republic National Fertility Survey 1975. *WFS Scientific Reports* no 14.

Hansen, M.H., W.N. Hurwitz and M.A. Bershad (1961). Measurement Errors in Censuses and Surveys. *Bull. Int. Statist. Inst. 38/2*: 359-74.

Hansen, M.H., W.W. Hurwitz and W.G. Madow (1953). *Sample Survey Methods and Theory.* New York: Wiley.

Hobcraft, J. (1980). Illustrative Analysis: Evaluating Fertility Levels and Trends in Colombia. *WFS Scientific Reports* no 15.

Kish, L. (1962). Studies of Interviewer Variants for Attitudinal Variables. *J. Amer. Statist. Assoc. 57*: 92-115.

Landis, J.R. and G. Koch (1975). A Review of Statistical Methods in the Analysis of Data Arising from Observer Reliability Studies. *Statistica Neerlandica 29*: 101-23 and 151-61.

O'Muircheartaigh, C.A. (1977). Response Errors. *The Analysis of Survey Data,* vol II. London: Wiley.

Potter, J.E. (1975). The Validity of Measuring Change in Fertility by Analyzing Birth Histories Obtained in Surveys. Doctoral dissertation, Princeton University.

Somoza, J.L. (1980). Illustrative Analysis: Infant and Child Mortality in Colombia. *WFS Scientific Reports* no 10.

Trussell, J. (1980). Illustrative Analysis: Age at First Marriage in Sri Lanka and Thailand. *WFS Scientific Reports* no 13.

Verma, V.K., C. Scott and C.A. O'Muircheartaigh (1980). Sample Designs and Sampling Errors for the World Fertility Survey. *Jour. Roy. Statist. Soc., A, 143*: 3.