Scientific Reports

NUMBER 45 JANUARY 1984

C. A. O'MUIRCHEARTAIGH

The Magnitude and Pattern of Response Variance in the Peru Fertility Survey

INTERNATIONAL STATISTICAL INSTITUTE
Permanent Office. Director: E. Lunenberg
428 Prinses Beatrixlaan, PO Box 950
2270 AZ Voorburg
Netherlands

WORLD FERTILITY SURVEY
Project Director: Halvor Gille
35–37 Grosvenor Gardens
London SW1W 0BS
United Kingdom

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

The Magnitude and Pattern of Response Variance in the Peru Fertility Survey

C. A. O'MUIRCHEARTAIGH

London School of Economics and Political Science

The recommended citation for this publication is:

O'Muircheartaigh, C. A. (1984). The Magnitude and Pattern of Response Variance in the Peru Fertility Survey. *WFS Scientific Reports* no 45. Voorburg, Netherlands: International Statistical Institute.

Typeset and printed in Great Britain by H Charlesworth & Co Ltd

Contents

FIGURES

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 error – 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 error 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. The first report, 'Methodology of the Response Errors Project' (*WFS Scientific Reports* no 28) described the methodology used, common to all the four country studies. This is the first of the reports which present the results from each of the country studies. 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.

HALVOR GILLE
Project Director

1 Introduction

This is the second of a set of six publications which report on 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 errors in the data collected in WFS surveys; to estimate the magnitudes of these errors; and to examine their implications for the design of the surveys and 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 first report in this series (O'Muircheartaigh 1982) describes the underlying methodology which is common to all the four country studies which comprise the project – Peru, Lesotho, Turkey and Dominican Republic. This second report presents the design, implementation and analysis of the study in Peru. A separate report will be published on each of the four country studies; the final report will present a comparative analysis and evaluate the success of the project.

A general assessment of the quality of WFS data is presented in Chidambaram, Cleland and Verma (1980). The data from the household 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 displacements 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 omission of infant and child deaths in the maternity histories. Hence, for most countries, errors of displacement and omission 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.

These methods of data evaluation do not, however, 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: (i) the nature of the errors; and (ii) 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.

It is important that the analysis of response errors in this context should be geared to the overall aims of evaluation programmes. Three important areas can be distinguished.

1 To report methodological advances In the case of this project there are three types of contribution. First, it is important to demonstrate that an ambitious evaluation programme can successfully be carried out in developing countries. The successes and failures of its implementation have significance for future work. Secondly, the results themselves augment knowledge of the nature and magnitude of some kinds of errors. Thirdly, some new methodological features have been incorporated into the studies and the results of these analyses extend our understanding of the survey process.

2 To improve survey design and execution The project assesses the impact of response errors on the quality of the survey data. By estimating the magnitude of this impact and relating it to various aspects of survey design and implementation, the project provides information which can assist survey designers in reaching informed decisions on the optimum allocation of resources. An important aspect of this is the relative importance of sampling variance and response variance, both for the sample as a whole and for domains of study within the sample.

3 To provide guidance to users of the data This is particularly important since the WFS is entering the data utilization and analysis phase on a large scale. The project provides, for the countries in which it was conducted, a detailed evaluation of the reliability of the survey responses. Users of the data will therefore be in a position to take these additional components of error into account in deciding on the structure of their analysis and in making inferences on the basis of the analysis.

Underlying all three areas there are two general issues. The first is the nature of the total variance of survey estimates. One of the principal contributions of this project is that it permits the decomposition of the total variance into a number of components and permits the assessment and estimation of each of these components for the major variables in the surveys. The implications of each of the components are different both in terms of impact and in terms of possible treatment. The second major issue is the robustness of the findings of the project. It is desirable that the conclusions reached should provide a basis for general statements about variability in the survey results. It is for this reason that the four country studies were made comparable in design and analysis, thus allowing regularities in the pattern of results across countries to be used to strengthen the conclusions reached in the analysis of each country's data separately.

2 Structure of the Study

The over-riding 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.

All of the survey data are based on information collected for a sample of the population of interest. Thus all 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 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 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 the analysis can take this into account. There is however a second source of variability which would persist 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 retrospective 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.

In defining the concept of error it is necessary to postulate 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 affect the individual's 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 1983 as 30 years, the individual response error would be seven 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 as the *expected survey value*. 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 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 has not been 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 analysis therefore is of the *reliability* of the data, rather than of their validity.

There are 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 could 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, that of interpenetration, 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, 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 difference between the results obtained by different interviewers may be due to differences between the individuals whom the interviewer happened to interview rather than to differences caused by 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 (at least within certain defined limits) 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 is 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 were 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 also makes 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 under-estimate 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 section 5. The practical features are discussed in section 3.

The project has two principal objectives. 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*).

3 The Peru Design

The Peru Fertility Survey, conducted by the National Statistics Office during 1977–78, was based on a three 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, 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 present (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) used in 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 in 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 as follows:

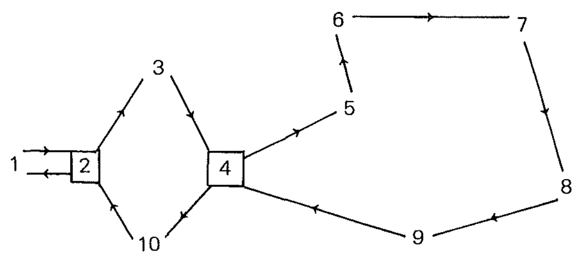
SSU 1	9 households	A B C D E F G A B
SSU 2	10 households	C D E F G A B C D E
SSU 3	7 households	F G A B C D E
SSU 4	15 households	F G A B C D E F G A B C D E F
SSU 5	4 households	G A B C

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 in urban areas 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) was 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 clusters 1-10 and conducts re-interviews in the purposively selected clusters 2 and 4 during its return trip.



A rotation 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 were to be recorded by the supervisor and reported to headquarters at the end of the fieldwork.

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-interviews	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

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 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 by the supervisor 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 (1) that at least two interviewers should simultaneously visit an SSU, with interviews within the cluster allocated randomly between the interviewers, and (2) that re-interviewing in the cluster 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 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 taken 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 the purposively designated 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 the 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. The analysis is described in section 5.

4 Measures of Consistency

For each individual interviewed in the re-interview survey we have two separate observations for each variable. The differences within and between the pairs of observations provide the raw material for the investigation. In general, reliability can be defined as the extent to which a measurement remains constant as it is repeated under conditions taken to be constant. Thus a useful measure of reliability should take into account variations in the individual observations. At a basic level, the most illuminating presentation is that which describes the set of deviations between the observations on the two occasions. This approach has the further advantage that it applies to all types of variables and that the magnitudes of the individual response deviations can be interpreted substantively. In addition, it is applicable to the whole set of variables, regardless of the level of measurement - nominal, ordinal or metric.

4.1 THE BASIC DATA

In this section we consider some examples of this basic procedure. In examining the response obtained on the two occasions for a particular variable, the data can be represented by a cross-classification of the two sets of responses. Tables 1–4 are examples of such cross-tabulations.

Table 1 presents the data for the variable *Ever-use of contraception.* This is a binary variable and thus all the information is contained in a simple 2 X 2 table.

Approximately 20 per cent of the women gave inconsistent responses on the two occasions. The observed response variability stems at least in part from the fact that the basic condition of comparability - the 'essential survey conditions' being the same for the two interviews - was violated. The method of questioning in the two interviews differed. In the original interview, the respondent was asked to name the contraceptive methods she had 'heard of' and for each method mentioned she was asked whether she had ever used it; this was followed by the interviewer reading out a description of a number of other methods one by one and repeating the question on use in each case. This extra probing was not done in the second interview, and a substantial proportion of respondents may consequently have failed to report contraceptive use. The level of ever-use of contraception reported in the first interview was 12 per cent higher than in the second interview, with 15 per cent of all respondents reporting use in the original interview and not in the re-interview, whereas only 3 per cent reported use in the re-interview and not in the original interview. The main illustrative point to be noted here, however, is that the direct presentation of the two sets of responses is a straightforward and comprehensive way of reporting the consistency of response for binary data.

Table 1 Ever-use of contraception as reported in the original interview and the re-interview

Original interview	Re-interview Yes	No	Total
Yes	519	184	704
No	41	453	494
Total	560	638	1198

Table 2 presents the cross-tabulation for *Level of education.* This is an ordinal variable where increasing values of the categories represent greater exposure to education.

The level of education reported differed for one in six respondents, ie for 195 women. For the great majority of those - 174 - the difference between the two responses amounted to a shift through one educational level; for 11 women the discrepancy was two educational levels; and for the remaining 10 women there was a difference of three levels.

This table provides a greater wealth of detail than table 1 because of the number of categories involved. The categories are also in rank order and the difference between the categories is of substantive significance. By observing the marginals of the table, we see that the pattern of results is broadly similar for the two interviews.

Table 3 deals with one of the variables of central importance in a fertility survey – the *Number of children ever born* to the respondent (parity).

Partly because of the size of the table (the number of categories) the pattern of results is striking. For the great majority of respondents the responses on the two occasions are identical. For a variable that seems as unequivocal as this, however, it is perhaps surprising that any observations differ on the two occasions. Most of the discrepant cases involve a difference of only one, but deviations as large as four are found in a number of cases. Overall one in eight

Table 2 Educational level as reported in the original interview and the re-interview

Original interview	Re-interview 1	2	3	4	5	Total
1	259	46	0	4	0	309
2	27	191	18	1	1	238
3	4	35	115	9	1	164
4	1	3	24	178	3	209
5	0	4	2	12	260	278
Total	291	279	159	204	265	1198

NOTE: 1: no education; 2: 1–2 years; 3: 3–4 years; 4: 5–6 years; 5: 7 years or more

Table 3 Number of children ever born as reported in the original interview and the re-interview

Original interview	Re-interview																				
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	Total
0	40	5																			46
1	2	124	7	3																	136
2		1	148	8																	158
3			3	151	11	4	1														169
4			1	4	139	7	1														151
5					2	107	8	4		1											121
6					4	4	98	10													116
7						6	3	78	3												89
8							1	5	55	7	1		1								69
9								1	1	38	5										45
10									4		21	2	1								28
11								3		1	5	26	3								38
12												1	11								11
13												1	1	3	1						5
14											3			1	5						9
15												1				3					4
16																	2				2
17																					
18																					
19																				1	1
Total	43	131	159	165	156	177	111	100	62	47	34	30	17	4	6	3	2	0	0	1	1198

women reported inconsistently; the reporting is less consistent at higher parities than at lower parities, as might be expected. Nevertheless the marginal distributions are very similar, and the means for the original interview and the re-interview are almost identical.

The problems of providing a useful summary of the data are illustrated by this table. There are 361 cells in the table, of which 342 would indicate a discrepancy between the two observations. Only 46 of these cells contain observations, and the importance of these depends on the size of the discrepancy they represent. To discuss each of the occupied cells in turn, however, would be both lengthy and uninformative. This problem is exacerbated by the fact that we wish also to describe the reliability of the data for subclasses of the sample. Thus we will certainly be forced to condense the tables into some summary measures which contain the information necessary to evaluate the data.

One further table may be considered here to illustrate the difficulty. Table 4 gives the two sets of responses for one of the few attitudinal variables included in most WFS national surveys – *Number of children desired.* This table is dramatically different from table 3. We would expect an attitude variable to be particularly subject to response variability and table 4 confirms this expectation. Furthermore, this variable is different in kind from the variables considered in tables 1 to 3 in that the true value of the variable may change between the two interviews. In fact, fewer than half the women gave identical responses on the two occasions. The discrepancies are large and the overall impression is of very unreliable reporting. From a substantive point of view, this variable is of interest more as an indication of the desire for small or large families rather than as a precise measure of behaviour, and it is encouraging that 70 per cent report the number desired within one child in the two interviews. The marginal distributions are relatively stable and the means of the two distributions are very close.

The four tables presented in this section illustrate both the strengths and the weaknesses of this kind of analysis. To some extent it is only by examining the response deviations in detail that we can obtain an understanding of the underlying process. But the tables are relatively unwieldy and cannot realistically be presented for every variable for every subclass of interest. It is therefore necessary to consider how the information may be condensed and summarized to make it more manageable and more easily interpretable. In the next section the simplest summary measures are presented.

4.2 SIMPLE MEASURES OF RELIABILITY

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 all observations 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 matrix $[p_{ij}]$ can be obtained from tables such as tables 1–4 by dividing the frequency in each cell by the total sample size. The simplest measure of reliability is the *index of crude agreement.*

$$A = \Sigma p_{ii} \tag{4.1}$$

which is the proportion of the cases classified identically by the two interviews. This index has considerable descriptive

Table 4 Number of children desired as reported in the original interview and the re-interview

Original interview	Re-interview															
	0	1	2	3	4	5	6	7	8	9	10	11	12	98	99	Total
0	1	2	1	2	3	1	1								1	11
1	1	10	19	1	4	0	0						1		9	45
2	3	12	134	46	35	7	4	1			1	1	0	2	9	253
3	6	3	62	129	48	14	13	0	5		0	0	1	5	12	298
4	1	6	31	34	145	10	25	1	1		1	0	2	1	10	268
5		2	15	18	11	16	9	1	5		3	0	0	6	10	96
6		1	7	21	16	16	29	2	3	1	1	1	4		9	110
7			0	1	1	3	2	2	3	0	0		0		3	15
8			1	4	3	7	4	1	2	0	0		0		3	25
9			0			0	2			0	0		0		6	8
10	1		3			1				1	3		1		5	15
11																0
12				2		1	0		1		4		7		1	10
16						1										1
20			3													3
98	1		3	4	1					1					6	15
99			3	7	1	4								3	8	25
Total	12	36	281	270	268	81	88	7	19	1	12	2	10	17	93	1198

value, as does its complement, the index of crude disagreement

$$D = 1 - A \tag{4.2}$$

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

$$\{p_{i.} (= \sum_j p_{ij})\} \text{ and } \{p_{.j} (= \sum_i p_{ij})\}$$

One approach, due to Cohen (1960), is to define an index of consistency, κ, of the form:

$$\kappa = 1 - \frac{\text{observed disagreement}}{\text{expected disagreement}}$$

$$= 1 - \frac{1 - p_o}{1 - p_e} = \frac{p_o - p_e}{1 - p_e} \tag{4.3}$$

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

$$\sum_i (p_{ii} - p_{i.} \, p_{.i})/(1 - \Sigma p_{i.} \, p_{.i}) \tag{4.4}$$

While (4.4) is a more appropriate measure of reliability than A, especially 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 (4.4) 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 is:

$$B = \frac{2}{\pi} \cos^{-1} \left[\sum_i (p_{i.} \, p_{.i})^{\frac{1}{2}} \right] \tag{4.5}$$

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

The measures (4.1)–(4.5) 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 (4.1) by defining 'agreement' to mean that the two interviews obtain values within some acceptable distance (k units) of each other.

$$A_k = \sum_{|i-j| \leq k} p_{ij} = 1 - D_k \tag{4.6}$$

A modified form of κ can also be used 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:

$$\kappa_w = \frac{p_o^* - p_e^*}{1 - p_e^*} \tag{4.7}$$

Table 5 Values of D, A, κ and κ_w

Variable	D	A	κ	κ_w
Level of education	16	84	79	94
Status of first union	05	95	83	80
Children ever born	12	88	86	98
Ever-use of contraception	19	81	63	63
Age	46	54	53	98
Age in 5 year groups	15	85	83	97
Age at marriage	54	46	41	79
Year of marriage	49	51	50	96
Marital duration (years)	57	43	41	96
Births in past 5 years	16	84	78	91
Worked since marriage (binary)	13	87	70	70
Worked since marriage	28	72	66	–
No. of children desired	56	44	31	42
First birth interval (months)	65	35	32	43
Last closed birth interval (months)	55	45	44	76
Year of first birth	29	71	70	97
Year of last birth	25	75	72	97
Year of next to last birth	36	64	61	94

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. For metric variables, the weights used here are:

$$w_{ij} = 1 - (i - j)^2 \tag{4.8}$$

Under observed marginal symmetry, κ_w with weights (4.8) 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. These measures are discussed in more detail in Landis and Koch (1976).

Table 5 presents the values of D, A, κ and κ_w for eighteen variables. For most variables the index of crude agreement, A, is very close to the supposedly more refined measure κ. This is probably due to the fact that for most of the variables considered the number of categories involved is large, with no dominant category. For an approximately uniform distribution across a large number L of categories, $p_e = 0\ (1/L)$ and it follows from equations (4.1) and (4.3) that for a reasonably consistent set of data, A = $p_o \gg p_e$ so that $\kappa \doteq A$. Hence little is gained by introducing κ in such cases.

Where the range of the variable is very wide, as it is for many of these variables, the discrepancies, while substantively serious, are small compared to the range, and consequently κ_w is a rather insensitive index of consistency. Furthermore, since the marginal distributions are in general fairly close, κ_w will tend to be almost identical to the correlation between the two sets of respondents (ie those on the original interviews and the re-interviews).

In terms of the measures used in table 5 the four most unreliable variables are the *Number of children desired*, *Age at marriage*, the *First birth interval* and the *Last closed birth interval.* The last three of these are composite variables derived from two or more questions, each of which is subject to error. The other is one of the few attitudinal items in the questionnaire, and may be expected to be particularly sensitive to response variability. Even so, the degree of unreliability gives cause for concern. For the *First birth interval*, for example, only one third of the respondents provided the same data on both occasions, and the correlation between the two sets of responses is only 0.4.

Among the variables least affected by response variability are the two measures of fertility which are central to much of the WFS analysis. These are the number of *Children ever born* and *Births in past 5 years.* This is reassuring, although even for these variables the responses from the two interviews are by no means perfectly consistent. Another variable which performs well is *Age in 5-year groups.* It is worth noting, however, that even for this variable one in seven of the women is classified in a different age group in the two interviews.

One further point may be worth noting. The two dichotomies included in the table – *Ever-use of contraception* and *Worked since marriage* – perform reasonably well except in terms of κ_w. Since these variables have only two categories, each discrepancy receives considerable weight in the computation of κ_w. This may be contrasted with the converse cases where the value of κ_w is high relative to A and κ. *Age*, *Children ever born*, *Age at marriage* and *Last closed birth interval* are examples where the value of κ_w is relatively high. This is because in these cases the discrepancies, though they may be substantively serious, are small in relation to the possible range of values for the variable.

Part of the difficulty in evaluating table 5 arises from the fact that the measures considered in this section do not fit easily into the framework of survey analysis and are either too crude, as in the cases of A and D, or unsatisfactory in terms of substantive interpretation, as with κ and κ_w. In section 5 a more general approach is described.

5 Components of the Total Variance

5.1 INTRODUCTION – SIMPLE VARIANCE

The conventional measures of reliability described and used in section 4 do not enable us to fit the examination of the consistency of reporting into the general framework of statistical inference. The total variability of the estimates obtained from the survey is the sum of the sampling variability and the non-sampling variability. In this section we partition the total variance of the estimators into four components, each of which has different implications for survey design. The full model is described in O'Muircheartaigh (1982). In this report response biases are excluded from the model.

A particular survey is regarded as a single trial, ie the survey is regarded as conceptually repeatable. An observation for the *j*th element in the population for trial t is denoted by y_{jt} where j denotes the individual and t denotes the trial.

The observation y_{jt} can be partitioned as follows:

$$y_{jt} = y_j + \epsilon_{jt} \qquad (5.1)$$

where y_j is the true value for element j and ϵ_{jt} is the *variable response error* (or *response deviation*) obtained for element j at trial t. This model ignores fixed response errors (*response biases*). Once we have specified the distribution of the $\{\epsilon_{jt}\}$ the model is completely specified. The distribution of the $\{\epsilon_{jt}\}$ is called the η-distribution. The objective of the survey is to estimate the population mean

$$\bar{y} = \sum_{j=1}^{N} y_j \qquad (5.2)$$

The sample mean of the observations is

$$\bar{y}_{.t} = \frac{1}{n} \sum_{j \epsilon s} y_{jt} \qquad (5.3)$$

Simple sampling variance (SSV)

One of the sources of variation in the results of a survey is the variation among the true values for different individuals in the population. These true values are the quantities of interest in the survey itself. The true value for each individual is fixed. The variation between these values, usually measured by the population variance σ_y^2, is also 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 simplest sample design is a simple random sample. Although such a design is extremely rare in practice, it provides a useful benchmark for the evaluation of other sample designs. For a simple random sample of size n from a population of size N, the variance of the sample mean $\bar{y}_{.t}$ is

$$V_p(\bar{y}_{.t}) = (1 - f')\frac{\sigma_y^2}{n}$$

$$\text{where } f' = \frac{n-1}{N-1} \qquad (5.4)$$

If the finite population correction $(1 - f')$ is ignored, this gives

$$V_p(\bar{y}_{.t}) = \frac{\sigma_y^2}{n} \qquad (5.5)$$

The subscript p in $V_p(\bar{y}_{.t})$ indicates that this is the sampling variance of $\bar{y}_{.t}$, and the variability is a function of the sample design p and its associated sampling distribution. The variance in (5.4) is the *simple sampling variance* (SSV).

In the case of the Peru Fertility Survey, as in all other WFS surveys, the sample design was not a simple random sample. It is however possible to obtain a good estimate of σ_y^2 from the data. The most accurate procedure (given for example in Kish (1965)) involves the use of the correctly estimated sampling variance for the design. In practice an acceptable approximation can be obtained by treating the sample observations as though they had arisen from a simple random sample.

Simple response variance (SRV)

The second important source of variation in the results is the set of response deviations (the $\{\epsilon_{jt}\}$ in (5.1)). The value of an observation is determined not only by the true value for individuals but also by errors of measurement. The presence of these errors makes the estimates derived from the survey observations less stable and less precise than they would otherwise be.

The simplest situation is that in which the only distortion of the true values is a random disturbance term; in other words the response deviations are not correlated with the true values or with each other. In terms of the model this is equivalent to specifying that

$$E(\epsilon_{jt}) = 0 \qquad [\text{all j}]$$

$$V_\eta(\epsilon_{jt}) = \sigma_j^2 = \sigma^2 \qquad [\text{all j}]$$

$$\text{Cov}_\eta(\epsilon_{jt}, \epsilon_{j't'}) = p_{jj'}\, \sigma_j \sigma_{j'} = 0 \qquad [\text{all j}]$$

The component of the variance contributed by these uncorrelated response errors is

$$V_\eta(\bar{y}_{.t}) = \frac{\sigma_\epsilon^2}{n} \qquad (5.6)$$

The variance in (5.6) is a function of the sizes of the response deviations and the size of the sample, and is the *simple response variance* (SRV).

We do not have any direct means of observing the values of the response deviations. In order to estimate σ_ϵ^2 we need to have at least two observations on each individual in the sample. The set of differences $\{y_{j_1} - y_{j_2}\}$ provides for us the values of $\{\epsilon_{j_1} - \epsilon_{j_2}\}$, ie the *difference* between the response deviations for individual j on the first and second occasions. The variance of $(\epsilon_{j_1} - \epsilon_{j_2})$ 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

$$\sigma_{\epsilon_{1,2}}^2 = 2\sigma_\epsilon^2\,(1 - \rho_{\epsilon_1\epsilon_2}) \qquad (5.7)$$

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 $(1 - \rho_{\epsilon_1,\epsilon_2})$. 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. This investigation is described in section 6.1.

Simple total variance (STV)

The simple response variance is a measure of the variability of the response deviations. The simple sampling variance is a measure of the variability of the true values in the population. The sum of these two quantities is

$$\frac{\sigma_y^2}{n} + \frac{\sigma_\epsilon^2}{n} \qquad (5.8)$$

and can be called the *simple total variance* (STV). This is the variance of the mean of a simple random sample of size n from the population when the response deviations $\{\epsilon_{jt}\}$ are uncorrelated. The STV can be estimated directly from the data by taking the observed variance of the observations. Ignoring the finite population correction

$$E\left(\frac{s^2}{n}\right) = \left(\frac{\sigma_y^2}{n}\right) + \left(\frac{\sigma_\epsilon^2}{n}\right) \qquad (5.9)$$

where

$$s^2 = \sum_{j=1}^{n} [(y_{jt} - \bar{y}_{.t})^2/(n-1)] \qquad (5.10)$$

A useful measure of the reliability of the data is the *index of inconsistency*, I, where

$$I = \frac{\sigma_\epsilon^2}{\sigma_y^2 + \sigma_\epsilon^2} \qquad (5.11)$$

This index measures the proportion of the simple total variance (5.8) which may be attributed to the simple response variance (5.6). Thus, in effect, the index I enables us to partition the simple total variance into two constituent parts: 'true' variability in the underlying values of the variable in the population and the random disturbance (noise) introduced into the observations by the measurement process itself.

Estimation of the components of the simple total variance

As the preceding section demonstrates, the simple variance estimated from a sample of observations automatically includes the simple sampling variance and the simple response variance. With repeated observations we obtain in effect two estimates of this simple total variance, one from the original interviews and one from the re-interviews. The simple sampling variance and the simple response variance, however, can only be estimated from the two sets of observations together, as described on pages 16 and 17. This section gives an example of the estimation of the components of the simple total variance and of the index of consistency, I.

On the basis of the data in table 6, the parameters of the three frequency distributions can be estimated. The distributions of the responses in the original interviews and the re-interviews provide estimates of the simple total variance. The distribution of the deviations provides an estimate of the simple response variance. Table 7 gives the estimates.

Both 1.037 and 1.034 are estimates of the simple total variance $\sigma_y^2 + \sigma_\epsilon^2$, whereas 0.182 is an estimate of $2\sigma_\epsilon^2$. Thus the estimate of σ_ϵ^2 is

$$\hat{\sigma}_\epsilon^2 = \tfrac{1}{2}(0.182) = 0.091$$

The best available estimate of $\sigma_y^2 + \sigma_\epsilon^2$ is

$$\hat{\sigma}_y^2 + \hat{\sigma}_\epsilon^2 = \tfrac{1}{2}(1.037 + 1.034) = 1.0355$$

Consequently

$$\hat{I} = \frac{0.091}{1.0355} = 0.08788$$

This procedure makes use of all the available data. Instead of using the matrix containing the full cross-classification of the responses from the two interviews (examples are given in tables 1–4 in section 4) which becomes unwieldy when the number of categories is large, the data are used in the form given in table 6. All the components of the simple total variance can be derived from these distributions.

Figure 1 presents the components of the simple total variance for twelve key variables, arranged in order of increasing values of $\hat{I}$. The variables in figure 1 show a very wide range of values for $\hat{I}$. For the variable 'Age' the data show a very high degree of reliability, with only 2 per cent of the simple total variance being attributable to simple response variance. At the other extreme, the variables, 'First birth interval' (FBINT) and 'Number of children desired' (DESFAM) show a high degree of unreliability,

Table 6 Data for the estimation of simple variance components for the variable *Births in the past five years*

Value	Original interview	Re-interview	Deviation between original interview and re-interview	Frequency
0	407	440		
1	389	381	−2	4
2	283	266	−1	58
3	103	97	0	1009
4	16	10	1	120
5	0	4	2	7
Total	1198	1198	Total	1198

Table 7 Components of the simple variance for *Births in the past five years*

	Original interviews	Re-interviews	Deviations
Mean	1.110	1.054	0.056
Standard deviation	1.018	1.017	0.427
Standard error	0.029	0.029	0.012
Variance	1.037	1.034	0.182

with values of $\hat{I}$ of 0.56 and 0.58 respectively. In the case of these two variables more than half the simple total variance is due to the response deviations – in other words, of the total variability in the responses less than half can be attributed to genuine differences in the underlying values of the variable in the population; the remainder is due to disturbances introduced into the observations by the measurement process itself.

Inconsistency of different categories of respondents

Although the results in figure 1 give an overall impression of the degree of reliability of responses for the variables considered it is important to bear in mind that most of the analysis of the data will be carried out on subsets of the whole sample, ie subclasses of the population. This section looks at three important sets of subclasses: age groups, education subclasses, and city size. The classification value of each of these is taken as reported in the first interview.

Table 8 gives the values of $\hat{I}$ for six selected subclasses – the youngest and oldest age groups; residents of Lima and rural residents; and those with no education and those with seven or more years of education. The results in table 8 show the values of $\hat{I}$ for the extreme subclasses for each characteristic.

In general the results show a consistent pattern. For older women, for women with little education and for women residing in rural areas, the values of $\hat{I}$ are generally higher and in some cases much higher than for younger, better educated urban women. The background characteristics are not unrelated, of course; better educated women are generally younger and tend to live in urban areas. The number of cases in the sample (n = 1198) does not, however, permit good estimation of the degree of inconsistency for cells of a two-way or three-way classification.

A number of important conclusions emerge from the table. The value of $\hat{I}$ for the total sample is not sufficient to evaluate all estimates based on a particular variable. In the case of *Age at marriage*, for instance, the value of $\hat{I}$ for

Table 8 gives the values of $\hat{I}$ for six selected subclasses -

Variable	Age <25	Age ≥45	Lima	Rural	No educ.	Educ. ≥7 years	All
Age	06	29	01	03	04	01	02
Children ever born	02	03	01	03	03	01	02
Year of first birth	28	15	01	03	03	04	02
Age in 5 year groups	24	–	02	04	06	02	03
Year of last birth	09	03	01	06	04	02	03
Year of marriage	10	25	02	06	11	01	04
Marital duration	12	24	02	06	11	01	04
Education	06	06	08	16	–	–	06
Year of next to last birth	59	07	02	08	06	07	06
Births in past 5 years	05	28	05	13	11	06	09
Last closed birth interval	81	31	13	18	17	34	20
Age at marriage	16	29	11	36	40	07	20
Worked since marriage	31	40	21	43	53	21	30
Ever-use of contraception	30	36	34	62	45	37	35
First birth interval	57	66	29	77	82	42	56
No. of children desired	40	75	47	67	60	36	58

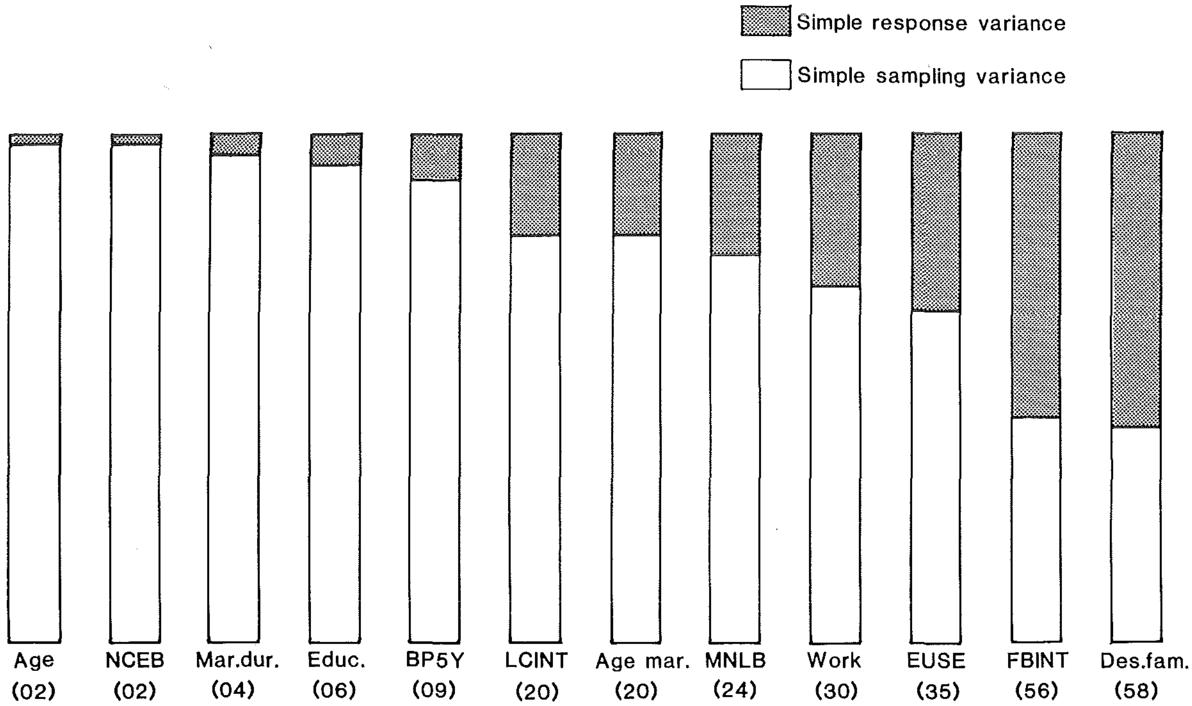


Figure 1 Simple total variance

the total sample is 0.20. In Lima the value is only 0.11 whereas in rural areas the value is 0.36. For the women with 7 or more years of education $\hat{I}$ is only 0.07 whereas for those with no education it is 0.40. Thus the degree of confidence we can have in the reliability of the data may be very different for different parts of the sample.

Figures 2A and 2B show in graphical form the partitioning of the simple total variance for the urbanization and age subclasses of two variables – *Marital duration* (MARDUR) and *Number of children desired* (DESFAM). Marital duration is a relatively reliable variable - the overall value of $\hat{I}$ is 0.04 - whereas number of children desired is extremely unreliable, $\hat{I} = 0.58$.

These figures indicate the implications and the pattern of the variation on reliability for different subclasses. The pattern is consistent - unreliability increases with age, rurality and lack of education. The first part of figure 2B has particular significance. It can be seen clearly from the figure that each and every age subclass has a higher value of $\hat{I}$ than the sample as a whole. This is not the case for any variable for any subclass other than age subclasses.

This occurs because the variable concerned (marital duration) is itself age-related. Since the sample is a cross-section of the population the variance of the true values (σ_y^2) of the observations within a particular age subclass is much lower than the variance of the true values for the population as a whole. The response variance (σ_ϵ^2), on the other hand, must be, on average across subclasses, equal to the response variance for the whole sample. This affects the three variables *age*, *marital duration* and *year of marriage* very strongly, and is an important factor to bear in mind when considering the impact of response errors on analysis within age group.

Overall the analysis in this section indicates that the values of measures of reliability for the total sample provide only a rough guide to the reliability of results for subclasses of the population. This reservation is particularly noteworthy for the analysis of age-related variables, but is also relevant to analysis for any variables when dealing with subclasses containing a high proportion of rural, uneducated or older respondents.

5.2 CORRELATED VARIANCE – DESIGN EFFECT AND INTERVIEWER EFFECT

Section 5.1 discusses the partitioning of the simple total variance, which is the sum of the simple sampling variance and the simple response variance. The simple sampling variance is a function of the variability among the true values in the population, and is the variance of the mean of a simple random sample of size n selected from the population of true values. In practice, however, simple random samples are rarely if ever used. The actual sampling variance is thus not adequately measured by the simple sampling variance. The complexity of the design of the sample, usually involving both stratification and clustering, has an impact on the sampling variance, and a realistic presentation of the sampling variance must take these complexities into account. Similarly the simple response variance is the variance of the response deviations when it is assumed that the response deviation for each individual in the sample is independent of the response deviations of the other respondents. This would be realistic only if there were no factor in the field execution which affected different groups of respondents in different ways. Any interrelationship between the response deviations within groups of respondents may lead to an increase in the response variance over that estimated by the simple response variance. In this section the estimation of the correlated variance is discussed.

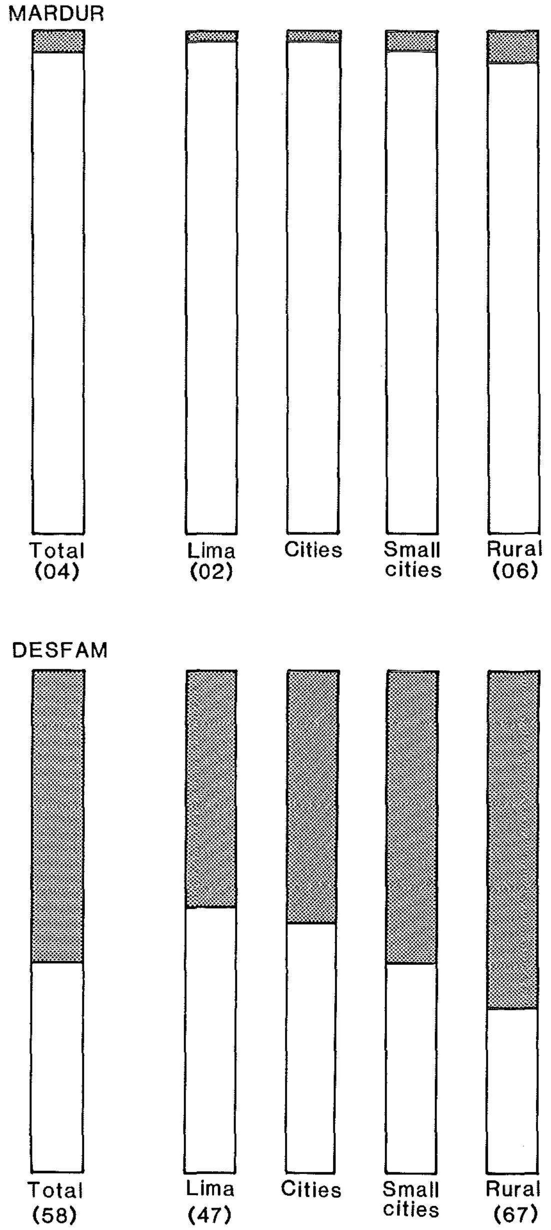


Figure 2A Simple total variance: urbanization subclasses

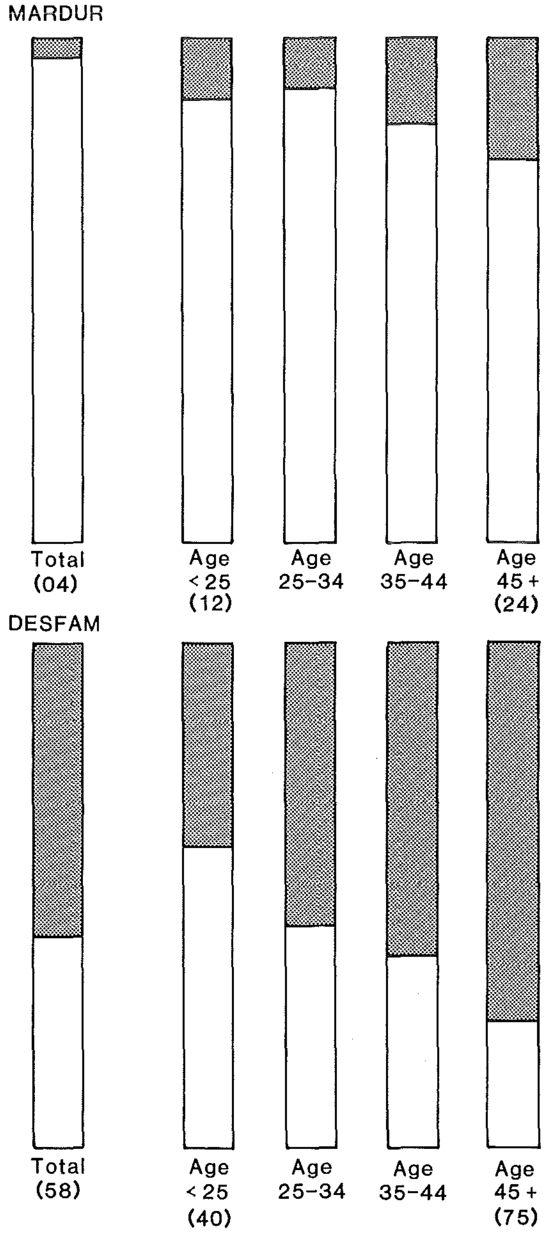


Figure 2B Simple total variance: age subclasses

Correlated sampling variance (CSV)

Simple random samples are rarely if ever found in practice in field surveys. Most sample designs are stratified multistage designs and the sampling variance of such designs is normally greater than the sampling variance of a simple random sample of the same size. Typically, although stratification leads to a reduction in variance, this effect is dominated by the increase in variance due to the clustering of the sample. The effect of clustering arises from the positive correlation between the true values for individuals in the same cluster. The impact of the sample design on the sampling variance in WFS surveys is presented for twelve countries in Verma, Scott and O'Muircheartaigh (1980). The total sampling variance can be expressed, ignoring the finite population correction, as

$$V_p(\bar{y}_{.t}) = \frac{\sigma_y^2}{n}\{1 + \text{roh}\,(\bar{b} - 1)\} \qquad (5.12)$$

The synthetic intra-cluster correlation coefficient, *roh*, is a measure of the internal homogeneity of the clusters used in the sample design. This coefficient gives an indication of the relative similarity of individuals within a cluster compared to the similarity of individuals in the population as a whole. The more similar individuals are to one another within a cluster, the larger the value of *roh* will be.

The quantity $\bar{b}$ is the average number of individuals interviewed in each cluster. The increase in the variance over the simple sampling variance given by (5.5) is

$$\frac{\sigma_y^2}{n}\{\text{roh}\,(\bar{b} - 1)\} \qquad (5.13)$$

and may be called the *correlated sampling variance* (CSV).

It is clear from (5.12) and (5.13) that the size of $\bar{b}$ will have an important impact on the correlated sampling variance and thus on the *total sampling variance* (TSV) which is given by (5.12).

In the presentation of sampling variance, the concept of the *design effect* (Kish 1965) is frequently used. The design effect, usually denoted by Deff, is the ratio of the total sampling variance (5.12) to the simple sampling variance (5.5) and is

$$\text{Deff} = \frac{\frac{\sigma_y^2}{n}\{1 + \text{roh}(\bar{b} - 1)\}}{\frac{\sigma_y^2}{n}}$$

$$= 1 + \text{roh}(\bar{b} - 1) \tag{5.14}$$

The total sampling variance is thus a function of both the variability among the true values of the individuals in the population and the degree of clustering introduced into the sample by the sample design.

Correlated response variance (CRV)

The analysis of response deviations presented in section 5.1 treats these deviations as uncorrelated; in other words, for each particular variable the response deviation for one individual is assumed not to be dependent on, or related to, the response deviation for another individual. There is, however, one important element of the survey operation which may tend to invalidate this assumption, at least for some variables. The possible intercorrelation arises from the fact that each interviewer carries out a set of interviews and may have a systematic effect on the responses of those whom she interviews, in addition to the random (haphazard) disturbances in the responses. If this is the case, then the estimates of variance obtained ignoring this factor may seriously underestimate the actual variance of the estimators. The situation is analogous to that of the sampling variance where the simple sampling variance would underestimate the total sampling variance.

The simple model in section 5.1 can be modified to take the possibility of intercorrelated errors into account. The assumptions given in (5.6) can be changed to:

$$E_\eta(\epsilon_{jt}) = 0 \qquad [\text{all j}]$$

$$V_\eta(\epsilon_{jt}) = \sigma_\epsilon^2 \qquad [\text{all j}]$$

$$\left.\begin{aligned} \text{Cov}_\eta(\epsilon_{ijt}, \epsilon_{i'j't}) &= \rho_1\sigma_\epsilon^2 \quad \text{if } i = i' \\ &= \rho_2\sigma_\epsilon^2 \quad \text{if } i \neq i' \end{aligned}\right\} \tag{5.15}$$

In (5.15) ρ_1 is the correlation between the response deviations for individuals interviewed by the same interviewer. The subscript i denotes the interviewer. For completeness ρ_2, the correlation between response deviations for individuals interviewed by different interviewers, is included although typically ρ_2 will be negligibly small.

Under the model (5.15) the contribution of the response deviations to the total variance will be:

$$V_\eta(\bar{y}_{.t}) = \sigma_\epsilon^2\{1 + \rho_1(m - 1) + \rho_2 m(k - 1)\}$$

Ignoring ρ_2, this becomes

$$V_\eta(\bar{y}_{.t}) = \frac{\sigma_\epsilon^2\{1 + \rho_1(m - 1)\}}{n} \tag{5.16}$$

where m is the size of each interviewer's workload. If workloads vary in size the formula can be used as an approximation with the average workload size for the interviewers. The increase in the variance over the simple response variance given by (5.7) is

$$\frac{\sigma_\epsilon^2}{n}\{\rho_1(m - 1)\} \tag{5.17}$$

and may be called the *correlated response variance* (CRV).

The intra-interviewer correlation coefficient, ρ_1, is a measure of the homogeneity imposed on the responses by the consistent or systematic effect of each interviewer. There is a striking similarity between the form of the expression (5.16) for the *total response variance* and the expression (5.12) for the *total sampling variance.*

In order to estimate the correlated response variance due to the interviewers the survey design must be modified. The procedure is discussed in detail in O'Muircheartaigh (1982). 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. The implementation of the allocation procedure for Peru has been described in section 3.

From the data we 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, ignoring ρ_2,

$$\left.\begin{aligned} &E_pE_\eta\{C\} = \sigma_y^2 + \sigma_\epsilon^2\{1 + \rho_1(m - 1)\} \\ &\text{and} \\ &E_pE_\eta\{F\} = \sigma_y^2 + \sigma_\epsilon^2(1 - \rho_1) \end{aligned}\right\} \tag{5.18}$$

Hence

$$\frac{1}{m}\{C - F\}$$

provides a possible estimator of $\rho_1\sigma_\epsilon^2$. In fact, under this model,

$$E\left\{\frac{1}{m}\{C - F\}\right\} = \sigma_\epsilon^2(\rho_1 - \rho_2)$$

but 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).

5.3 THE TOTAL VARIANCE

The partitioning of the total variance of the estimator is presented in figure 3 below. The total variance is shown to be composed of four components: the simple and correlated sampling variances and the simple and correlated response variances. The implications of each of these components are different in terms of survey design and execution.

The simple sampling variance can be affected only by changing the sample size. The correlated sampling variance is due to, and can be modified by, the choice of sample design. The intra-cluster correlation coefficient is determined by the choice of clusters (sampling units) for the design: the more homogeneous the clusters the larger the clustering effect. The average subsample size within the selected clusters is the other determining factor, and for a given sample size depends on the number of clusters included in the sample.

The simple response variance is to some extent a measure of the quality of the data collection process. It is a measure of the degree to which the responses obtained represent the true values of the variables for the respondents. With a perfect measurement process the simple response variance would be zero. The simple response variance represents the effects of all the factors which cause the responses to deviate in a variable or non-systematic way from the true values. The correlated response variance is the additional variance due to the interrelationships between the response deviations. The most important cause of such intercorrelation, and the one dealt with in this paper, is the interviewer. If the interviewers have consistent but different effects on the respondents whom they interview, this will produce an additional component of variance which is analogous to the additional sampling variance produced by the selection of clusters of elements in a cluster sample.

The two components of the simple total variance - the simple sampling variance and the simple response variance - represent the basic underlying components of the variance. The true values of the individuals in the population, which underlie the simple sampling variance by determining σ_y^2, are fixed regardless of the survey design. The response variability among the individuals, which underlies the simple response variance, is the result of the field execution, the questionnaire and the characteristics of the respondents themselves, and cannot be changed unless we change either the questionnaire or the quality of the field execution. Thus figure 1 (see page 19) represents the basic situation with regard to the variance of the estimators.

5.4 ESTIMATES OF THE CORRELATED RESPONSE VARIANCE – INTERVIEWER VARIANCE

It is not normally possible to estimate the correlated response variance in a survey. In order to do so the fieldwork design must be modified by allocating the respondents randomly to interviewers at least within sampling units. In Peru, as in the three other countries in which the Response Error Project was carried out, the overall design included both random allocation of respondents to interviewers and re-interviewing of the respondents. The design used was based on Fellegi (1964) and is extremely powerful in terms of the components of the total variance which it enables the analyst to estimate. In this paper, however, we do not make use of all the features of the design. In the case of Peru, due to the interruptions and practical difficulties in the execution of the fieldwork referred to in section 3, the

Figure 3 Total variance by source and type

Source	Type: Simple	Correlated	Total
Sampling	$\boxed{\text{SSV}}$ Simple sampling variance: the variance of a simple random sample of size n	$\boxed{\text{CSV}}$ Correlated sampling variance: the additional variance due mainly to the clustering of the sample	$\boxed{\text{TSV}}$ Total sampling variance: a function of the sample design and the variability among the true values in the population
Measurement (response)	$\boxed{\text{SRV}}$ Simple response variance: due to random (haphazard) response. deviations caused by the observation or measurement process	$\boxed{\text{CRV}}$ Correlated response variance: the additional variance due to interrelationship between the response deviations caused by, for example, a common interviewer for each group of respondents	$\boxed{\text{TRV}}$ Total response variance: a function of the data collection process
Total	$\boxed{\text{STV}}$ Simple total variance: discussed in section 5.1; the variance of the estimator for a simple random sample with uncorrelated response deviations	$\boxed{\text{CTV}}$ Correlated total variance: the additional sampling and response variance neglected by the analysis in section 5.1	$\boxed{\text{TV}}$ Total variance: the actual variance of the estimators

allocation of respondents to interviewers diverged substantially from that planned, particularly in the re-interview survey. This makes it impracticable to carry out the full analysis described in O'Muircheartaigh (1982). Indeed a number of approximations were needed to permit the analysis described below. The position for Lesotho, Dominican Republic and Turkey is much more satisfactory.

The estimation is consequently confined to that described under the heading of Correlated Response Variance (page 21). The interviews from the main survey and those for the re-interview survey must be analysed separately. The magnitude of the correlated response variance can be estimated separately for the same set of variables in each case.

The correlated response variance - in this case the *interviewer variance* - is of the form given by (5.17) and is:

$$\frac{\rho_1\sigma_\epsilon^2 (m-1)}{n} \tag{5.17}$$

In this case we cannot, for each set of interviews, estimate ρ_1 directly. We can, however, estimate $\rho_1\sigma_\epsilon^2$. The term $(m-1)$, where m is the interviewer workload, is an artefact of the design. A good index of the potential impact of the interviewer variance is

$$\rho_1 I = \rho_1\sigma_\epsilon^2/(\sigma_y^2 + \sigma_\epsilon^2) \tag{5.19}$$

where I is the index of inconsistency, defined on page 17. The denominator in (5.19) can be estimated simply by using s^2, where s^2 is defined in (5.10).

Table 9 gives the estimated values of $\rho_1 I$ for the five variables which produced significant results for Peru. Our estimation procedure provides separate estimates of $\rho_1 I$ for the two phases in Peru, and although these values are all estimates it is interesting to note that with one exception the same five variables emerged in both analyses as those most sensitive to interviewer effect. The exception is *Ever-use of contraception* which had the largest estimated interviewer effect in the main survey, an effect which disappears completely in the re-interviews.

It is possible, but unlikely, that this is due to the imprecision of our estimates of $\rho_1 I$ and a more reasonable explanation is that it arises from the difference in procedure for this question between the first and second interviews. In this first interview this question included detailed probing by the interviewer, which was not repeated in the re-interview. The results indicate that this probing may be particularly sensitive to interviewer effect.

The values of $\rho_1 I$ provide an index of the susceptibility of variables to interview effect. The magnitude of the variance component may be expressed either as

$$\frac{\rho_1\sigma_\epsilon^2}{n}(m-1) \tag{5.17}$$

or alternatively as

$$\rho_1 I \frac{(\sigma_y^2 + \sigma_\epsilon^2)}{n}(m-1) \tag{5.20}$$

which has the advantage that it uses as a base the value of

$$\frac{(\sigma_y^2 + \sigma_\epsilon^2)}{n}$$

which is the simple total variance. The simple total variance is easily and directly estimable from the survey data and also provides the base against which the sampling variance is measured in most survey work.

Whichever form is used, the most important point to note is that the average interviewer workload m is critical in determining the magnitude of the variance component. Even a relatively small value of $\rho_1 I$ will have a considerable impact on the total variance if the value of m is large. With a value of $\rho_1 I = 0.02$, for example, and $m = 100$, the effect of the correlated response variance would be to increase the total variance by an amount equal to twice the simple total variance.

A large value of ρ_1 would not in itself be sufficient to imply a large increase in the total variance. The size of the simple response variance (σ_ϵ^2) is also important. If σ_ϵ^2 is small - in particular if it is small relative to the simple total variance - even a large value of ρ_1 will have little impact. This is indicated also by the fact that four of the five variables in table 9 are the four variables with the highest values of $\hat{I}$ in figure 1.

The central point is that the correlated response variance is an additional contribution to the total variance due to intercorrelations between the response deviations. Thus, in principle, if a variable is not subject to fluctuations in response - if there is no simple response variance - there cannot be any correlated response variance. Similarly if the simple response variance is very small, a very high degree of intercorrelation among the response deviations would be necessary before the correlated response variance could make a substantial contribution to the total variance. If however for a variable with non-negligible simple response variance, the response deviations are sensitive to the behaviour or other characteristics of the particular interviewer who conducts the interview, then the interviewer variance may be an extremely important component of the total variance and could in some cases dominate the total variance.

Table 9 Estimates of $\rho_1 I$

Variable	Main survey	Re-interview survey
Ever-use of contraception	0.10	0.00
Whether worked	0.04	0.05
Education	0.06	0.05
No. of children desired	0.03	0.15
First birth interval	0.02	0.04

5.5 PARTITIONING THE TOTAL VARIANCE

In this section two variables are considered in detail - *Ever-use of contraception* and *First birth interval*, both of which are subject to considerable interviewer effect. For each the total variance is presented in terms of its four components: simple sampling variance, simple response variance, correlated sampling variance and correlated response variance (interviewer variance). In order to put the results into perspective, two additional variables are considered in section 5.6 - *Children ever born* and *Age at marriage* - neither of which shows any evidence of interviewer effect.

Figure 4 presents the results for *First birth interval.* Column I of figure 4A represents the estimate of the total variance for the Peru Fertility Survey. In the main survey the average interviewer workload (m) was 101 and the average number of individuals interviewed in each primary sampling unit (b) was 14.

The simple total variance is the sum of the two bottom components in the bar chart. This quantity can be estimated directly from the survey data and is:

$$\frac{\sigma_y^2 + \sigma_\epsilon^2}{n} \tag{5.9}$$

As part of the routine analysis of the WFS surveys, the sampling variance is estimated using the CLUSTERS program. The estimate is actually an estimate of the simple total variance plus the correlated sampling variance. In the notation of this section it is an estimate of:

$$\frac{\sigma_y^2}{n}\{1 + \text{roh}\,(b-1)\} + \frac{\sigma_\epsilon^2}{n} \tag{5.21}$$

The correlated sampling variance is

$$\frac{\sigma_y^2}{n}\{\text{roh}\,(b-1)\}$$

and it is represented by the third component (from the bottom) of the bar chart. The interviewer variance is:

$$\frac{\sigma_\epsilon^2}{n}\{\rho_1\,(m-1)\} \tag{5.17}$$

and it is represented by the top component of the bar chart.

The total variance is:

$$\frac{\sigma_y^2}{n} + \frac{\sigma_\epsilon^2}{n} + \frac{\sigma_y^2}{n}\{\text{roh}\,(b-1)\} + \frac{\sigma_\epsilon^2}{n}\{\rho_1\,(m-1)\} \tag{5.22}$$

All the components of variance are affected by the sample size, but their relative magnitudes are not dependent on the sample size. Of the factors in (5.22) only two (apart from n) are subject to manipulation through the survey design. These are the interviewer workload size m and the average cluster 'take' b.

The effect of changing the value of m can be seen from columns I to V of figure 4A.

Column I gives the estimate of the actual total variance and its components for the survey design used in Peru. The magnitude of the correlated sampling and response variances are based on the values of b and m actually used, ie b = 14 and m = 101. It is clear from column I that the total variance is dominated by the interviewer variance, which accounts for some 61 per cent of the total variance. The correlated sampling variance accounts for 9 per cent, the simple sampling variance for 13 per cent and the simple response variance for 17 per cent.

Column VI shows only the simple total variance. This is the estimate of the total variance that would be obtained if s^2/n were used as the estimator, ie if the variance were estimated as though for a simple random sample. In this case the total variance would be under-estimated by a factor of more than three.

Column V shows the quantity actually estimated in practice for WFS surveys. This is the estimate provided by using the correct formula for the sampling variance. In fact it estimates the total sampling variance plus the simple response variance. The only component of the total variance neg-

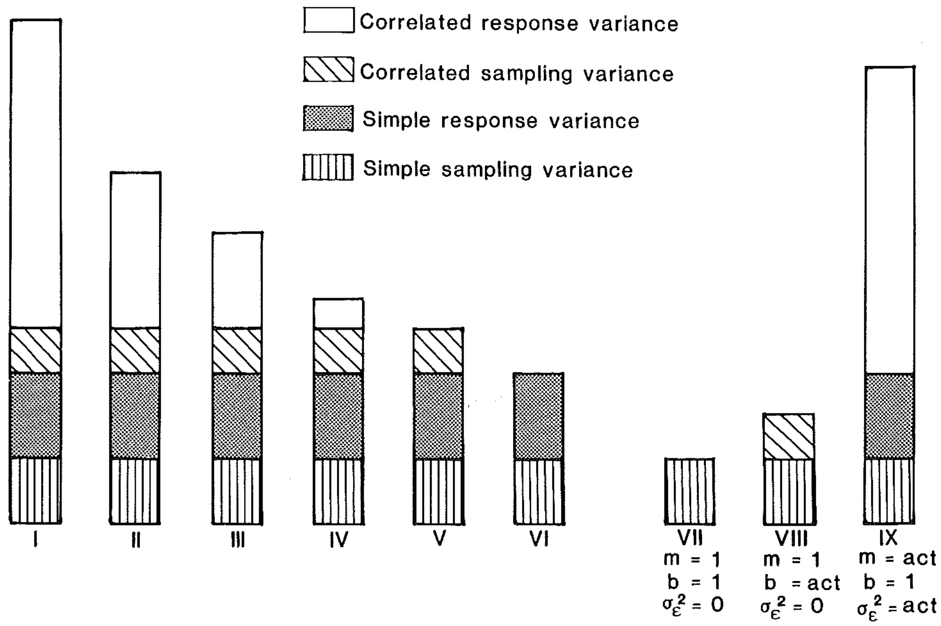


Figure 4A Total variance: first birth interval lected by this estimate is the correlated response variance.

Columns II, III and IV give an indication of the way in which the total variance could be reduced by changing the field strategy, within a fixed total sample size. Column II gives the total variance for a design in which the number of interviewers is doubled, keeping the sample size unchanged. The effect is to cut by half the contribution of interviewer variance to the total variance, due to the reduction of the interviewer workload m and consequently of the term $\rho_1\sigma_\epsilon^2$ (m – 1)/n. It is assumed in this case that the quality of the interviewers is not affected by increasing their number. Columns III and IV indicate the effect of reducing the interviewer workload to 31 and 11 respectively under the same assumption. In principle column V is the variance obtained when m = 1, ie when each respondent is interviewed by a different interviewer.

Column VII is the minimum variance possible for a sample of size n (assuming no stratification). This would be the case if a simple random sample of size n were selected and if the measurement were perfect, ie if there were no response errors of any kind. Column VIII represents the actual total sampling variance for the design used.

Figure 4B is an alternative way of looking at the information in columns I to IV. Each bar shows the relative contribution of the four components of variance for one of the six sets of circumstances. Column VI is identical to the representation (FBINT) in figure 1.

The results for *ever-use of contraception* are given in figure 5. The situation is even more dramatic in this case. By comparison with the simple sampling variance and the simple response variance the correlated variance components are overwhelming, and between them they account for more than 90 per cent of the total variance. The difference between column VI and column V highlights the necessity of proper estimation of sampling variance. Ignoring the effect of the clustering in the sample design would lead to an under-estimation by a factor of three. The contrast between columns V and I shows that for this variable also the total variance is dominated by the interviewer variance, accounting as it does for almost 75 per cent of the total variance. This situation of course is due not only to the intercorrelation between the response deviations but also to the large average workload size. Columns II, III and IV show the effect of reducing the workload size. Figure 5B presents these results in percentage terms and demonstrates how this dominance by the interviewer variance can be radically altered. With an average workload size of m = 11, for instance, the interviewer variance - other things being equal - would account for less than a quarter of the total variance.

5.6 SUMMARY MEASURES AND CONFIDENCE INTERVALS

The results in section 5.5 are not typical of all variables in the Peru Fertility Survey. The two variables described there are those for which the impact of response variance is greatest. In order to put these results in perspective a set of four variables is considered in this section which includes all types of variables in terms of the relative magnitude of the different components of the total variance.

In order to simplify the presentation some manipulation of the terms used in the earlier sections is required, particularly for the components of the correlated variance. Instead of using σ_y^2/n as a base for the correlated sampling variance and σ_ϵ^2/n as a base for the correlated response variance it is possible to use the simple total variance $(\sigma_y^2 + \sigma_\epsilon^2)/$ n as a base for both.

Thus the correlated sampling variance which has previously been written as

$$\frac{\sigma_y^2}{n}\{\text{roh}\,(b-1)\} \tag{5.13}$$

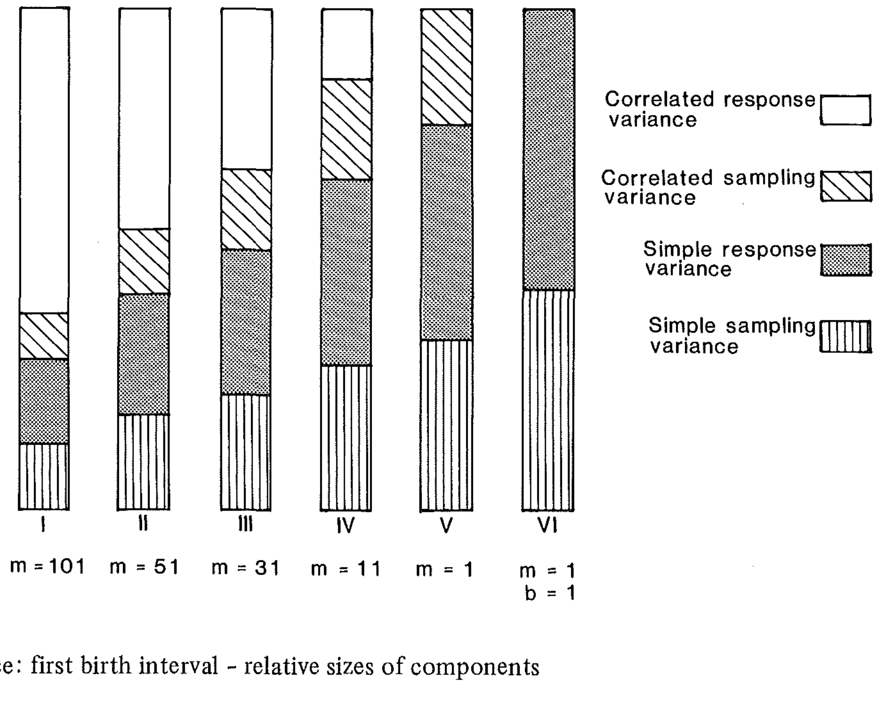


Figure 4B Total variance: first birth interval - relative sizes of components

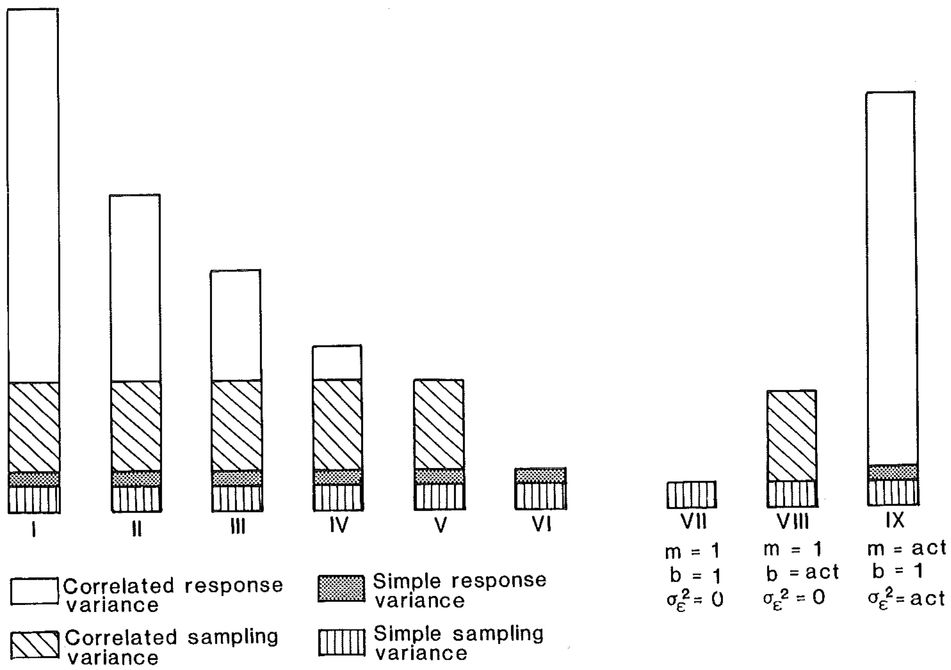


Figure 5A Total variance: ever-use of contraception

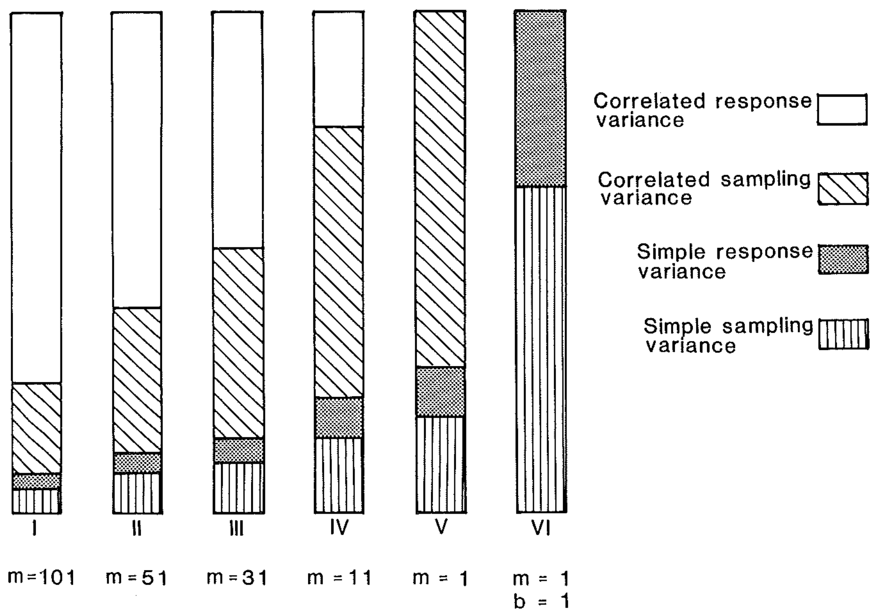


Figure 5B Total variance: ever-use of contraception - relative sizes of components

can alternatively be written as

$$\frac{\sigma_y^2 + \sigma_\epsilon^2}{n} \cdot \frac{\sigma_y^2}{\sigma_y^2 + \sigma_\epsilon^2} \{\text{roh}\,(b-1)\}$$

$$= \frac{\sigma_y^2 + \sigma_\epsilon^2}{n} (1 - I)\{\text{roh}\,(b-1)\}$$

$$= \frac{\sigma_y^2 + \sigma_\epsilon^2}{n} \rho_{cl}(b-1) \qquad (5.23)$$

where ρ_{cl} is a synthetic intra-cluster correlation coefficient which takes into account the presence of the simple response variance. The quantity roh which is estimated by standard sampling error programs is in fact ρ_{cl} and not the pure roh in (5.13). The estimate of the design effect, deff, is in fact an estimate of $1 + \rho_{cl}(b-1)$.

Similarly the interviewer variance component can be expressed either as

$$\frac{\sigma_\epsilon^2}{n} \{\rho_I\,(m-1)\} \qquad (5.17)$$

or, using $(\sigma_y^2 + \sigma_\epsilon^2)/n$ as a base, as

$$\frac{\sigma_y^2 + \sigma_\epsilon^2}{n} \cdot \frac{\sigma_y^2}{\sigma_y^2 + \sigma_\epsilon^2} \{\rho_1 (m-1)\}$$

$$= \frac{\sigma_y^2 + \sigma_\epsilon^2}{n} I\rho_1 (m-1)$$

$$= \frac{\sigma_y^2 + \sigma_\epsilon^2}{n} \rho_{int}(m-1) \qquad (5.24)$$

where ρ_{int} is equal to $\rho_1 I$.

The total variance (5.22) can now be written as

$$\frac{\sigma_y^2 + \sigma_\epsilon^2}{n} \{1 + \rho_{cl}(b-1) + \rho_{int}(m-1)\} \qquad (5.25)$$

The design effect becomes

$$\text{Deff} = 1 + \rho_{cl}(b-1)$$

and by analogy, the interviewer effect is

$$\text{Inteff} = 1 + \rho_{int}(m-1)$$

The design factor is

$$\text{Deft} = \sqrt{\text{Deff}}$$

and the interviewer factor is

$$\text{Inteft} = \sqrt{\text{Inteff}}$$

In the results presented here, Deff and Inteff (and consequently Deft and Inteft) are estimated and their estimates will be denoted by deff, inteff, deft and inteft. The choice between using variances and standard errors depends on the purpose for which the results are presented. Table 10 provides both for the four variables concerned. The variables are *Children ever born*, *Age at marriage*, *First birth interval* and *Ever-use of contraception.*

In order to make the first and fourth columns of the table comparable to the others, $1/\hat{I}$ is presented instead of $\hat{I}$. This quality, the reciprocal of the index of inconsistency, measures the factor by which the simple sampling variance must be multiplied to give the simple total variance.

The variable least affected overall is *Children ever born.*

Table 10 Summary measures of the variance components and the standard errors for four variables

Variable	$1/\hat{I}$	deff	inteff	$\sqrt{1/\hat{I}}$	deft	inteft
Children ever born	1.02	1.14	1.00	1.01	1.07	1.00
Age at marriage	1.25	1.10	1.00	1.12	1.05	1.00
First birth interval	2.27	1.30	2.99	1.51	1.14	1.73
Ever-use of contraception	1:54	3.39	11.02	1.24	1.84	3.32

It has a very small component of simple response variance relative to the simple sampling variance; the effect of the clustering of the sample on the variance is slight - an increase of only 14 per cent; and there is no evidence of interviewer effect. Taking the simple sampling variance as a base, the total effect of all the other components is to multiply the variance by a factor of 1.16. If the simple total variance is taken as a base, the multiplying factor is 1.14.

The second variable is similarly dominated by the simple sampling variance, although the simple response variance in this case accounts for 20 per cent of the simple total variance. The effect of the clustering of the sample is to multiply the simple total variance by 1.10. There is no evidence of any interviewer variance. The overall ratio of the actual total variance to the simple sampling variance is 1.375.

The two remaining variables are very different. In both cases the simple response variance is a substantial element in the simple total variance. Furthermore the design effect and the interviewer effect are large for both variables. The ratio of the total variance to the simple sampling variance is 7.49 for the *First-birth interval* and 20.65 for *Ever-use of contraception*; the ratios of the total variance to the simple total variance are 3.29 and 13.41 respectively.

These ratios are easily calculable from the figures given in table 10. From (5.25) we have that the total variance is:

$$\frac{\sigma_y^2 + \sigma_\epsilon^2}{n} \{1 + \rho_{cl}(b-1) + \rho_{int}(m-1)\} \qquad (5.25)$$

$$= \frac{\sigma_y^2}{n} \cdot \frac{1}{I} \cdot \{1 + (\text{deff} - 1) + (\text{inteff} - 1)\}$$

$$= \frac{\sigma_y^2}{n} \cdot \frac{1}{I} \{\text{deff} + \text{inteff} - 1\} \qquad (5.26)$$

The expression (5.26) illustrates the advantage of working directly with the variances and the ratios of variances. Whereas it is more relevant in some respects to use standard errors and the ratios of standard errors, they cannot be presented in such simple additive form. The expression corresponding to the relationship (5.26) is that the total standard error is

$$\sqrt{\left[\frac{\sigma_y^2 + \sigma_\epsilon^2}{n} \{1 + \text{deft}^2 - 1 + \text{inteft}^2 - 1\}\right]}$$

$$= \sqrt{\left[\frac{\sigma_y^2}{n} \cdot \frac{1}{I} \{1 + \text{deft}^2 - 1 + \text{inteft}^2 - 1\}\right]} \qquad (5.27)$$

A possible solution would be to base our measures of the design effect and the interviewer effect not on the ratio of the variances but on the increment in the variance due to each component. Thus if we define

$$d^2 = \text{deff} - 1$$

and

$$i^2 = \text{inteff} - 1$$

we would have the total variance equal to

$$\frac{\sigma_y^2 + \sigma_\epsilon^2}{n} \{1 + d^2 + i^2\} \qquad (5.26')$$

and the total standard error equal to

$$\sqrt{\frac{\sigma_y^2 + \sigma_\epsilon^2}{n} \{1 + d^2 + i^2\}} \tag{5.27'}$$

To illustrate the effects of the variance components on the four variables table 11 gives the width of the 95 per cent confidence intervals using the three possible estimation procedures. The sample mean is also given for each variable, and column (1) gives the simple sampling error (ie σ_y^2/n).

The possible estimates of the standard error are used in columns (2), (3) and (4) in calculating the width of the confidence interval. Column (2) is calculated using s^2/n as the estimate of the total variance; column (3) uses the appropriate calculation for a complex sample design where the data are free of correlated response variance; and column (4) gives the correct estimate of the total error.

The variable *Children ever born* illustrates the position when neither the correlated sampling variance nor the correlated response variance has much impact. Similarly the various estimates for *Age at marriage* differ little from one another. It should be noted however that for some variables not given in table 11, the design effect is important even though there is no interviewer effect.

For the remaining two variables the situation is very different. For the *First birth interval* the width of the confidence interval estimated using the correct procedure to estimate sampling error would be 2.40. When the interviewer effect is taken into account the width of the confidence interval is seen to be 3.85 - an increase of 60 per cent over the usual estimate. For *Ever-use of contraception* the adjustment is even more striking. Column (3) gives a confidence interval of width 0.1052; the correct confidence interval is of width 0.2094 - an increase of almost 100 per cent.

The importance of the interviewer effect may be illustrated also by considering the true confidence level for the intervals constructed using the estimates of standard error from column (3). For the *First birth interval* the apparent 95 per cent confidence interval is actually a 78 per cent confidence interval; for *Ever-use of contraception* the true confidence level is 68 per cent.

The first two variables in table 11 are more representative of variables from WFS surveys than are the last two. Furthermore the figures in table 11 are based on estimates of the variance components and these estimates are themselves subject to sampling error. The problem of estimating the variance of the estimates of the variance components will be dealt with in later reports in this series.

5.7 EFFECTS ON CROSSCLASSES

In common with many other surveys, one of the main objectives of the WFS is to produce separate estimates for subgroups or subclasses of the study population, such as particular demographic, socio-economic or geographic categories. While the number of substantive variables involved may not be very large, the subclasses of interest tend to be much more numerous; each cell of the multiway cross-tabulations of the survey results forms a subclass. Further, much of the analysis of survey results may take the form of comparing and contrasting estimates for different subclasses, resulting in an even larger number of *subclass differences* of interest.

In practice it therefore becomes necessary to confine computation of variances to a selection of subclasses and subclass differences. This approach was used in Verma, Scott and O'Muircheartaigh (1980) in the presentation and analysis of sampling errors for the WFS. In that paper three groups of subclasses were used: (i) subclasses defined in terms of *demographic* characteristics (age, marriage duration, etc); (ii) subclasses defined in terms of *socio-economic* characteristics (woman's literacy, husband's level of education, occupation, etc); and (iii) a small number of *geographic* subclasses (regional and urbanization classes, for instance). These different subclasses correspond to the major categories by which WFS surveys are cross-tabulated.

Subclasses in the three groups tend to differ in the way in which the elements in the subclasses are distributed across the primary sampling units in the sample. Demographic subclasses are generally fairly uniformly distributed across clusters and form what may be called *crossclasses.* Socio-economic subclasses have a less uniform spread; higher educational groups and non-farming occupations tend to be concentrated in urban areas, for example. These may be called *mixed classes.* By contrast, geographic subclasses are in most cases completely *segregated* - either all or none of the elements in a sample cluster will belong to a subclass. This terminology is due to Kish, Groves and Krotki (1976).

For several purposes it is useful to investigate the relationship between the total variance for an estimator based on the whole sample and the total variance for subclasses and subclass differences: (a) to extrapolate results computed for a particular set of subclasses to numerous other subclasses of interest; (b) to simplify the presentation of results; and (c) to seek stable relationships between the total variance for the whole sample and the total variance for subclasses of particular kinds. In this context, if a stable pattern is found for the relationship, this may provide a

Table 11 Width of 95 per cent confidence interval for four variables using different estimates of the total error

Variable	Mean	Simple sampling error (1)	Simple total error (2)	(2) × deft (3)	Correct standard error (4)
Children ever born	4.66	0.356	0.360	0.384	0.384
Age at marriage	19.9	0.428	0.476	0.500	0.500
First birth interval	11.3	1.40	2.12	2.40	3.85
Ever-use of contraception	0.56	0.0460	0.0572	0.1052	0.2094

better procedure for estimating the total variance for a subclass than direct computation, since each individual estimation is itself subject to a (possibly) large sampling variance.

Three models have been used in the past for the relationship between the variance for the whole sample and the variance for a subclass. The work in this area has been done for sampling variance only and is described in Kish *et al* (1976) and Verma *et al* (1980). The empirical results obtained have suggested that for crossclasses the intracluster correlation coefficient is approximately stable, although it may increase slightly as the relative size of the crossclass decreases.

In this section the analysis is extended to the more complex case of the total variance. The data available for Peru do not permit empirical testing of the model described, but this will be done in a later report in this series. The purpose of this section is to investigate the implications of a simple approximate model for the total variance for a crossclass. The algebraic presentation is illustrated by applying it to the total variance found in Peru for the four variables discussed in section 5.6.

A model

The total variance of the sample mean is:

$$V(y) = (\sigma_y^2 + \sigma_\epsilon^2)\ \{1 + \rho_{cl}(b-1) + \rho_{int}(m-1)\} \tag{5.25}$$

The model proposed here makes a number of assumptions. In particular σ_y^2, σ_ϵ^2, ρ_{cl} and ρ_{int} are assumed to be the same for the total sample and for the crossclasses. Denoting characteristics of the total sample by the subscript t and those for a crossclass by the subscript s, we therefore have:

$$\left.\begin{aligned} V_t(\bar{y}) &= \frac{\sigma_y^2 + \sigma_\epsilon^2}{n_t}\ \{1 + \rho_{cl}(b_t - 1) + \rho_{int}(m_t - 1)\} \\ &\text{and} \\ V_s(\bar{y}) &= \frac{\sigma_y^2 + \sigma_\epsilon^2}{n_s}\ \{1 + \rho_{cl}(b_s - 1) + \rho_{int}(m_s - 1)\} \end{aligned}\right\} \tag{5.28}$$

For a subclass of size $n_s = M_s n_t$ (ie using M_s to denote n_s/n_t)

$$\frac{V_s}{V_t} = \frac{\{1 - \rho_{cl} - \rho_{int}\} + M_s\{\rho_{cl}b_t + \rho_{int}m_t\}}{M_s\{1 - \rho_{cl} - \rho_{int}\} + M_s\{\rho_{cl}b_t + \rho_{int}m_t\}}$$

$$> 1 \text{ if } M_s < 1 \text{ and } 1 - \rho_{cl} - \rho_{int} \geqslant 0 \tag{5.29}$$

A limiting case of some interest is that of a simple random sample with no correlated response variance, ie $\rho_{cl} = \rho_{int} = 0$. In this case

$$\frac{V_s}{V_t} = \frac{1}{M_s} \tag{5.30}$$

In general, (5.29) can be written as:

$$\frac{V_s}{V_t} = \left\{1 + \frac{1 - M_s}{M_s} \cdot \frac{K_1}{K_1 + K_2}\right\} \tag{5.31}$$

where $K_1 = 1 - \rho_{cl} - \rho_{int}$

and $K_2 = \rho_{cl}b_t + \rho_{int}m_t$

Thus we see that the relationship between the total variance for the whole sample and the total variance for a crossclass can be reduced to a very simple form. The quantity $(1 - M_s)/M_s$ is fixed for a subclass making up the proportion M_s of the whole sample. The only quantity which needs to be calculated is $K_1/(K_1 + K_2)$ where K_1 is a simple function of ρ_{cl} and ρ_{int}, and K_2 takes into account also the average cluster take in the whole sample (b_t) and the average workload size for the whole sample (m_t).

An application

The model described immediately above is applied to the four variables previously described. The derivation shows that the important factors are ρ_{cl}, ρ_{int} and $K_1/(K_1 + K_2)$. Table 12 gives the values of these parameters for the four variables.

The first two variables are examples of the simple case when there is no interviewer variance and the correlated sampling variance is also relatively small. The effect of this is to give values of $K_1/(K_1 + K_2)$ close to 1, which is the limiting value for the situation where there is no correlated variance. The third variable is an intermediate case where both components of correlated variance are present and non-negligible. The last variable is an extreme case where the data are subject to large correlated sampling variance and large correlated response variance. The effect of this is seen in the extremely small value of $K_1/(K_1 + K_2)$ - the absolute minimum value for this factor is zero.

The implications of the parameters in table 12 can be seen from table 13, which gives the relative magnitude of V_t and V_s - the values of V_s/V_t are presented for three different subclass sizes. The subclass sizes chosen are M_s = 0.5, 0.3 and 0.1. The first corresponds to a subclass which makes up half of the sample, the second to a subclass comprising 30 per cent of the sample, and the third comprising one tenth of the sample. Most subclasses used in practice fall in this range, although for multiway classifications even smaller subclasses may be involved.

In evaluating the figures in table 13 it is important to remember that the ratio V_s/V_t must be between $1/M_s$ and 1, where the value $1/M_s$ corresponds to the case where there is no correlated variance, and the total variance is inversely proportional to sample size. For reference, the last row of the table gives the values of $1/M_s$.

As we would expect, the variables *Children ever born* and *Age at marriage* have values of V_s/V_t close to the upper limit. This is because there is no interviewer variance for these variables and the correlated sampling variance is relatively small. The results for the variable *First birth interval*

Table 12 Values of ρ_{cl}, ρ_{int} and $K_1/(K_1 + K_2)$

Variable	ρ_{cl}	ρ_{int}	$K_1/(K_1 + K_2)$
Children ever born	0.0113	0.00	0.8638
Age at marriage	0.0078	0.00	0.9021
First birth interval	0.0234	0.02	0.2899
Ever-use of contraception	0.1875	0.10	0.0532

Table 13 Relative magnitude of V_s and V_t (values of V_s/V_t for three subclass sizes)

Variable	Subclass size (M_s)		
	0.5	0.3	0.1
Children ever born	1.86	3.02	8.77
Age at marriage	1.90	3.10	9.12
First birth interval	1.29	1.68	3.61
Ever-use of contraception	1.05	1.12	1.48
No correlated variance	2.00	3.33	10.00

show how unwise it would be to apply this limit to a case where either of the correlated variance components is reasonably large. Under the assumptions of this model, using the upper limit for the variance would lead to overestimating the total variance by 50 per cent when $M_s = 0.5$; by almost 100 per cent when $M_s = 0.3$; and by 177 per cent when $M_s = 0.1$.

The last variable in the table shows even more dramatic results. This variable is atypical since both the correlated sampling variance and the interviewer variance are extremely large. In such a situation, however, the effects are astonishing. For a crossclass with $M_s = 0.5$ the total variance is almost identical to the total variance for the whole sample, although the sample size for the subclass is only half the size of the whole sample. The further reduction of sample size for $M_s = 0.3$ and $M_s = 0.1$ leads only to relatively small increases in the variance. For $M_s = 0.1$ (a crossclass comprising one tenth of the sample) the ratio of V_s/V_t is only 1.48. For a variable with no correlated variance this ratio would be 10.00.

The results in tables 12 and 13 can also be presented in a form closer to the approach used in discussing sampling variance. Table 14 gives the values of *deff*, *inteff* and *toteff*, where

$$\text{toteff} = \text{deff} + \text{inteff} - 1 \tag{5.32}$$

and toteff is the ratio of the *total variance* (5.22) to the *simple total variance* (5.9).

The results in table 14 conform to the pattern observed in the sampling literature for crossclasses. Under the assumptions of the model the effect of the correlated variance components decreases as the proportion of the population in the crossclass decreases. The larger the effect of the correlated variance components, the more dramatic the reduction as M_s decreases.

Finally, to illustrate the practical implications of these results for the evaluation of survey estimates, table 15 gives the width of the 95 per cent confidence intervals for crossclasses of different sizes. The same four variables are presented and the width of the confidence interval for the estimate based on the whole sample is also given for comparison.

The relationship between the standard error for a subclass and the standard error for the whole sample is determined by two factors: (i) the size of the sample for the subclass. The smaller the sample size (ie the smaller M_s) the larger the standard error will be - this applies to all components of the total variance; (ii) the relative size of the correlated errors. In the absence of correlated errors, the only influence will be the relative sizes of the total sample and the subclass. However when there are correlated errors, either sampling or response, the relationship becomes more complex. For crossclasses, the model described on page 29 implies that there will be a considerable dilution of the effect of the reduction in sample size. This is because the impact of the correlated errors depends critically on the size of the 'clusters' within which the errors are correlated; for correlated sampling errors the cluster take is the dominant factor, for correlated interviewer errors the interviewer workload size is the critical consideration. For small crossclasses both these sizes are greatly reduced, with a consequent reduction in the correlated components.

The final column of table 15 encapsulates the results of this section. For the two variables *Children ever born* and *Age at marriage* the ratio of the standard errors (and thus of the confidence intervals) is close to that expected on the basis of sample size alone - the correlated errors are relatively unimportant. For the *First birth interval* the confidence interval for the smaller crossclasses is a good deal narrower than would be expected if sample size were the only consideration. For *Ever-use of contraception* the confidence interval for the crossclass with $M_s = 0.1$ (ie based on one tenth of the total sample) is only 22 per cent wider than the confidence interval based on the total sample. This is because the dominance of the correlated error

Table 14 Values of *deff*, *inteff* and *toteff* for different values of M_s

Variable	Measure	Total sample	$M_s = 0.5$	$M_s = 0.3$	$M_s = 0.1$
Children ever born	deff	1.14	1.06	1.04	1.00
	inteff	1.00	1.00	1.00	1.00
	toteff	1.14	1.06	1.04	1.00
Age at marriage	deff	1.10	1.04	1.02	1.00
	inteff	1.00	1.00	1.00	1.00
	toteff	1.10	1.04	1.02	1.00
First birth interval	deff	1.30	1.16	1.08	1.01
	inteff	2.99	1.99	1.59	1.18
	toteff	3.29	2.12	1.66	1.19
Ever-use of contraception	deff	3.39	2.10	1.59	1.08
	inteff	11.02	6.00	4.00	1.91
	toteff	13.41	7.03	4.51	1.98

Table 15 Width of 95 per cent confidence intervals for crossclasses of different sizes

Variable	Mean	Subclass	Simple sampling error (1)	Total simple error (2)	(2) X deft (3)	Correct standard error (4)
Children ever born	4.66	Total sample	0.356	0.360	0.354	0.384
		$M_s = 0.5$	0.504	0.508	0.524	0.524
		$M_s = 0.3$	0.648	0.656	0.668	0.668
		$M_s = 0.1$	1.124	1.136	1.136	1.136
Age at marriage	19.9	Total sample	0.428	0.476	0.500	0.500
		$M_s = 0.5$	0.604	0.672	0.684	0.684
		$M_s = 0.3$	0.776	0.872	0.880	0.880
		$M_s = 0.1$	1.348	1.508	1.508	1.508
First birth interval	11.3	Total sample	1.40	2.12	2.40	3.85
		$M_s = 0.5$	1.968	2.972	3.212	4.389
		$M_s = 0.3$	2.540	3.836	3.992	5.005
		$M_s = 0.1$	4.400	6.644	6.676	7.315
Ever-use of contraception	0.56	Total sample	0.0460	0.0572	0.1052	0.2094
		$M_s = 0.5$	0.0652	0.0812	0.1176	0.2136
		$M_s = 0.3$	0.0840	0.1048	0.1320	0.2220
		$M_s = 0.1$	0.1456	0.1812	0.1884	0.2555

in the standard error of estimates based on the total sample becomes progressively weaker as the crossclass size decreases.

Discussion

A model is presented (page 29) which describes the total variance of an estimate in terms of five factors: the simple total variance; the synthetic intracluster correlation coefficient for the sample design; the synthetic intra-interviewer correlation coefficient for the fieldwork design; the average cluster take; and the average interviewer workload size. The model is analogous to that generally used to describe the total sampling variance. The implications of this model for the total variance of estimates based on crossclasses were presented, and a simple expression was derived for the relationship between the total variance for the total sample and the total variance for a crossclass. A number of important assumptions are made in the model. First, it is assumed that the crossclasses are uniformly distributed across clusters and interviewers; in the context of WFS surveys, age subclasses are likely to satisfy this condition at least approximately. Secondly, it is assumed that the intra-cluster and intra-interviewer correlation coefficients remain constant for crossclasses. The evidence on this is less convincing, although it seems a useful approximation in practice. In particular the evidence for the intra-cluster correlation coefficient suggests that it is reasonably stable. Further investigation of the behaviour of the intra-interviewer correlation is desirable.

The application above (page 29) provides an illustration of the theoretical implications of the model. The results are presented for four variables which represent the different situations which might arise. For two of the variables the total variance is primarily due to the simple sampling variance and the simple response variance, with a small correlated sampling variance component. In this case the relative size of crossclass variance is determined largely by the crossclass size. For the third variable there is a more substantial correlated sampling variance component and also a correlated response variance component. The total correlated variance dominates the total variance for estimates based on the whole sample. For crossclasses, however, this dominance is reduced as the crossclass size decreases. For small crossclasses the simple variance predominates and the effects of the correlated variances almost disappear. The situation is even more striking for the fourth variable - the total effect (the ratio of the total variance to the simple total variance) is 13.41 for estimates based on the total sample and is only 1.98 for estimates based on a crossclass representing one tenth of the total sample. This is an extension of the results obtained for sampling variance in other studies - the effects of the design are diminished as the size of the crossclass is reduced.

Although the results (pages 29–31) are not based on direct computations of the variances, values of the parameters on which the calculations are based are obtained from computations carried out on the data from the Peru study. For technical reasons it was not possible to estimate the total variance for the crossclasses and thus to examine the extent to which the theoretical results provide a good fit to the data. It is hoped, however, to test the model directly using the data from the other studies in this project.

The choice of sample design and field design for a survey tends to be determined by material and practical constraints imposed by the data collection operation. Nevertheless, data relating to sampling and response errors can provide a more rational basis for making decisions about the design. The findings of this section, however, illustrate a particular difficulty. A basic consideration in evaluating the design is the relative importance attached to estimates based on the whole sample compared with those for sample subclasses and subclass differences. Generally, the smaller a subclass the less sensitive is the associated variance to specific features of the design. In particular the less is the effect of the correlated components of the variance and the more ill-defined is the 'optimal' solution to the problem of survey design.

6 Further Analysis

6.1 ESTIMATION OF σ_ϵ^2

One of the difficulties involved in estimating σ_ϵ^2 is that our estimator σ_ϵ^2 will under-estimate the true variance if there is a positive correlation between the response deviations of the same individual in the two interviews (see section 5). In practice this correlation will tend to be positive - the respondent may, for example, remember some of the responses from the first interview and tend to give the same answers in the re-interview. The problem with our general model is that we have no way of obtaining an estimate of this correlation directly. It is only by using information from outside the model that we can get any indication of the possible size of this correlation. In the case of Peru, we have one further source of information. Due to the long period over which the fieldwork extended, it is possible to compare the magnitudes of estimated response deviations for varying time intervals between the interview and the re-interview; for the rural sector in particular we have time intervals ranging from 1 month to 10 months. A simple regression analysis, taking the squared response deviation as the dependent variable and the time interval in months as the independent variable, was carried out for the set of variables. Statistically significant results were obtained for four variables - *Children ever born*, *Age group*, *Births in past 5 years*, and *Month of last birth.* Assuming that the effect of recalling the first response has disappeared after 10 months, the results indicated that the estimates of σ_ϵ^2 should be inflated by a factor of 1.3, approximately.

An examination of the residuals from the simple linear regression suggested that for some variables a more appropriate model would include a quadratic term in the time interval. For *First birth*, *Age at marriage* and *Year of marriage* this modified model produced a good fit, but the indicated under-estimation was only about 10 per cent in each case.

The evidence provided by this analysis is not by any means conclusive, particularly since, although an attempt was made to exclude real changes over time from the data, there was some difficulty in achieving this for *Births in the past five years.* Refinement of the estimation procedure and evidence from other surveys would be necessary before any substantial modification of the estimated σ_ϵ^2 should be introduced. Data from the United States (Bailar 1968) suggest that the effect is negligible for many items, which is the situation we found for nine out of the 13 variables considered. We hope to investigate this further when data from the later surveys become available.

6.2 RELATIONSHIPS BETWEEN RESPONSE DEVIATIONS FOR DIFFERENT VARIABLES

Up to this point we have considered reliability separately for each variable. Since much of the substantive analysis of the data will, however, use composite variables, it is of considerable importance to examine whether the response deviations for different variables are related. In particular, even if large errors are present in each of the variables used to construct a composite variable or to measure a time interval, if these errors are compensating, the derived variables may be considerably more stable than the variables used in its construction. Age at marriage, for instance, is derived using the date of birth and date of marriage and if each of these is displaced by the same amount, age at marriage will be unaffected. Similarly the first birth interval is constructed from the date of marriage and the date of the first birth.

Analysis carried out on the data from Nepal (Goldman *et al* 1979) has shown that those women who do not report dates of births or who report ages at heaped numbers are also more likely to omit births from the fertility histories and to misreport duration of marriage. Hence it is suggested that different kinds of misreporting may be strongly correlated with one another. Both the direction and the magnitudes of these correlations are of importance to the substantive data analysis.

We calculated all the correlations for the set of variables previously given in table 6 and found that in fact the response deviations for many of the variables were correlated. The pattern of the correlations was also consistent with the hope that the errors would be compensating. The magnitudes of the coefficients were not, however, sufficient to explain much of the variability in the individual variables. The correlation between the response deviations for the date of birth and date of marriage, for instance, was 0.2. This implies that the errors in the two variables will cancel out to a negligible extent. The situation for the first birth interval was similar. Although the response deviations for the date of marriage and date of first birth were correlated, the magnitude of the correlation (about 0.3) was insufficient to reduce substantially the response variability of the first birth interval.

Data were not available to examine the individual dates in the birth history and the situation in that context may be more favourable since these questions are part of an integrated set of questions. Some evidence to support this contention is however available from the relatively high reliability of the *Last closed birth interval.*

Thus the results so far suggest that although the errors in individual variables do compensate to some extent in mitigating the unreliability of derived variables, the mitigation is slight in relation to the total variability.

The presence of uncorrelated response deviations affects estimates of coefficients of association and correlation in a manner different from the effect on estimates of means and totals. In measuring the correlation between two variables x and y, the estimator of the correlation ρ_{xy} is attenuated by the factor $(\alpha_x\alpha_y)^{\frac{1}{2}}$ where α_x and α_y are the correlations between the original responses and the responses from the re-interviews for variables x and y respectively.

In fact α is directly related to the index of inconsistency I, since

$$\alpha = 1 - I \tag{6.1}$$

Thus the usual well-known result on attenuation may be re-expressed as:

$$\sqrt{(1 - I_x)} \quad \sqrt{(1 - I_y)} \tag{6.2}$$

This attenuation factor depends, however, on the assumption that the response deviations for different variables are uncorrelated with each other. This is the issue considered above in the context of variables derived from the difference between two variables. The same issue arises when considering correlation and association. In particular the covariance between response deviations for different variables is

$$\begin{aligned} \text{Covar}(\epsilon_y, \epsilon_x) &= E(y_{jt} - y_j)(x_{jt} - x_j) \\ &= \rho_{\epsilon(xy)}\,\sigma_{\epsilon(x)}\,\sigma_{\epsilon(y)} \end{aligned}$$

where $\rho_{\epsilon(xy)}$ is the correlation between the response deviations for x and y.

A measure analogous to the index of inconsistency I can be defined which describes the proportion of the covariance of the observations on x and y due to the covariance between the response deviations:

$$\text{Covar}(x_t y_t) = \text{Covar}(xy) + \text{Covar}(\epsilon_x \epsilon_y)$$

and

$$\begin{aligned} I_{xy} &= \frac{\text{Covar}(\epsilon_x \epsilon_y)}{\text{Covar}(xy) + \text{Covar}(\epsilon_x \epsilon_y)} \\ &= \frac{\rho_{\epsilon(xy)}\,\sigma_{\epsilon(x)}\,\sigma_{\epsilon(y)}}{\rho_{(xy)}\sigma_x\sigma_y + \rho_{\epsilon(xy)}\,\sigma_{\epsilon(x)}\,\sigma_{\epsilon(y)}} \end{aligned} \tag{6.3}$$

where ρ_{xy} is the actual correlation between x and y, both measured without response error.

The correlation between the observations measures

$$\text{Corr}(x_t y_t) = \frac{\text{Covar}(xy) + \text{Covar}(\epsilon_x \epsilon_y)}{\sqrt{\text{Var}(x) + \text{Var}(\epsilon_x)}\sqrt{\text{Var}(y) + \text{Var}(\epsilon_y)}}$$

which can be shown to be equal to

$$\begin{aligned} &\frac{\text{Covar}(xy)}{\text{Var}(x)\,\text{Var}(y)} \quad \frac{\sqrt{1 - I_x}\sqrt{1 - I_y}}{1 - I_{xy}} \\ &= \rho_{xy}\frac{\sqrt{1 - I_x}\,\sqrt{1 - I_y}}{1 - I_{xy}} \end{aligned} \tag{6.4}$$

Thus the correct attenuation factor taking into account between-variable correlation for the deviations is

$$\frac{\sqrt{1 - I_x}\,\sqrt{1 - I_y}}{1 - I_{xy}} \tag{6.5}$$

When $I_{xy} = 0$, the simple standard result (6.2) holds. The evidence from the Peru data indicates that $I_{xy} > 0$ for some combinations of variables, although it tends to be smaller than I_x or I_y for the variables most seriously affected by similar response variance.

In estimating the regression coefficient β_1 in the model $y = \beta_0 + \beta_1 x$ a similar result holds; in this case, the attenuation is

$$\frac{1 - I_x}{1 - I_{xy}} \tag{6.6}$$

It is worth emphasising that the results (6.5) and (6.6) arise from a consideration of the effect of the simple response variance on estimates of correlation and regression coefficients. The attenuation is expressed in terms of the simple total variance and covariance of the observations. The presence of correlated sampling variance and correlated response variance may have an equal or greater impact on the estimates. There are two kinds of effects which may occur. The correlated components of variance will almost certainly increase the variance of the estimates; some results on the impact of the correlated sampling variance on measures of association (see, for example, Fellegi 1980) suggest that the effect may be severe. The correlated errors may also cause a change in the expected value of the estimator analogous to the attenuation described above.

6.3 INTERVIEWERS' ASSESSMENT OF RESPONSES

At two stages during the course of the interview the interviewers are instructed to record their observations on an aspect of the respondent's replies to the question. Immediately after completing the birth history section of the questionnaire, and before asking the questions dealing with contraception, the interviewer is asked to tick one of the three boxes indicating the *reliability of the answers given in the birth history section*; the three categories given are GOOD, FAIR and POOR. The interviewer's instructions suggest guidelines for completing this question. If considerable probing was necessary for determination of the dates of births and pregnancies, or if inconsistencies arose in the answers, or if the interviewer got the impression that the respondent was unsure of the answers, then the POOR box was appropriate. If the interviewer felt that the respondent was not telling the truth, then again the reliability was to be classified as POOR. In the opposite case, the reliability was to be classified as GOOD. In intermediate cases, involving a moderate amount of probing or correcting, the FAIR box was to be used. Once the interview has been completed the interviewer is asked to tick one of four boxes indicating the respondent's *degree of co-operation*; the four categories given are: BAD, AVERAGE, GOOD and VERY GOOD. The interviewer is instructed *not* to complete this section in the presence of the respondent.

In this subsection we look at the extent to which the interviewer's assessments of the respondents are reflected in the magnitudes of the response deviations. For this purpose we use the absolute value of the difference between the responses obtained from the two interviews for an individual as a measure of the response error. The magnitude is therefore the difference in units (months, years, births, etc) between the responses at the first and second interviews. The response deviations themselves would be unsatisfactory since, by definition, they tend to cancel out over groups of individuals. The interviewer's assessments are taken from the first interview in each case.

The results for the total matched sample of 1198 cases in Peru are given in table 16. In the case of the *co-operation* variable no assessment was available for 13 cases. These are excluded from the table.

A very clear pattern emerges from the table. With the exception of only the two binary variables, *whether worked* and *ever-use*, there is a perfect *rank* correlation between the interviewer's assessments and the magnitudes of the response deviations. The differences between the categories are also statistically significant, the level of significance being less than 0.001 in 20 of the 24 comparisons. It is interesting to note that both the reliability and co-operation assessments are effective in differentiating between respondents. Furthermore the assessment of reliability, which is based on the responses in the birth history section, seems also to be relevant to the background variables such as age and age at marriage and even to the attitudinal question on number of children desired.

The number of individuals classified as POOR is small for both the criteria used by the interviewers - less than 4 per cent in each case - but the AVERAGE/FAIR category (REGULAR in the Spanish version of the questionnaire) is also effective in identifying a group with high response variability.

The same analysis was carried out for the urban and rural subclasses and for the five education subclasses described previously. Since the sample sizes are considerably smaller for the subclasses, the BAD group was amalgamated with the FAIR/AVERAGE group for the analysis. The pattern of results persisted for the urban and rural subclasses, and the differences were statistically significant for the fertility variables despite the smaller sample sizes. For all the education subclasses the same pattern emerged, although *reliability* differentiated better than *co-operation* in general. The results were least convincing in the lowest education group.

On balance, the results indicated that the interviewer's assessments are strongly related to the quality of the responses. There is, however, evidence of association between assessments and education, age and place of residence of the respondent. It is not possible to determine completely the extent to which these are the characteristics on which the interviewers base their judgements, but the results for the subclasses suggest that the interviewers' assessments provide a useful further indicator of the quality of the responses.

It would appear that interviewers are reluctant to classify respondents as either POOR or FAIR on either criterion; more than 70 per cent of the respondents were classified as GOOD or better for each assessment.

In the case of *co-operation*, however, where two positive categories GOOD and VERY GOOD were provided, the interviewers were quite successful in differentiating between the two. This suggests that there is scope for extending the categorization used in the assessment of reliability and that a wider choice, particularly of positive categories, might increase the usefulness of the indicator.

A note of caution may be appropriate here. Although the differences observed are large and of substantive significance, the proportion of the total variability in the response deviations which they explain is generally small.

6.4 VARIANCE OF THE VARIANCE ESTIMATORS

It has been emphasized throughout this report that the values of the measures presented in the tables of results are themselves estimates based on the observations in the sample. These values are subject to sampling variability and it is desirable that the magnitude of this variability should be estimated.

The procedure used in this section is the jackknife, first proposed as a method of reducing bias in ratio estimators

Table 16 Magnitude of response deviations cross-tabulated by interviewers' assessments

Variable	Reliability			Co-operation			
	GOOD	FAIR	POOR	VERY GOOD	GOOD	FAIR	POOR
First birth interval	13.6	24.0	25.4	10.0	15.7	22.9	28.1
Last closed interval	6.3	10.9	14.5	5.1	7.2	9.1	23.0
Year of first birth	0.45	0.98	1.53	0.30	0.50	1.08	1.53
Children ever born	0.12	0.28	0.76	0.09	0.15	0.26	0.79
Month of last birth	3.52	7.97	10.63	1.8	4.4	6.8	18.4
Year of last birth	0.30	0.64	0.65	0.16	0.37	0.54	1.26
Desired no of children	1.14	1.47	1.88	0.96	1.25	1.46	2.16
Current age	0.58	0.95	1.31	0.30	0.65	1.05	1.37
Age at marriage	1.20	1.76	2.19	0.93	1.36	1.53	2.50
Year of marriage	0.98	1.84	2.33	0.71	1.21	1.56	2.37
Age group	0.12	0.20	0.20	0.05	0.14	0.21	0.23
Whether worked	0.13	0.13	0.13	0.14	0.12	0.11	0.23
Ever-use of contraception	0.19	0.22	0.09	0.14	0.21	0.21	0.08
Births in past 5 years	0.14	0.22	0.44	0.10	0.15	0.25	0.38
Sample size	932	223	42	280	667	202	35

and now widely used to estimate variances (see, for example, Kish and Frankel 1974, Kalton 1977). The basic steps are as follows:

(i) Divide the sample into a number k of random subgroups. These subgroups could be primary sampling units, or groups of primary sampling units.

(ii) Calculate the value of the measure, u say, leaving out each subgroup in turn. This will give a set of k values of the measure u. Denote these by $u_{-1}, u_{-2}, \ldots, u_{-i}, \ldots, u_{-k}$ where u_{-i} is the value of u for the data ignoring subgroup i.

(iii) Calculate the pseudo-values $(u_{*i}: i = 1, \ldots, k)$

where $u_{*i} = ku - (k-1)u_{-i}$

(iv) Calculate

$$u_* = \frac{1}{k} \sum_i u_{*i}$$

(v) The variance of u_* can be estimated by

$$\text{var}(u_*) = \frac{1}{k(k-1)} \sum_i (u_{*i} - u_*)^2$$

(vi) The estimated standard error of u_* is

$$\text{se}(u_*) = \sqrt{\text{var}(u_*)}$$

(vii) We use $\text{var}(u_*)$ as an estimate of the variance of the measure u.

This procedure can be applied to measures based on the whole sample and also to measures based on subclasses.

The simple response variance σ_ϵ^2

One of the basic measures of response error used in this report is the simple response variance. Table 17 presents the estimates of σ_ϵ^2, and the estimated variance, the estimated standard error and the estimated coefficient of variation of these estimates for the 16 variables previously considered. The results in this table are reassuring. The coefficient of variation of $\hat{\sigma}_\epsilon^2$ is remarkably stable across variables, with most values close to 0.17. The level of the values is satisfactory in that it provides reasonable confidence in the estimated values of σ_ϵ^2.

The estimates of σ_ϵ^2 in table 17 are based on the whole sample of n = 1198 individuals. Many of the estimates used in the report and in WFS analysis are based on subclasses of the sample, where the number of individuals is much smaller. We would therefore expect the variance estimates to be less precise in these cases. Table 18 presents the results of the jackknife estimation of the variance for five important subclasses: rural areas; metropolitan Lima; respondents with no formal education; respondents under 25; and respondents over 45. The five variables presented are chosen to represent different levels of sensitivity to response errors.

An interesting feature of table 18 is the variation in the values of the simple response variance, σ_ϵ^2, across subclasses. For five of the six variables (the exception is *Ever-use of contraception*) the simple response variance is much larger for the rural, uneducated and over 45 subclasses than for the under 25 and metropolitan subclasses. This is in keeping with the results previously discussed in section 5.1, and served as a reminder of the need for caution in extending the results for the total sample to particular subclasses of interest.

The second point about table 18 is that the coefficient of variation σ_ϵ^2 is in general larger for the subclasses than for the total sample. This is not surprising as the estimate of σ_ϵ^2 is based on fewer observations in the case of subclasses than in the case of the total sample, and consequently the variance (or the standard error) of σ_ϵ^2 might be expected to be correspondingly larger.

What is perhaps worth noting is that the coefficient of variation of σ_ϵ^2 is much more stable across subclasses than the simple response variance itself. This is particularly noticeable in the case of *Marital duration*, *Age at marriage*, and *First birth interval*. In fact this is reassuring since it

Table 17 $\hat{\sigma}_\epsilon^2$, var $(\hat{\sigma}_\epsilon^2)$, se $(\hat{\sigma}_\epsilon^2)$, cv $(\hat{\sigma}_\epsilon^2)$ for the 16 variables

Variable	$\hat{\sigma}_\epsilon^2$	var $(\hat{\sigma}_\epsilon^2)$	se $(\hat{\sigma}_\epsilon^2)$	cv $(\hat{\sigma}_\epsilon^2)$
Age	1.1929	0.0335	0.1829	0.15
Children ever born	0.1485	0.0016	0.0395	0.27
Year of first birth	1.5746	0.1939	0.4404	0.28
Age in 5 year groups	0.0819	0.0001	0.0092	0.11
Year of last birth	0.6183	0.0121	0.1100	0.18
Year of marriage	3.0258	0.2833	0.5323	0.18
Marital duration	3.1238	0.2781	0.5274	0.17
Education	0.1270	0.0005	0.0220	0.17
Year of next to last birth	1.4970	0.0613	0.2476	0.17
Births in past 5 years	0.0897	0.0001	0.0076	0.08
Last closed birth interval	163.5427	1388.1861	37.2584	0.23
Age at marriage	3.4571	0.2924	0.5407	0.16
Worked since marriage	0.0664	0.0001	0.0091	0.14
Ever-use of contraception	0.0870	0.0000	0.0057	0.07
First birth interval	182.4751	1020.7824	31.9497	0.18
No. of children desired	2.2091	0.1919	0.4380	0.20

Table 18 $\hat{\sigma}_\epsilon^2$, se $(\hat{\sigma}_\epsilon^2)$ and cv $(\hat{\sigma}_\epsilon^2)$ for six variables for five subclasses

Variable	Children ever born			Year of last birth			Marital duration		
Subclass	$\hat{\sigma}_\epsilon^2$	se $(\hat{\sigma}_\epsilon^2)$	cv $(\hat{\sigma}_\epsilon^2)$	$\hat{\sigma}_\epsilon^2$	se $(\hat{\sigma}_\epsilon^2)$	cv $(\hat{\sigma}_\epsilon^2)$	$\hat{\sigma}_\epsilon^2$	se $(\hat{\sigma}_\epsilon^2)$	cv $(\hat{\sigma}_\epsilon^2)$
Rural	0.2397	0.1252	0.52	0.8097	0.3070	0.38	4.8137	1.0742	0.22
Lima	0.0429	0.0110	0.56	0.3830	0.2024	0.53	1.4974	0.3437	0.23
No education	0.2471	0.1275	0.52	0.8325	0.1230	0.15	6.6647	1.9418	0.29
Under 25	0.0257	0.0081	0.32	0.1019	0.0396	0.39	0.8473	0.1797	0.21
Over 45	0.3176	0.1890	0.59	0.9285	0.2757	0.30	5.3354	1.1533	0.22
All	0.1485	0.0395	0.27	0.6183	0.1100	0.18	3.1238	0.5274	0.17

Variable	Age at marriage			Ever-use of contraception			First birth interval		
Subclass	$\hat{\sigma}_\epsilon^2$	se $(\hat{\sigma}_\epsilon^2)$	cv $(\hat{\sigma}_\epsilon^2)$	$\hat{\sigma}_\epsilon^2$	se $(\hat{\sigma}_\epsilon^2)$	cv $(\hat{\sigma}_\epsilon^2)$	$\hat{\sigma}_\epsilon^2$	se $(\hat{\sigma}_\epsilon^2)$	cv $(\hat{\sigma}_\epsilon^2)$
Rural	5.4189	1.1915	0.22	0.0881	0.0134	0.15	309.6	70.4	0.23
Lima	1.9583	0.5118	0.26	0.0627	0.0104	0.17	96.2	20.5	0.21
No education	6.9489	1.7698	0.25	0.0775	0.0156	0.20	369.8	121.1	0.33
Under 25	0.8628	0.1460	0.17	0.0749	0.0164	0.22	87.9	25.1	0.29
Over 45	6.4465	1.5927	0.25	0.0803	0.0147	0.18	374.4	187.4	0.50
All	3.4571	0.5407	0.16	0.0870	0.0057	0.07	182.5	31.9	0.17

conforms to the theoretical expectation for the variance of a variance estimator of this kind.

In general, if $\hat{\sigma}^2$ is an estimator of σ^2 based on n − 1 degrees of freedom, then

$$\text{var}\,(\hat{\sigma}^2) = \frac{\mu_4 - \mu_2^2}{n} + \frac{2}{n(n-1)}\mu_2^2$$

$$= \sigma^4 \left[\frac{\beta_2 - 1}{n} + \frac{2}{n(n-1)}\right]$$

where $\beta_2 = \mu_4/\sigma^4$ and μ_2, μ_4 are the second and fourth moments of the parent distribution.

For large n,

$$\text{var}\,(\hat{\sigma}^2) \doteq \frac{(\beta_2 - 1)\sigma^4}{n}$$

and

$$\text{se}\,(\hat{\sigma}^2) \doteq \sqrt{\frac{\beta_2 - 1}{n}} \cdot \sigma^2$$

Consequently

$$\text{cv}\,(\hat{\sigma}^2) \doteq \sqrt{\frac{\beta_2 - 1}{n}}\ \alpha \frac{1}{\sqrt{n}} \qquad (6.7)$$

This derivation implies that the ratio of the cv (σ^2) for the subclasses to the cv (σ^2) for the total sample should be between 1.5:1 and 2.75:1, similar to the ratios found in table 18. Equally satisfying is the fact that even for the subclasses, the coefficients of variation are of the order of 0.2 to 0.3, except for *Children ever born* and *Year of last birth* for which they are larger and less stable. This is understandable since these are the variables with the lowest degree of response variance.

The index of inconsistency, I

The index of inconsistency, I (defined by $\sigma_\epsilon^2/(\sigma_y^2 + \sigma_\epsilon^2)$), measures the proportion of the simple total variance which is due to the simple response variance. The estimates $\hat{I}$ of I obtained from the data are used extensively in section 5.1 to describe the sensitivity of variables to response errors. In figure 1 the values of $\hat{I}$ for the total sample are presented, while table 8 and figures 2A and 2B give the value of $\hat{I}$ for major subclasses. The validity of the conclusions drawn from these tables and figures depend on the precision of the estimates of $\hat{I}$.

Table 19 presents the results of the jackknife estimation of the variance of $\hat{I}$ for the six variables and five subclasses previously considered. The variables span the range of observed values of $\hat{I}$ and the subclasses represent the extremes of the characteristics considered.

The pattern of variation in the values of $\hat{I}$ is similar to that for σ_ϵ^2. The variables are arranged in order of increasing $\hat{I}$ overall - for *Children ever born* $\hat{I}$ is 0.02; for *Year of last birth*, 0.03; for *Marital duration*, 0.04; for *Age at marriage*, 0.20; for *Ever-use of contraception*, 0.35; and for *First birth interval*, 0.56.

The coefficients of variation for the estimates of I for the subclasses are of the same order of magnitude as those for σ_ϵ^2. The least stable estimates are those for the first two variables - the level and variability of the cv's are high for these variables. For the remaining four variables the situation is more satisfactory. The range of the cv's is 0.11 to 0.34 with an average value near 0.22. These are comparable to the corresponding values for σ_ϵ^2, and justify some confidence in the conclusions reached on the basis of a comparison of the $\hat{I}$ values for subclasses. Three examples are given below; these are differences commented on in the text of section 5.1 after table 8. In general, the variance of the difference between two random variables x_1 and x_2 is

$$\text{var}\,(x_1 - x_2) = \text{var}\,(x_1) + \text{var}\,(x_2) - 2\,\text{cov}\,(x_1, x_2)$$

Table 19 $\hat{I}$, se ($\hat{I}$) and cv ($\hat{I}$) for six variables and five subclasses

Variable	Children ever born			Year of last birth			Marital duration		
Subclass	$\hat{I}$	se ($\hat{I}$)	cv ($\hat{I}$)	$\hat{I}$	se ($\hat{I}$)	cv ($\hat{I}$)	$\hat{I}$	se ($\hat{I}$)	cv ($\hat{I}$)
Rural	0.023	0.0156	0.68	0.056	0.0220	0.39	0.066	0.0110	0.17
Lima	0.005	0.0017	0.32	0.129	0.0067	0.52	0.020	0.0044	0.22
No education	0.023	0.0145	0.63	0.040	0.0098	0.25	0.115	0.0347	0.30
Under 25	0.013	0.0057	0.43	0.081	0.0353	0.44	0.116	0.0293	0.25
Over 45	0.275	0.0164	0.60	0.029	0.0098	0.33	0.236	0.0569	0.24
Variable	**Age at marriage**			**Ever-use of contraception**			**First birth interval**		
Subclass	$\hat{I}$	se ($\hat{I}$)	cv ($\hat{I}$)	$\hat{I}$	se ($\hat{I}$)	cv ($\hat{I}$)	$\hat{I}$	se ($\hat{I}$)	cv ($\hat{I}$)
Rural	0.363	0.081	0.22	0.542	0.145	0.27	0.763	0.084	0.11
Lima	0.109	0.030	0.28	0.329	0.059	0.18	0.289	0.071	0.25
No education	0.402	0.074	0.18	0.434	0.107	0.25	0.833	0.139	0.17
Under 25	0.157	0.034	0.22	0.304	0.066	0.22	0.569	0.086	0.15
Over 45	0.267	0.092	0.34	0.372	0.076	0.20	0.672	0.227	0.34

For the differences discussed here, the model for the simple response variance implies that the covariance term is zero. Hence,

$$\text{var}(\hat{I}_1 - \hat{I}_2) = \text{var}(\hat{I}_1) + \text{var}(\hat{I}_2).$$

Table 20A gives the computations for those comparisons of values of $\hat{I}$. The last two columns give the difference in $\hat{I}$ for the subclasses and the estimated standard error of this difference.

For the three contrasts given in table 20A the estimated precision of the estimated difference is sufficiently high to warrant the conclusion that there is a real difference in the values of the index of inconsistency in these cases. It would be unwise however to have too much faith in the absolute value of the difference estimated. If we were justified in constructing a normal 95 per cent confidence interval for the difference in I between women in rural areas and those in Lima for the first birth interval, the confidence interval would be 0.474 ± 0.216 or (0.258, 0.690).

Furthermore, not all the apparent differences in values of $\hat{I}$ are estimated precisely enough to justify much confidence. An example is given in table 20B.

The difference in the values of $\hat{I}$ is 0.12; in fact the estimate of I for the youngest subclass is less than half that for the oldest subclass. The estimated variance for the difference, however, suggests that this apparent difference may result simply from the sampling variance of the estimates involved. The estimated standard error is equal to more than half the estimated difference, and thus a 95 per cent normal confidence interval would be:

$$0.120 \pm 0.125 \text{ or } (-0.005, 0.245).$$

This does not mean that there is no difference between the values of the index of inconsistency for the two subclasses for *marital duration*. It does mean, however, that additional evidence would be necessary before the presence of the difference can be established beyond reasonable doubt. In the case of this particular difference the consistency of the

Table 20A Standard errors of contrasts of $\hat{I}$ values for subclass pairs

Variable	Subclass	$\hat{I}$	se ($\hat{I}$)	var ($\hat{I}$)	var ($\hat{I}_1 - \hat{I}_2$)	se ($\hat{I}_1 - \hat{I}_2$)	$\hat{I}_1 - \hat{I}_2$
Age at marriage	Rural	0.363	0.081	0.00659	0.00750	0.087	0.0254
	Lima	0.109	0.030	0.00091			
Age at marriage	No education	0.402	0.074	0.00546	0.00570	0.075	0.332
	7+ years	0.070	0.015	0.00024			
First birth interval	Rural	0.763	0.084	0.00713	0.01211	0.110	0.474
	Lima	0.289	0.071	0.00498			

Table 20B A counter-example to table 20A

Variable	Subclass	$\hat{I}$	se ($\hat{I}$)	var ($\hat{I}$)	var ($\hat{I}_1 - \hat{I}_2$)	se ($\hat{I}_1 - \hat{I}_2$)	$\hat{I}_1 - \hat{I}_2$
Marital duration	Under 25	0.116	0.029	0.00086	0.00409	0.064	0.120
	Over 45	0.236	0.057	0.00323			

pattern of response variance across age subclasses is so marked that the contrast for a particular variable receives support from the contrast for other variables. Indeed the only exceptions to the pattern of variation in $\hat{I}$ are explicable in terms of the constraints on the simple sampling variance in these cases.

Discussion

All the measures of response variability presented in this report are estimates based on the sample of respondents observed in the main survey and the re-interview survey, and are thus themselves subject to sampling variance. In this section two of the basic measures of response variability are considered - the simple response variance σ_ϵ^2 and the index of inconsistency I. The procedure used to estimate the variance of the estimates is the jackknife, a general method applicable to any measure.

The results are encouraging and indicate that the precision of the estimates is sufficiently high to justify statements about the general level of response errors and to confirm broad patterns of variation across variables and across subclasses. It is clear from the computations however that not all apparent differences in level are sufficiently supported by the evidence - table 20B provides an example.

No attempt has been made to derive estimates of variance for the estimated components of correlated variance. The technical problems caused by the disruption of the fieldwork execution in Peru complicated the estimation of the correlated response variance and made the estimation of precision impracticable. The data from Lesotho, Turkey and Dominican Republic will provide an opportunity to rectify this omission.

References

Bailar, B.A. (1968). Recent Research in Reinterview Procedures. *J. Amer. Statist. Assoc. 63*: 41-63.

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.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educ. and Psych. Meas. 20*: 92-102.

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

Fellegi, I.P. (1980). Approximate Tests of Independence and Goodness of Fit Based on Stratified Multistage Samples. *J. Amer. Statist. Assoc. 75*: 261-8.

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

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.N. Hurwitz and W.G. Madow (1953). *Sample Survey Methods and Theory*. New York: Wiley.

Kalton, G. (1977). Practical Methods for Estimating Survey Sampling Errors. *Bull. Int. Statist. Inst. 47/3.*

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

Kish, L. (1965). *Survey Sampling.* New York: Wiley.

Kish, L. and M.R. Frankel (1974). Inference from Complex Samples. *J. Roy. Statist. Soc. B 36*: 1-37.

Kish, L., R.M. Groves and K. Krotki (1976). Sampling Errors for Fertility Surveys. *WFS Occasional Papers* no 17.

Landis, J.R. and G. Koch (1976). 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. (1982). Methodology of the Response Errors Project. *WFS Scientific Reports* no 28.

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