Scientific Reports

NUMBER 70 DECEMBER 1984

C.A. O'MUIRCHEARTAIGH

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

INTERNATIONAL STATISTICAL INSTITUTE
Permanent Office. Director: E. Lunenberg

Mailing address:
428 Prinses Beatrixlaan, PO Box 950
2270 AZ Voorburg
Netherlands

WORLD FERTILITY SURVEY
Project Director: Halvor Gille

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 Lesotho 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 Lesotho Fertility Survey. *WFS Scientific Reports* no 70. Voorburg, Netherlands: International Statistical Institute.

Printed in Great Britain at the Alden Press
Oxford London and Northampton

Contents

Preface

Right from the start, the WFS programme 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, was 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. The second report (*WFS Scientific Reports* no 45) dealt with response variance in Peru and the present report covers Lesotho.

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 made important contributions at different stages. In particular, I wish to acknowledge the contribution of the late 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 third 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 error in the data collected in WFS surveys; to estimate the magnitude 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 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. The second report (O'Muircheartaigh 1984) presents the design, implementation and analysis of the study in Peru. This third report presents the corresponding results for Lesotho. The reader is referred to the two earlier reports for the general background and objectives of the project.

1.1 Structure of the Study

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

In the context of the WFS, methodological experimentation is by and large excluded by the very nature of the operation. The primary objective has been to assist countries in obtaining the best possible data from a single operation, which necessarily requires the choice of a study design considered *a priori* to be the most suitable. Thus it is not possible in general to compare different survey procedures in order to ascertain which is superior. A severe constraint is therefore imposed on any investigation of response errors in WFS surveys. It is not possible to interfere with the principles laid down for conduct of the survey by introducing any new or experimental procedures which might reduce the quality of the data collected. Furthermore, given the absence of an external source of validation data, it is not possible to examine the absolute magnitudes of the individual response errors. There are, however, two possible approaches which can provide some information on the magnitude and impact of the errors: *re-enumeration* and *interpenetration*.

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

Certain characteristics of the survey would be constant for the two observations: the subject matter, the questions asked, the field force, the procedures for 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 the factors would be necessary. The re-interviews could, however, provide an opportunity to examine the reliability of the data – the extent to which the application of the same essential survey conditions on two occasions would produce the same or different results. Thus, they would enable us to partition the variability observed in the data into two components, one due to the inherent variability in the variable being measured, the other to the haphazard disturbances introduced into the recorded responses by the observation process itself.

The second approach involves a modification of the survey design. It has been established in other contexts that interviewers may influence in a systematic way the responses they obtain. If this is so for WFS surveys, then the estimates of variability obtained in the usual way for statistics calculated from the sample observations may seriously underestimate the true variance. This component of variance – the correlated response variance due to interviewers – will be present in any statistics calculated from the survey data, but the difficulty in practice is that there is usually no way of estimating it. The problem arises because respondents are usually allocated purposively (or haphazardly) to interviewers and any difference between the results obtained by different interviewers may be due to differences between the individuals they interview rather than 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 the respondents must be allocated randomly to interviewers, so that no systematic difference between the workloads of the interviewers can contaminate the comparison of their results. 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 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 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 procedure of data collection, it would be possible to modify the execution of the survey so that the contribution of the correlated response variance due to interviewers could be estimated and its impact on the survey results assessed.

The basic approach of this project thus involves two elements:

1 *Re-enumeration* A subsample of the respondents in the main survey are re-interviewed under the same (or similar) essential survey conditions. This permits the partitioning of the observed variability of the response 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 is possible to estimate the extent to which the usual estimates of variance underestimate the true variance, and thus to provide a more valid estimate of the total variance of the survey.

The particular design used in the project combines the two procedures of interpenetration and re-enumeration in a way which permits the estimation of some additional parameters of the response errors. The technical aspects of the design, suggested in a paper by Fellegi (1964), are described briefly in section 3. The practical features are discussed in section 2. A full description of the methodology is given in the earlier reports (O'Muircheartaigh 1982 and 1984).

2 The Lesotho Design

The Lesotho Fertility Survey, conducted by the Central Bureau of Statistics in 1977–78, was based on a national two-stage probability sample. Census enumeration areas (of which there are 1066 in the country) were the primary sampling units (PSUs). One hundred PSUs were selected with probabilities proportional to size and a sample of households selected within each PSU such that each household in the population had an equal probability of selection. The PSUs were stratified by ecological zone, population density and size before the first stage of selection. All ever-married women residing (on a *de facto* basis) in the selected households were eligible to be interviewed. In all, 3603 individual interviews were successfully completed, giving an overall response rate of around 88 per cent.

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

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

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

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

No of interviewers in the team	4	5
Interviewers for main survey	1234	12345
Interviewers for re-interviews	4321	54231

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

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

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

2.1 IMPLEMENTATION

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

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

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

3 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 the 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 variable 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.

3.1 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 × 2 table.

About one in five of the respondents (19 per cent) gave inconsistent responses on the two occasions. This 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 14 per cent higher than in the second interview, with 16 per cent of all respondents reporting use in the first interview and not in the re-interview, whereas only 3 per cent reported use in the re-interview and not in the original interview.

This table illustrates two strengths of this direct presentation. First, a comparison of the marginals provides an indication of whether there is any major difference between the results of the two interviews for the whole sample, which serves as a check on the constancy of the essential survey conditions. Secondly, the cells of the cross-tabulation give a vivid picture of the scale of the response deviations for the individual respondents.

Tables 2A and 2B present the cross-classification for *Education*. In the case of table 2A the data are presented for the categories of educational level. The categories are in rank order and the differences between them are of substantive significance.

The level of education reported differed for almost one in four respondents, ie for 142 women. For the great majority of those – 135 – the difference between the two responses amounted to a shift through one educational level. In only seven cases was there a shift through two levels. By observing the marginals of the table, we see that the pattern of results is broadly similar for the two interviews, providing some reassurance that for this variable the essential survey conditions remained constant.

In table 2B the data are presented for number of years' education completed. This is a metric variable for which the size of the discrepancy in each case has a clear and unambiguous meaning. For 63 per cent of the respondents the two observations agree. For only 11 per cent did the discrepancy exceed one year. The pattern of the marginals is still broadly similar – the original interviews and the re-interviews produce comparable distribu-

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

Original interview	Re-interview		
	Yes	No	Total
Yes	67	99	166 (27.3%)
No	16	427	443 (72.7%)
Total	83 (13.7%)	526 (86.3%)	609

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

Original interview	Re-interview			
	1	2	3	Total
1	158	323	3	190
2	29	192	32	253
3	4	42	119	165
Total	190	265	154	609

NOTES: 1: 0–4 years; 2: 5–6 years; 3: 7 years or more. The apparent small discrepancies in the totals in this table and in tables 2B, 3 and 4 are due to rounding of weights.

Table 2B Education in years as reported in the original interview and the re-interviews

Original interview	Re-interview											
	0	1	2	3	4	5	6	7	8	9	10	Total
0	41	2	0	1	1	1	0	0	0	0	0	46
1	0	2	1	2	0	1	0	0	0	0	0	5
2	2	2	4	2	0	1	1	1	0	0	0	12
3	2	0	8	23	6	7	3	2	0	0	0	50
4	2	1	3	7	46	9	9	0	0	0	0	77
5	0	0	1	1	22	76	15	5	0	0	0	120
6	1	0	1	2	1	16	85	27	0	0	0	133
7	0	1	0	1	0	8	32	93	2	0	0	137
8	0	0	0	0	0	0	2	3	2	2	0	9
9	0	0	0	0	0	0	0	1	4	6	0	11
10	0	1	0	0	1	0	0	0	0	1	5	8
Total	47	9	18	39	77	118	147	132	8	9	5	609

tions of years of education completed. Table 2B provides more information about the two sets of responses than does table 2A but the additional detail also has the effect of making the information more difficult to assimilate.

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

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. However, for a variable that seems as ambiguous as this, it is perhaps surprising that any observations differ on the two occasions. Most of the discrepant cases involve a difference of only one, but much larger deviations occur for some respondents. Overall, one in five women reported inconsistently; the greatest inconsistency occurred among respondents at higher parities. 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 an adequate and useful summary of the data are illustrated by this table. There are 196 cells in the body of the table, of which 182 would indicate a discrepancy between the observations. Only 54 of these cells contain observations and the importance of each of these depends on the size of the discrepancy it represents and the number of cases in the cell. To discuss each of the occupied cells in turn would, however, be both lengthy and uninformative. The inappropriateness of such a procedure is confirmed by the fact that we wish also to describe the reliability of the data for subclasses of the sample. Thus we will certainly be obliged 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 strikingly 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

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	Total
0	64	5		1			1	1							72
1	2	92	12	4	1	1	1		1						114
2	2	4	84	4	2		1	1	1						99
3	1	1	2	67	5	1	1								78
4	1	1		1	52	11			1						67
5			1		2	46	6			1					56
6				1	1	5	28	6	1			1			43
7	1					1	3	24	3	1					33
8		1		2				3	11	1					18
9									3	11	1				15
10									1	1	5	2	2		11
11												1			1
12													2		2
13														1	1
Total	71	104	99	80	63	65	41	35	22	15	6	4	4	1	609

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

Original interview	Re-interview																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	16	17	98	Total
1	1		1			1				1								4
2		5		1	1	1	2	1										11
3		4	12	10	2	8		2		1	1							40
4		4	6	68	18	18	5	4	3	5	1						4	136
5		2	2	6	28	13	3	2	1	5		1			1		6	70
6	1	2	3	22	7	60	11	9	1	7	2	4	1	1			4	134
7	1			3	1	7	16	6	2	4	1	1		1			7	50
8			2	2	2	11	3	22	5	11		1					1	59
9				1		1		1	6	2								11
10		1	1	3	2	4	1	5	1	33	1	6	3			1	4	66
11						1					0							1
12									1	3	1	3		1				9
13												1	1					2
14			1				1							1				3
20				1													1	2
98			1	1	1	1		3		1							4	12
Total	3	18	29	118	62	125	41	55	20	73	6	18	5	4	1	1	30	609

variables considered in tables 1–3 in that its true value may change between the two interviews. In fact, only two in five women gave identical responses on the two occasions. The discrepancies are large and the overall impression is of very unreliable reporting. Of course 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. About 60 per cent of the respondents reported the number of children desired within one child in the two interviews. It is also worth noting that the marginal distributions are relatively stable and the means of the two distributions are very close.

The tables presented in this section illustrate both the strengths and the weaknesses of this kind of analysis. For the subject matter specialist it is clearly important to look in detail at the pattern of individual response deviations. The only satisfactory way of doing this is to cross-tabulate the two sets of responses for each variable. Tables such as tables 1–4 provide an opportunity to examine the deviations in the context of the values obtained from the two interviews and thus allow the analyst to investigate the underlying response process. But the tables are relatively unwieldy and cannot in practice be presented and examined 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. There is conflict between detail and assimilation. In the next subsection the simplest summary measures are presented.

3.2 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 the 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*

$$A = \sum p_{ii} \tag{3.1}$$

which is the proportion of the cases classified identically by the two interviews. This index has considerable descriptive value, as does its complement, the index of crude disagreement

$$D = 1 - A. \tag{3.2}$$

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

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

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

$$\kappa = 1 - \frac{\text{observed disagreement}}{\text{expected disagreement}} \tag{3.3}$$

$$= 1 - \frac{1 - p_o}{1 - p_e} = \frac{p_o - p_e}{1 - p_e}.$$

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

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

giving

$$\sum_i (p_{ii} - p_{i.}p_{.i})/(1 - \Sigma p_{i.}p_{.i}). \qquad (3.4)$$

While (3.4) is a more appropriate measure of reliability, 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 (3.4) 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:

$$B = \frac{2}{\pi}\cos^{-1}\left[\sum_i (p_{i.}p_{.i})^{\frac{1}{2}}\right] \qquad (3.5)$$

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

The measures (3.1)–(3.5) 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 (3.1) by defining 'agreement' to mean that the two interviews obtain values within some acceptable distance (k units) of each other.

$$A_k = \sum_{|i-j| \leqslant k} p_{ij} = 1 - D_k \qquad (3.6)$$

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:

$$\kappa_w = \frac{p_o^* - p_e^*}{p_l - p_e^*} \qquad (3.7)$$

where

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

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

$$w_{ij} = 1 - (i - j)^2. \qquad (3.8)$$

Under observed marginal symmetry, κ_w with weights (3.8) 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 (3.1) and (3.3) that for a reasonably consistent set of data $A \doteq p_o \gg p_e$ so that $\kappa \doteq A$. Hence little is gained by introducing κ in such cases.

Table 5 Values of D, A, κ and κ_w (values $\times$ 100)

Variable	D	A	κ	κ_w
Years of education	35	65	58	87
Children ever born	19	81	78	92
Ever-use of contraception	19	81	42	42
Current age	40	60	58	94
Age in five-year groups	17	83	80	93
Age at marriage	51	49	43	69
Year at marriage	29	71	70	93
Marital duration (years)	40	60	59	93
Births in past five years	18	82	74	85
No of children desired	54	46	36	51
First birth interval (months)	52	48	47	41
Last closed birth interval (months)	50	50	49	66
Year of first birth	25	75	74	94
Month of first birth	36	64	64	94
Year of last birth	23	77	73	96
Month of last birth	35	65	65	96
Year of next to last birth	28	72	69	94
Month of next to last birth	44	56	56	94

Among the variables in this class in table 5 the three most unreliable are *Age at first marriage*, the *First birth interval* and the *Last closed birth interval*. All three are composite variables derived from two or more questions, each of which is subject to error. Even so, the degree of unreliability gives cause for concern. Fully half of the respondents gave different responses on the two occasions, and the correlation between the two sets of responses is between 0.4 and 0.7.

For a further set of six variables the level of disagreement between the responses is also high, with about one-third of the individuals giving inconsistent responses. These variables are essentially dependent on single dates reported in the interviews.

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 the past five years*. This is reassuring, although even for these variables the responses for the two interviews are by no means perfectly consistent. Almost one in five differs on the two occasions. Another variable which performs well is *Age group*, where the respondents are classified in five-year age groups. The difference between the apparent reliability of *Age* and *Age group* arises from the fact that many of the discrepancies in the age variable nevertheless do not cause the individual to cross the boundary of the age group. It is worth noting that even for this variable one in six of the women is classified in a different age group in the two interviews.

Two variables deserve special mention. The dichotomy included in the table – *Ever-use of contraception* – performs well except in terms of κ and κ_w. Since this variable has only two categories, each discrepancy receives considerable weight in the computation. This is appropriate since a discrepancy represents a complete misclassification on one occasion. The other interesting variable is *Number of children desired.* This is one of the few attitudinal items in the questionnaire, and may be expected to be particularly sensitive to response variability. The full cross-tabulation of the responses for this variable was given in table 4.

The measure κ_w does not seem from the table to be particularly useful. Where the range of the variable is very wide, as it is for many of these variables, the discrepancies, while substantively serious, are small in comparison. In such cases 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 responses (ie those for the original interviews and the re-interviews).

For eleven of the variables the value of κ_w is greater than 0.9. Of the remaining seven variables two have values of κ_w around 0.85; in both cases the distribution of responses is skewed and this also affects the correspondence between A and κ. For the remaining five variables the value of κ_w is relatively low. However, the complexity of the measure means that no single straightforward general proposition can be put forward as an interpretation of the values of κ_w. In general, though, a low value for κ_w does imply that the variable concerned is subject to considerable unreliability in the responses.

4 Components of the Total Variance

4.1 INTRODUCTION – SIMPLE VARIANCE

The conventional measures of reliability described and used in section 3 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.

A particular survey is regarded as a single trial, ie the survey is regarded as conceptually repeatable. An observation for the jth element in the 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:

$$y_{jt} = y_j + \varepsilon_{jt} \tag{4.1}$$

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 (ε_{jt}) the model is completely specified. The distribution of the (ε_{jt}) is called the ξ-distribution. The objective of the survey is to estimate the population mean

$$\bar{y} = \sum_{j=1}^{N} y_j/N. \tag{4.2}$$

The sample mean of the observations is

$$\bar{y}_{.t} = \frac{1}{n}\sum_{j\varepsilon s} y_{jt}. \tag{4.3}$$

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 $y_{.t}$ is

$$V_p(\bar{y}_{.t}) = (1 - f')\frac{\sigma_y^2}{n} \tag{4.4}$$

where

$$f' = \frac{n-1}{N-1}.$$

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

$$V_p(\bar{y}_{.t}) = \frac{\sigma_y^2}{n}. \tag{4.5}$$

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 (4.4) is the *simple sampling variance* (SSV).

In the case of the Lesotho 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 (ε_{jt}) in (4.1)). 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

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

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

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

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

$$V_\eta(\bar{y}_{.t}) = \frac{\sigma_\varepsilon^2}{n}. \quad (4.6)$$

The variance in (4.6) 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 $(\varepsilon_{j_1} - \varepsilon_{j_2})$, ie the *difference* between the response deviations for individual j on the first and second occasions. The variance of $(\varepsilon_{j_1} - \varepsilon_{j_2})$ can be estimated simply and is

$$\sigma_{\varepsilon_{1,2}}^2 = \sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2 - 2\sigma_{\varepsilon_1\varepsilon_2}.$$

If we assume, not unreasonably, that $\sigma_{\varepsilon_1}^2 = \sigma_{\varepsilon_2}^2 = \sigma_\varepsilon^2$, this gives

$$\sigma_{\varepsilon_{1,2}}^2 = 2\sigma_\varepsilon^2(1 - \rho_{\varepsilon_1\varepsilon_2}). \quad (4.7)$$

We estimate σ_ε^2 by

$$\hat{\sigma}_\varepsilon^2 = \tfrac{1}{2}\hat{\sigma}_{\varepsilon_{1,2}}^2 .$$

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 $\hat{\sigma}_\varepsilon^2$ underestimates the simple response variance in the survey by a factor $(1 - \varrho_{\varepsilon_1,\varepsilon_2})$. 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.

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

$$\frac{\sigma_y^2}{n} + \frac{\sigma_\varepsilon^2}{n} \quad (4.8)$$

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 (ε_{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

$$E\left(\frac{s^2}{n}\right) = \left(\frac{\sigma_y^2}{n}\right) + \left(\frac{\sigma_\varepsilon^2}{n}\right) \quad (4.9)$$

where

$$s^2 = \sum_{j=1}^{n} \{(y_{jt} - \bar{y}_{.t})^2/(n - 1)\}. \quad (4.10)$$

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

$$I = \frac{\sigma_\varepsilon^2}{\sigma_y^2 + \sigma_\varepsilon^2}. \quad (4.11)$$

This index measures the proportion of the simple total variance (4.8) which may be attributed to the simple response variance (4.6). 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. This section gives two examples of the estimation of the components of the simple total variance and of the index of inconsistency, 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.

From table 7A, both 0.782 and 0.836 are estimates of the simple total variance, whereas 0.243 is an estimate of the simple response variance multiplied by two. Thus the estimate of σ_ε^2 is

$$\hat{\sigma}_\varepsilon^2 = \tfrac{1}{2}(0.243) = 0.1215.$$

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

$$\widehat{\sigma_y^2 + \sigma_\varepsilon^2} = \tfrac{1}{2}(0.782 + 0.836) = 0.809.$$

In order to obtain a single estimate of σ_y^2 from the two sets of observations σ_y^2 is obtained by subtracting $\hat{\sigma}_\varepsilon^2$ from the estimate of $\sigma_y^2 + \sigma_\varepsilon^2$.

Thus,

$$\hat{\sigma}_y^2 = \tfrac{1}{2}(0.782 + 0.836) - \tfrac{1}{2}(0.243) = 0.6875.$$

From table 7B, the corresponding estimates for ever-use of contraception are:

$$\hat{\sigma}_\varepsilon^2 = \tfrac{1}{2}(0.170) = 0.085$$

$$\widehat{\sigma_y^2 + \sigma_\varepsilon^2} = \tfrac{1}{2}(0.198 + 0.111) = 0.1545$$

$$\hat{\sigma}_y^2 = 0.1545 - 0.085 = 0.0695.$$

The estimates above provide all the information required for the estimation of the index of inconsistency, I. For births in the past five years:

$$\hat{I} = \frac{0.1215}{0.809} = 0.150.$$

Table 6 Data for estimation of simple variance components for two variables

Value	Original interview	Re-interview	Deviation between original interview and re-interview	Frequency
A *Births in the last five years*				
0	215	204	−3	1
1	217	209	−2	4
2	155	168	−1	69
3	22	26	0	501
4	1	3	1	30
			2	4
			3	1
B *Ever-use of contraception*				
0	446	531	−1	14
1	163	78	0	496
			1	98

For ever-use of contraception:

$$\hat{I} = \frac{0.085}{0.1545} = 0.550.$$

This procedure makes use of all available data. Instead of using the matrix containing the full cross-classification of the responses from the two sets of interveiws (examples are given in tables 1–4 in section 3), which becomes unwieldy when the number of categories is large, the data are used in the form given in tables 6 and 7. All the characteristics of the simple total variance can be derived from these distributions.

Table 8 presents the components of the simple total variance for eighteen key variables, arranged in order of increasing values of $\hat{I}$. The variables show a very wide range of values, but some interesting conclusions may be drawn from the table.

In assessing the meaning of the values of $\hat{I}$ presented in table 8 it is important to bear in mind that $\hat{I}$ is a ratio of two variances. The numerator $\hat{\sigma}_\varepsilon^2$ is the simple response variance and is a measure of the magnitude of the inconsistencies in the responses. The denominator $\hat{\sigma}_y^2 + \hat{\sigma}_\varepsilon^2$ is the simple total variance, which measures the total variability in the observations. The size of the ratio $\hat{I}$ therefore depends critically on the size of $\hat{\sigma}_y^2$, and values of $\hat{I}$ may be misinterpreted unless the analyst is aware that the value of $\hat{\sigma}_y^2$ may well be different even for variables which are measured in the same units. The index $\hat{I}$ measures the proportion of the total variability in the responses which is due to disturbances introduced into the observations by the measurement process itself.

Somewhat surprisingly, the six variables with the lowest values of $\hat{I}$ are the individual dates obtained for the first, next to last and last births. The values of $\hat{I}$ range from 0.037 to 0.062. There is a consistent increase in $\hat{I}$ as the births become more distant in time. In fact, the deterioration in reporting is more severe than that indicated by $\hat{I}$ itself. An inspection of the simple response variance ($\hat{\sigma}_\varepsilon^2$) shows that for the dates in years the value rises from 1.1 through 1.6 to 4.5 and for the dates in months from 171 through 244 to 632. The relatively slight increase in $\hat{I}$ for date of first birth is due to the large value of $\hat{\sigma}_y^2$ for the first birth. This is inevitable, since the possible range of values for date of first birth is considerably wider for the total sample of women aged 15–49 than the range of values for date of last birth. However, overall, these six variables seem to be reliably reported.

The next four variables are central to much of the analysis of WFS data. Age, age in five-year groups, year of first marriage and marital duration are all used widely as classification variables. All four have values of $\hat{I}$ near 0.07 and seem on this basis to be measured reliably. This issue, however, is considered in more detail in the following sub-section.

Table 7 Components of the simple variance for two variables

Measure	Original Interviews	Re-interviews	Deviations
A *Births in the last five years*			
Mean	0.977	1.041	−0.063
Standard deviation	0.884	0.914	0.493
Standard error	0.036	0.037	0.020
Variance	0.782	0.836	0.243
B *Ever-use of contraception*			
Mean	0.267	0.129	0.138
Standard deviation	0.445	0.333	0.412
Standard error	0.018	0.014	0.017
Variance	0.198	0.111	0.170

Table 8 Components of the simple response variance for eighteen variables

No	Variable	$\hat{\sigma}_y^2$	$\hat{\sigma}_\varepsilon^2$	$\hat{I}$
1	Year of last birth	29.06	1.118	0.037
2	Month of last birth	4210	171.3	0.039
3	Year of next to last birth	28.77	1.649	0.054
4	Month of next to last birth	4147	244.0	0.056
5	Year of first birth	67.79	4.456	0.062
6	Month of first birth	9752	631.7	0.061
7	Age	78.55	5.168	0.062
8	Age in five-year groups	3.122	0.2325	0.070
9	Year of marriage	74.88	5.973	0.074
10	Marital duration	75.18	5.964	0.074
11	Children ever born	6.180	0.5580	0.083
12	Years of education	4.277	0.6460	0.131
13	Births in past five years	0.6875	0.1215	0.150
14	Age at first marriage	7.017	3.102	0.306
15	Last closed birth interval	324.9	185.8	0.364
16	No of children desired	3.523	3.248	0.480
17	Ever-use of contraception	0.0695	0.0850	0.550
18	First birth interval	234.9	330.8	0.585

The group of three variables following the date-based variables have moderate values of $\hat{I}$. Two of them – children ever born and births in the past five years – are variables of crucial importance to the analysis of fertility. It may seem a little surprising that the more recent data have a larger value of $\hat{I}$ than the more general variable children ever born. This arises from two factors. First, births in the past five years depends not only on births being reported but also on the dates of these births, thus adding a potential source of inconsistency. Secondly, the value of $\hat{\sigma}_y^2$ is much larger for children ever born, so that although the simple response variance is much lower for births in the past five years, the value of $\hat{I}$ is almost twice as large.

The remaining five variables have high values for $\hat{I}$. In each case the proportion of the simple total variance due to the simple response variance is over 30 per cent. Three of the variables are based on dates and it is remarkable that the values of $\hat{I}$ are so large given the apparent reliability of the individual date variables. Table 9 gives the relevant data.

The first of these variables is *Age at first marriage*, which is derived directly from year of marriage (9) and year of birth (7). It is worth examining in some detail the way in which the observed value of $\hat{I}$ arises.

The value of $\hat{\sigma}_y^2$ for age at marriage is the variance of the difference between the 'true' values of year of marriage and year of birth. Since these two variables are highly correlated, the value of $\hat{\sigma}_y^2$ for age at marriage is much lower than the value of $\hat{\sigma}_y^2$ for either of them. This automatically implies that the denominator in $\hat{I}$ for age at marriage will be much smaller than that for the two variables from which it is derived.

Similarly, the simple response variance for age at first marriage is the variance of the difference between the response deviations on the other two variables. In this case there are no *a priori* grounds for expecting a correlation between the two sets of response deviations. If there were no correlation we would obtain a value of $\hat{\sigma}_\varepsilon^2$ of approximately 11.1. In fact, $\hat{\sigma}_\varepsilon^2$ is 3.1. This indicates that to some extent the reporting of age at marriage is of higher reliability than could be expected from the quality of the two dates from which it is derived. The value of $\hat{I}$ obtained is, however, a valid measure of the reliability of the variable itself.

For the two birth interval variables the same general structure emerges. The values of $\hat{I}$ obtained are due to a dramatic reduction in $\hat{\sigma}_y^2$ counterbalanced in part by a decrease in $\hat{\sigma}_\varepsilon^2$. For both variables, but particularly for the first birth interval, the values of $\hat{I}$ are very large and indicate a high degree of unreliability.

The other two variables in table 8 are of a different

Table 9 Results relating to the reliability of age at first marriage, last closed birth interval and first birth interval

Variable	$\hat{\sigma}_y^2$	$\hat{\sigma}_\varepsilon^2$	$\hat{I}$	Variable	$\hat{\sigma}_y^2$	$\hat{\sigma}_\varepsilon^2$	$\hat{I}$
Year at marriage	74.88	5.973	0.074	Age at marriage	7.02	3.102	0.306
Year at birth	78.55	5.168	0.062				
Month of last birth	4210	171.3	0.039	Last closed birth interval	325	185.8	0.364
Month of next to last birth	4147	244.0	0.056				
Date of marriage	10707*	842.2*	0.073*	First birth interval	235	330.8	0.585
Month of first birth	9752	631.7	0.061				

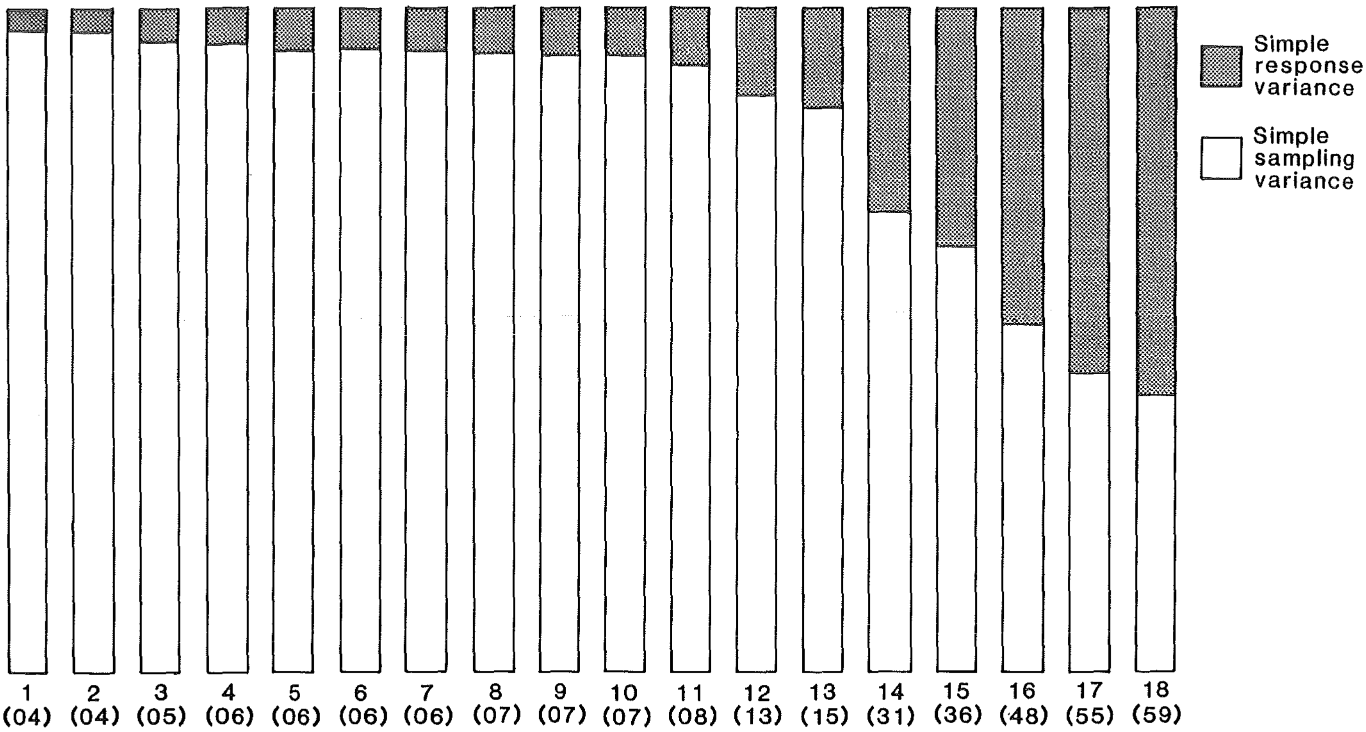

Figure 1 Simple total variance (the variable numbers correspond to those in table 8)

type. *Number of children desired* is an attitudinal variable and thus might be expected to be particularly sensitive to the measurement process. Both this and *Ever-use of contraception* are variables for which the true value could change in the time period between the two interviews. Because of the form in which the data were collected, it is not possible to adjust these variables to take changes into account. Both show a high degree of inconsistency between the two interviews.

Figure 1 gives a diagrammatic presentation of the components of the simple total variance for the variables in table 8.

Reliability of different categories of respondents

The results in the previous section give an overall view of the magnitude and pattern of the response variance for the variables considered. It is important, however, to bear in mind that most of the analysis of the survey data will be carried out on subsets of the sample, ie subclasses of the population. In this section two important sets of subclasses are examined: education subclasses and age groups. The classification value in each case is taken as reported in the first interview.

Education subclasses

Table 10 presents the values of $\hat{\sigma}_y^2$, $\hat{\sigma}_\varepsilon^2$ and $\hat{I}$ for three education subclasses – those with 0–4 years of education, those with 5–6 years and those with seven or more years. In some ways it would have been preferable to use instead the categories no education, 1–5 years and six or more years, but the number of cases in the no education category would then have been too small to give reasonably stable results.

In evaluating the results in this table two different criteria can be considered. Since we are comparing the reliability across subclasses within each particular variable the simple response variance σ_ε^2 is in some ways the most appropriate measure. However, the implications of a particular value of σ_ε^2 in analyses involving more than one variable arise from the magnitude of the index of inconsistency $\hat{I}$, which depends also on the variability of the true values in the subclass.

The most striking feature of table 10 is that for all but two variables the value of $\hat{\sigma}_\varepsilon^2$ is largest for the lowest education group. This is consistent with our expectation that the quality of responses rises with level of education. The value of $\hat{\sigma}_\varepsilon^2$ for those with little education is typically between 40 and 100 per cent larger than $\hat{\sigma}_\varepsilon^2$ for the total sample. In cases where the level of reliability overall is already low this suggests that there is even more cause for concern when the analysis is confined to this subclass. Two variables may be taken as examples. For the *First birth interval* the simple response variance for the total sample is estimated to be 331; for the low education subclass the value is 628, almost twice as large. For *Age at first marriage* the values are 3.1 and 4.2 respectively.

When the index of inconsistency is taken as the criterion the results follow the same pattern. This is reassuring since $\hat{I}$ depends not only on $\hat{\sigma}_\varepsilon^2$ but also on $\hat{\sigma}_y^2$, which is also an estimate. For twelve of the variables the value of $\hat{I}$ is largest for the lowest education subclass. Five variables for which this is not the case are the dates of the last and next to last births and the last closed birth interval, which is derived from them. The remaining variable, years of education, is a special case.

For the former set of twelve variables the value of $\hat{I}$ for the lowest education subclass is between 30 and 90 per cent higher than the value of $\hat{I}$ for the total sample. Generally the increase is due to the larger value of $\hat{\sigma}_\varepsilon^2$ for

Table 10 Components of the simple total variance for the total sample and three education subclasses

Variable	Total sample (n = 609)			0–4 years' education (n = 191)			5–6 years' education (n = 253)			7 or more years' education (n = 165)		
	$\hat{\sigma}_y^2$	$\hat{\sigma}_\varepsilon^2$	$\hat{I}$	$\hat{\sigma}_y^2$	$\hat{\sigma}_\varepsilon^2$	$\hat{I}$	$\hat{\sigma}_y^2$	$\hat{\sigma}_\varepsilon^2$	$\hat{I}$	$\hat{\sigma}_y^2$	$\hat{\sigma}_\varepsilon^2$	$\hat{I}$
Year of last birth	29.06	1.118	0.037	43.57	2.058	0.043	23.15	0.3384	0.014	18.30	1.123	0.058
Month of last birth	4210	171.3	0.039	6346	322.0	0.046	3347	52.88	0.015	2633	162.9	0.058
Year of next to last birth	28.77	1.649	0.054	42.50	2.010	0.043	22.86	1.275	0.049	18.92	1.746	0.084
Month of next to last birth	4147	244.0	0.056	6078	302.0	0.046	3337	186.8	0.049	2709	257.8	0.087
Year of first birth	67.79	4.456	0.062	68.20	9.059	0.113	66.87	1.414	0.021	56.63	3.705	0.061
Month of first birth	9752	631.7	0.061	9810	1285	0.112	9602	195.2	0.020	8189	530.1	0.061
Age	78.55	5.168	0.062	76.72	8.751	0.103	81.51	3.443	0.040	68.55	3.211	0.045
Age in five-year groups	3.122	0.2325	0.070	3.069	0.3717	0.109	3.199	0.1626	0.048	2.689	0.1470	0.052
Year of marriage	74.88	5.973	0.074	76.45	10.46	0.118	71.02	3.622	0.048	64.24	4.14	0.061
Marital duration	75.18	5.964	0.074	77.02	10.40	0.117	70.88	3.748	0.050	64.59	4.00	0.058
Children ever born	6.180	0.5580	0.083	6.243	0.8182	0.116	6.290	0.3505	0.053	5.533	0.5760	0.094
Years of education	4.227	0.6460	0.131	2.726	0.7925	0.223	0.2598	0.3332	0.571	0.4195	0.7870	0.652
Births in last five years	0.6875	0.1215	0.150	0.672	0.1668	0.196	0.7111	0.0924	0.114	0.666	0.1230	0.156
Age at first marriage	7.017	3.102	0.306	5.857	4.156	0.424	6.897	3.120	0.313	7.755	1.760	0.185
Last closed birth interval	324.9	185.9	0.364	410.0	148.2	0.255	330.1	233.0	0.411	206.7	155.3	0.429
No of children desired	3.523	3.248	0.480	3.047	4.694	0.639	3.302	2.536	0.433	3.680	2.853	0.437
Ever-use of contraception	0.0695	0.0850	0.550	0.0510	0.0677	0.564	0.0703	0.0782	0.524	0.0880	0.1075	0.550
First birth	234.9	330.8	0.585	121.1	628.2	0.856	335.8	255.5	0.462	244.4	57.52	0.191

Table 11 Components of the simple total variance for two sets of related variables

Variable	Total sample (n = 609)			0–4 years' education (n = 191)			5–6 year's education (n = 253)			7 or more years' education (n = 165)		
	$\hat{\sigma}_y^2$	$\hat{\sigma}_\varepsilon^2$	$\hat{I}$	$\hat{\sigma}_y^2$	$\hat{\sigma}_\varepsilon^2$	$\hat{I}$	$\hat{\sigma}_y^2$	$\hat{\sigma}_\varepsilon^2$	$\hat{I}$	$\hat{\sigma}_y^2$	$\hat{\sigma}_\varepsilon^2$	$\hat{I}$
Month of first marriage	10783	860.0	0.074	11009	1506	0.118	10227	552.0	0.048	9251	596.0	0.061
Month of first birth	9752	631.7	0.061	9810	1285	0.112	9602	195.2	0.020	8189	530.1	0.061
Correlation between response deviations		0.78			0.77			0.64			0.95	
First birth interval	234.9	330.8	0.585	121.1	628.2	0.856	335.8	255.5	0.462	244.4	57.52	0.191
Age	78.55	5.168	0.062	76.72	8.751	0.103	81.51	3.443	0.040	68.55	3.211	0.045
Year of first marriage	74.88	5.973	0.074	76.45	10.46	0.118	71.02	3.622	0.048	64.24	4.14	0.061
Correlation between response deviations		0.72			0.78			0.56			0.76	
Age at first marriage	7.017	3.102	0.306	5.857	4.156	0.424	6.897	3.120	0.313	7.755	1.760	0.185

the subclass. Furthermore, the correlation between the response deviations for related variables carries through to each of the subclasses. Table 11 illustrates this for two sets of variables.

The first set of three variables shows the way in which the reliability of *Age at first marriage* is determined. The two basic variables are *Age* (or year of birth) and *Year of first marriage*. The difference between these is *Age at first marriage*. The reliability of measurement for *Age* and *Year of first marriage* can be expressed as $\hat{\sigma}^2_\varepsilon$. In both cases this is largest for the least educated subclass. The two other subclasses have similar values for $\hat{\sigma}^2_\varepsilon$. When we consider *Age at first marriage* we find that $\hat{\sigma}^2_\varepsilon$ for this variable is much lower than we would expect on the basis of the values for the two variables from which it is constructed. If the response deviations for the two component variables were uncorrelated, then the response variance for *Age at first marriage* would be equal to the sum of the response variances for the other two. It is reassuring to note that this is not the case. For all three subclasses there is a strong positive correlation between the response deviations for *Year of birth* and *Year of first marriage*. Thus respondents are making compensating errors in reporting year of birth and year of first marriage. The absolute size of the response variance is in fact less for *age at first marriage* than for either *Year of birth* or *Year of first marriage* for all subclasses. The correlation between the response deviations for the two latter variables is between 0.6 and 0.8 in each case.

The implications for the reliability of reporting of *Age at first marriage* are considerable. In the absence of the correlation between the response deviations between the two component variables the value of $\hat{I}$ would be about 0.6 for the total sample and between 0.5 and 0.8 for the subclasses. The actual values of $\hat{I}$ are 0.31 for the total sample and 0.19, 0.31 and 0.42 for the most, middle and least educated subclasses respectively. The reason that all the values of $\hat{I}$ are larger for *Age at first marriage* than for *Year of birth* and *Year of first marriage* is that the variance of the true values ($\hat{\sigma}^2_y$) is much smaller for the former.

The other set of three variables shows the same underlying pattern. The situation here illustrates even more dramatically how unwise it is to assume anything about a composite variable on the basis of information about the component variables separately. The two component variables are *Month of first birth* and *Month of first marriage*. Both have low values of $\hat{I}$ – 0.07 for the total sample and between 0.02 and 0.12 for the subclasses. The two variables are used to calculate values of the *First birth interval*. There are two ways in which we can consider the reliability of measurement of the variables. The most basic measure is the simple response variance $\hat{\sigma}^2_\varepsilon$. In the case of all three variables $\hat{\sigma}^2_\varepsilon$ is largest for the least educated subclass – about twice the value for the total sample. Once again the response deviations for the two basic variables are highly correlated. There is a strong element of compensation in the errors in date of *First marriage* and date of *First birth*. In the absence of this correlation values of $\hat{\sigma}^2_\varepsilon$ of about 1500 for the total sample and 2800 for the least educated subclass might be expected. In fact, the values obtained were 331 and 628 respectively. This implies a correlation of about 0.8 between the two sets of response deviations.

The second measure of reliability is the index of inconsistency, $\hat{I}$. This depends not only on $\hat{\sigma}^2_\varepsilon$ but also on $\hat{\sigma}^2_y$. The situation is similar to that described above for the first set of variables. For the total sample $\hat{\sigma}^2_y$ is about 11 000 for date of first marriage and 10 000 for date of first birth. The variation in the first birth interval is, of course, much less – only 235 for the total sample. Consequently the values of $\hat{I}$ are far greater for the first birth interval than for the other two variables despite the smaller values of $\hat{\sigma}^2_\varepsilon$, ranging from 0.19 for the most educated group to 0.86 for the least educated.

These two sets of variables illustrate some important principles in assessing measurement error. First, the reliability – however measured – for a composite variable cannot be predicted in any uniform way from the reliabilities of the component variables. Secondly, there is a high degree of correlation between response deviations on related variables. Thirdly, the overall reliability of a variable is dependent on both the size of the measurement error and the extent of the variability among the true values for the variable concerned.

Table 12 presents the data for the deviant variables in table 10. The variables involved all relate to dates of more recent births. The two component variables are *Date of last birth* and *Date of next to last birth*. The composite variable is *Last closed birth interval*. The pattern of values of $\hat{I}$ across subclasses is inconsistent with that found for the other variables. The values of $\hat{I}$ for the least educated subclass are lower for all three variables than those for the most educated subclass.

The basic reason for this anomaly is the variation in the value of $\hat{\sigma}^2_y$ – the simple sampling variance, or the variation between the true values. The values of $\hat{\sigma}^2_y$ are largest for the least educated and smallest for the most educated, so that although the simple response variance is lower for the more educated respondents, the *relative* reliability is not.

The same strong positive correlation between response deviations illustrated in table 11 for the two component variables is present also in table 12. In this case this would have been expected (or at least hoped for), since the two variables are part of the birth history and are measured by an integrated set of questions in the questionnaire.

Age subclasses

Table 13 presents the values of $\hat{I}$ and of the components of the simple total variance for two age subclasses – those under 25 and those 45 and over. Age subclasses are widely used in the analysis of WFS data and it is important to assess the reliability of reporting for these subsets of the sample.

The most striking feature of table 13 is the contrast between the values of the simple response variance for the two subclasses. The older subclass provides substantially less reliable reporting than the younger one. The simple response variance for the older subclass is typically twice as large as that for the younger subclass, and in some cases the difference is even greater. A possible interpretation of this difference is that in recalling events, the extent of the unreliability is determined at least in part by the time elapsed between the event and the interview. This interpretation is supported by the internal

Table 12 Components of the simple total variance for a composite variable and its elements

Variable	Total sample (n = 609)			0–4 years' education (n = 191)			5–6 years' education (n = 253)			7 or more years' education (n = 165)		
	$\hat{\sigma}^2_y$	$\hat{\sigma}^2_\varepsilon$	$\hat{I}$	$\hat{\sigma}^2_y$	$\hat{\sigma}^2_\varepsilon$	$\hat{I}$	$\hat{\sigma}^2_y$	$\hat{\sigma}^2_\varepsilon$	$\hat{I}$	$\hat{\sigma}^2_y$	$\hat{\sigma}^2_\varepsilon$	$\hat{I}$
Month of last birth	4210	171.3	0.039	6346	322.0	0.046	3347	52.88	0.015	2633	162.9	0.058
Month of next to last birth	4147	244.0	0.056	6078	302.0	0.046	3337	186.8	0.049	2709	257.8	0.087
Correlation between response deviations		0.55			0.76			0.03			0.62	
Last closed birth interval	324.9	185.8	0.364	410.0	148.2	0.255	330.1	233.0	0.411	206.7	155.3	0.429

Table 13 Components of the simple total variance for the total sample and two age subclasses

Variable	Total sample (n = 609)			Age ⩽ 25 years (n = 169)			Age ⩾ 45 years (n = 64)		
	$\hat{\sigma}_y^2$	$\hat{\sigma}_\varepsilon^2$	$\hat{I}$	$\hat{\sigma}_y^2$	$\hat{\sigma}_\varepsilon^2$	$\hat{I}$	$\hat{\sigma}_y^2$	$\hat{\sigma}_\varepsilon^2$	$\hat{I}$
Year of last birth	29.06	1.118	0.037	0.8935	0.351	0.282	48.39	1.695	0.034
Month of last birth	4210	171.3	0.039	120.8	61.27	0.337	7043	259.5	0.035
Year of next to last birth	28.77	1.649	0.054	1.460	1.758	0.546	44.89	1.404	0.030
Month of next to last birth	4147	244.0	0.056	205.1	269.0	0.567	6368	198.5	0.030
Year of first birth	67.79	4.456	0.062	4.930	2.051	0.294	13.40	5.820	0.303
Month of first birth	9752	631.7	0.061	693.7	283.1	0.290	1949	799.2	0.291
Age	78.55	5.168	0.062	6.096	4.258	0.411	2.100	7.255	0.776
Age in five-year groups	3.122	0.2325	0.070	0.204	0.1965	0.491	d.n.a.	d.n.a.	d.n.a.
Year of first marriage	74.88	5.973	0.074	6.018	3.810	0.388	9.453	7.135	0.430
Marital duration	75.18	5.964	0.074	5.891	3.977	0.400	9.586	7.021	0.423
Children ever born	6.180	0.5580	0.083	1.019	0.3160	0.237	7.351	0.5950	0.075
Years of education	4.227	0.6460	0.131	3.190	0.7245	0.185	n.a.	n.a.	n.a.
Births in last five years	0.6875	0.1215	0.150	0.5370	0.1255	0.189	0.2970	0.0475	0.140
Age at first marriage	7.017	3.102	0.306	2.889	1.416	0.329	7.256	2.909	0.286
Last closed birth interval	324.9	185.8	0.364	196.5	64.45	0.247	363.1	296.4	0.449
No of children desired	3.523	3.248	0.480	3.383	2.175	0.391	3.248	3.029	0.480
Ever-use of contraception	0.0695	0.0850	0.550	0.0500	0.0575	0.535	0.0200	0.0655	0.766
First birth interval	234.9	330.8	0.585	196.6	74.27	0.274	167.0	592.8	0.780

evidence, which can be obtained by comparing the observed response variances for different events for the younger subclass. The simple response variance for year of last birth (a recent event) was 0.351; for year of next to last birth (a more distant event) 1.758; and for year of first birth (a still more distant event) 2.051.

An examination of the values of $\hat{I}$ provides significant evidence of the complex nature of the problem of evaluating response reliability. Overall, the pattern is that which could be expected. For ten of the seventeen variables the value of $\hat{I}$ is larger for the older group. Two examples of the expected pattern are *Desired family size* and *First birth interval*. The values of $\hat{\sigma}_y^2$ are approximately stable and the variation in the values of $\hat{I}$ is due to the greater unreliability of responses for the older group. This is essentially the pattern established in table 10 for the education subclasses. The pattern does not, however, hold for other variables.

The first major difference between the results of table 10 and those of table 13 can be illustrated by the variable *Year of first marriage*. The values of $\hat{I}$ for the younger and older groups are 0.39 and 0.43 respectively. The two intermediate age groups (not given in table 13) have similar values. The value of $\hat{I}$ for the total sample is 0.07. Figure 2 gives a diagrammatic representation. The problem here is that the value of $\hat{\sigma}_y^2$ is very different for the subclasses and for the total sample. This is because the restriction of a subclass to a particular age group necessarily reduces the possible variation in *Year of first marriage* considerably compared to the variation in the total sample of women aged 15–49. The value of $\hat{\sigma}_y^2$ for the total sample is 74.88; for the under 25 age group it is only 6.02; for the 45 and over age group it is 9.45. The same phenomenon occurs for *Age at first marriage*, *Marital duration*, *Age at first birth* and, of course, *Age*. These variables are all age-related, and when the subclasses are based on an age categorization the values of $\hat{I}$ for the total sample are no longer an 'average' of the subclass values as they were for the education subclasses considered earlier. The alternative diagrammatic representation in figure 3 illustrates this.

The discussion above raises a fundamental question about the apparent ranking of the variables in terms of reliability indicated by figure 1. The initial impression given by the $\hat{I}$ values is that *Age*, for example, is extremely reliably reported. Table 13 shows that this assessment is crucially dependent on the context in which the variable is used. When the total sample is being considered, or when a subclass is being used which is not age-related, the relative reliability of age reporting is high, as measured by the value of $\hat{I}$. When the analysis is restricted to an age group, however, the situation changes and the measure of a relationship between age and any other variable is severely affected by the response deviations.

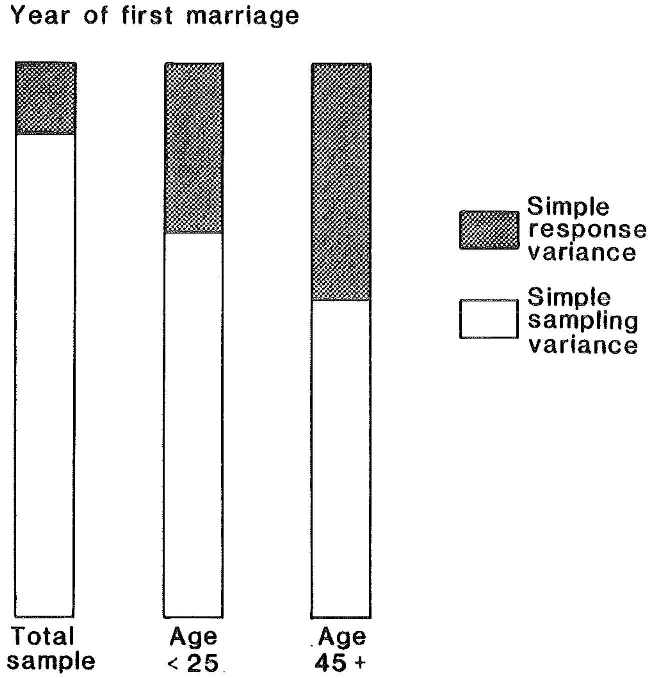

Figure 2 Relative sizes of the components of the simple total variance for two age groups

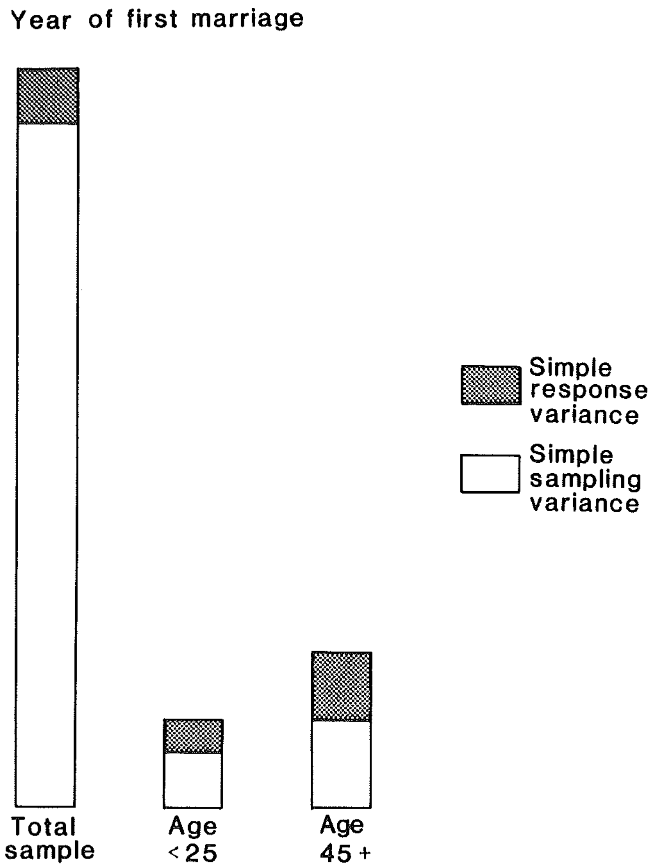

Figure 3 Absolute sizes of the components of the simple total variance for the total sample and two age groups

The same caveat applies to *Marital duration*. It is worth noting that the contrast between the reliability of *Age at first marriage* and *Marital duration* and that of *Age* is transformed by this change of context. For the restricted subclasses the $\hat{I}$ values are all about 0.40, whereas for the total sample both *Age* and *Marital duration* have $\hat{I}$ values near 0.07. In fact, for the age subclasses *Age at first marriage* is more reliable than the other two variables.

The second major difference between the age and the education subclasses can be seen in the group of variables dealing with dates of children's births. For *Date of first birth* and *Date of last birth* the reliability of reporting, as measured by the simple response variance, is much higher for the younger age group. But because of the range of dates to which the measurement refers, the values of $\hat{I}$ are approximately equal for first birth and much higher for the younger age group for date of last birth. The same contrast holds for date of next to last birth.

Two other variables are worth considering. *Number of children desired* has essentially the same pattern as the date of birth variables and the values of $\hat{I}$ reflect this. *Births in the past five years* is different insofar as it is a variable whose main relevance is to the younger group. This is the only variable for which the simple response variance is substantially larger for the younger respondents.

Conclusion

The analysis in this section establishes a clear pattern of response variability across subclasses but also suggests that great care must be taken in using measures of reliability outside the context in which they are calculated. The quality of the data is highest for the younger and for the more educated respondents. This is not necessarily reflected in the values of the index of inconsistency, $\hat{I}$, because of the dependence of $\hat{I}$ on the variation in the true values for the group of respondents under consideration. The most striking illustration of this is given by the analysis of age subclasses, where for age-related variables the values of $\hat{I}$ are all dramatically increased.

The other important finding is that there is a strong element of correlation between response deviations for related variables. Tables 11 and 12 give some examples of this and provide some reassurance about the quality of reporting of intervals.

4.2 CORRELATED VARIANCE – DESIGN EFFECT AND INTERVIEWER EFFECT

Section 4.1 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 factors 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.

Correlated sampling variance (CSV)

Simple random samples are rarely if ever found in practice in field surveys. Most sample designs are stratified multi-stage ones 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 sampling variance can be expressed, ignoring the finite population correction, as

$$V_p(\bar{y}_{.t}) = \frac{\sigma_y^2}{n}\{1 + \text{roh}\,(b - 1)\}. \tag{4.12}$$

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 b is the average number of individuals interviewed in each cluster. The increase in the variance over the simple sampling variance given by (4.5) is

$$\frac{\sigma_y^2}{n}\{\text{roh}\,(b - 1)\} \tag{4.13}$$

and may be called the *correlated sampling variance* (CSV). It is clear from (4.12) and (4.13) that the size of b will have an important impact on the correlated sampling variance and thus on the *total sampling variance* (TSV) which is given by (4.12).

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 (4.12) to the simple sampling variance (4.5) and is

$$\begin{aligned} \text{Deff} &= \frac{(\sigma_y^2/n)\{1 + \text{roh}\,(b - 1)\}}{(\sigma_y^2/n)} \\ &= 1 + \text{roh}\,(b - 1). \end{aligned} \tag{4.14}$$

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 4.1 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 4.1 can be modified to take the possibility of intercorrelated errors into account. The assumptions given in (4.6) can be changed to:

$$\begin{aligned} E_\eta(\varepsilon_{jt}) &= 0 \qquad [\text{all } j] \\ V_\eta(\varepsilon_{jt}) &= \sigma_\varepsilon^2 \qquad [\text{all } j] \\ \text{Cov}_\eta(\varepsilon_{ijt}, \varepsilon_{i'j't'}) &= \begin{cases} \rho_1\sigma_\varepsilon^2 & \text{if } i = i' \\ \rho_2\sigma_\varepsilon^2 & \text{if } i \neq i' \end{cases} \end{aligned} \tag{4.15}$$

In (4.15) 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 (4.15) the contribution of the response deviations to the total variance will be:

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

Ignoring ϱ_2, this becomes

$$V_\eta(\bar{y}_{.t}) = \frac{\sigma_\varepsilon^2\{1 + \rho_1(m - 1)\}}{n} \tag{4.16}$$

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 (4.7) is

$$\frac{\sigma_\varepsilon^2}{n}\{\rho_1(m - 1)\} \tag{4.17}$$

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

Figure 4 Total variance by source and type

Source	Type		
	Simple	Correlated	Total
Sampling	SSV Simple sampling variance: the variance of a simple random sample of size n	CSV Correlated sampling variance: the additional variance due mainly to the clustering of the sample	TSV Total sampling variance: a function of the sample design and the variability among the true values in the population
Measurement (response)	SRV Simple response variance: due to random (haphazard) response deviations caused by the observation or measurement process	CRV Correlated response variance: the additional variance due to inter-relationship between the response deviations caused by, for example, a common interviewer for each group of respondents	TRV Total response variance: a function of the data collection process
Total	STV Simple total variance discussed in section 4.1: the variance of the estimator for a simple random sample with uncorrelated response deviations	CTV Correlated total variance: the additional sampling and response variance neglected by the analysis in section 4.1	TV Total variance: the actual variance of the estimators

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 (4.16) for the *total response variance* and the expression (4.12) 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 their results. 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 Lesotho has been described in section 2.

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,

$$\left.\begin{aligned} E_p E_\eta\{C\} &= \sigma_y^2 + \sigma_\varepsilon^2\{1 + \rho_1(m-1)\} \\ \text{and} \quad & \\ E_p E_\eta\{F\} &= \sigma_y^2 + \sigma_\varepsilon^2(1 - \rho_1) \end{aligned}\right\} \quad (4.18)$$

Hence (1/m) (C-F) provides a possible estimator of $\rho_1\sigma_\varepsilon^2$. In fact, under this model, $E\{(1/m)(C\text{-}F)\} = \sigma_\varepsilon^2(\rho_1 - \rho_2)$, but it is usually recommended as an estimator of $\rho_1\sigma_\varepsilon^2$ since ρ_2 can generally be assumed to be small. See, for example, Hansen, Hurwitz and Bershad (1961), Fellegi (1964) and Kish (1962).

4.3 THE TOTAL VARIANCE

The partitioning of the total variance of the estimator is presented in figure 4. The total variance is shown to be made up 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

Table 14 Estimates of $\rho_1 I$

Variable	Main survey (total sample)	Main survey (designated sample)	Re-interview survey (designated sample)
Ever-use contraception	0.084	0.122	0.119
Years of education	0.015	0.077	0.077
No of children desired	0.041	0.041	0.048
First birth interval	0.045	0.163	0.084
Marital duration	0.000	0.016	−0.005
Age at marriage	0.005	−0.006	0.032

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 20) represents the basic situation with regard to the variance of the estimators.

4.4 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 Lesotho, as in the other countries in which the Response Errors Project was carried out, the overall design included both random allocation of respondents to interviewers and re-interviewing of 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 section the analysis is confined to the estimation of the quantities described in section 4.2. The interviews for the main survey and those for the re-interview survey were analysed separately. The magnitude of the correlated response variance can be estimated 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 (4.17) and is:

$$\frac{\rho_1\sigma_\varepsilon^2(m - 1)}{n} \qquad (4.17)$$

A good index of the potential impact of the interviewer variance is

$$\rho_1 I = \rho_1\sigma_\varepsilon^2/(\sigma_y^2 + \sigma_\varepsilon^2) \qquad (4.19)$$

where I is the index of inconsistency, defined on page 17. A simple estimator of the denominator is s^2, defined in (4.10). A more precise estimator is given by $(1/m)(C + F) + F$ from section 4.2.

Table 14 presents the estimated values of $\varrho_1 I$ for the variables for which significant results were found in Lesotho. The estimation procedure provides three separate estimates of $\rho_1 I$ – one for the whole of the main survey and one for each phase of the Response Errors Project for the respondents who were interviewed twice. The latter two sets of estimates are based on distinct data sets.

The values in table 14 are all estimates and are themselves subject to variance. It is noteworthy, however, that the same four variables emerge as the most sensitive to interviewer effect in all three analyses. The estimates in the second column are based on a subset of the responses considered in the first column, but the third column represents an entirely different set of responses. The result for *Years of education* for the total sample is surprisingly low, but otherwise the results provide reassurance on the representativeness of the designated sample.

The last two variables are included in the table as one significant value was obtained in each case. The weight of the evidence suggests, however, that neither variable is actually subject to interviewer effect but that the value is merely the result of chance variation.

The values of $\varrho_1 I$ provide an index of the susceptibility of the variables to interviewer effect. The magnitude of the variance component can be expressed either as

$$\frac{\rho_1\sigma_\varepsilon^2}{n}(m - 1) \qquad (4.17)$$

or alternatively as

$$\rho_1 I\,\frac{(\sigma_y^2 + \sigma_\varepsilon^2)}{n}(m - 1) \qquad (4.20)$$

which has the advantage that it uses as a base the value of $(\sigma_y^2 + \sigma_\varepsilon^2)/n$, 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 $\rho_I I$ will have a considerable impact on the total variance if the value of m is large. With a value of $\rho_I I = 0.04$, 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_I would not in itself be sufficient to imply a large increase in 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_I will have little impact. This is indicated by the fact that three of the first four variables in table 14 are those with the highest values of $\hat{I}$ in table 8 and 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 responses 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 be the dominant component.

4.5 PARTITIONING THE TOTAL VARIANCE

In this section two variables are considered in detail – *First birth interval* and *Ever-use of contraception* – 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 4.6 – *Age at marriage* and *Children ever born* – neither of which shows any evidence of interviewer effect.

Table 15 presents the results for *First birth interval.* The first column gives the sizes of the base variance components – $\hat{\sigma}_y^2$, $\hat{\sigma}_\varepsilon^2$, roh $\hat{\sigma}_y^2$ and $\rho_I \hat{\sigma}_\varepsilon^2$. The next three columns give the total variance for three special cases: (1) the actual design with average interviewer workload (m) of 100 and average cluster take of 36.8; (2) a design with average interviewer workload of 1 with cluster take unchanged; (3) a simple random sample (ie b = 1) with m = 1.

The simple total variance is the sum of the simple sampling variance and the simple response variance. This quantity can be estimated directly from the survey data by s^2.

As part of the routine analysis of 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:

$$\frac{\sigma_y^2}{n}\{1 + \text{roh}\,(b - 1)\} + \frac{\sigma_\varepsilon^2}{n}. \qquad (4.21)$$

The interviewer variance is:

$$\frac{\rho_I \sigma_\varepsilon^2}{n}(m - 1) \qquad (4.17)$$

The total variance is:

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

$$+ \frac{\sigma_\varepsilon^2}{n}\{\rho_I(m - 1)\}. \qquad (4.22)$$

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 the total variance 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.

Figure 5 presents a diagrammatic representation of the results in table 15. Column I gives the estimate of the actual total variance and its components for the survey design used in Lesotho. The magnitude of the correlated

Table 15 Total variance: first birth interval

	Base components	m = 100 (act) b = 36.8 (act)	m = 1 b = act	m = 1 b = 1
Correlated response variance (CRV)	25.46	2520.5	–	–
Correlated sampling variance (CSV)	6.788	243.0	243.0	–
Simple response variance (SRV)	330.8	330.8	330.8	330.8
Simple sampling variance (SSV)	234.9	234.9	234.9	234.9
Total variance (TV)	–	3329.2	808.7	565.7

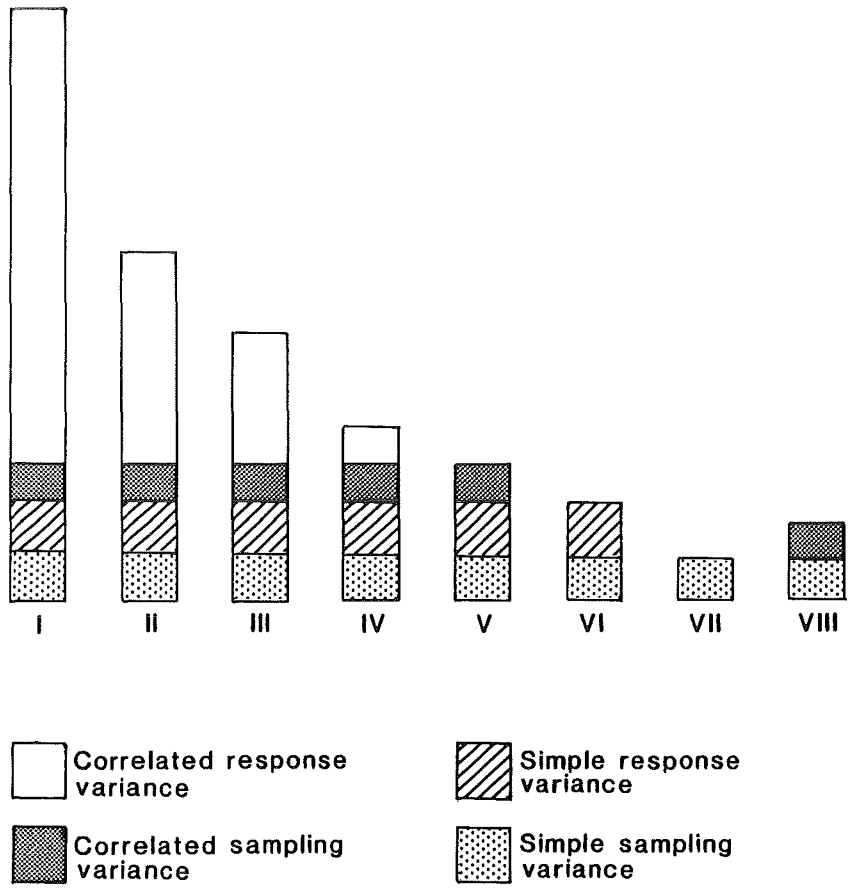

Figure 5 Total variance: first birth interval, Lesotho

sampling and response variances are based on the values of b and m actually used, ie b = 36.8 and m = 100. It is clear from column I that the total variance is dominated by the interviewer variance, which accounts for some 76 per cent of the total variance. The correlated sampling variance accounts for 7 per cent, the simple response variance for 10 per cent and the simple sampling variance for 7 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 underestimated by a factor of more than five.

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 neglected by this estimate is the correlated response variance. In this case the total variance would be underestimated by a factor of four.

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 $\rho_1 \sigma_\varepsilon^2(m - 1)/n$. 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 30 and 10 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.

The impact of the interviewer variance for *First birth interval* for this design is such that the change in field strategy indicated by column II would lead to a reduction of the actual total variance by 38 per cent. The further increase in the number of interviewers represented by column III would reduce the total variance by a further 15 per cent; an increase to ten times the original number of interviewers would give a total variance equal to less than one-third of the actual total variance.

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.

The results for *Ever-use of contraception* are given in table 16 and in figure 6. 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

Table 16 Total variance: ever-use of contraception

	Base components	m = 100 b = 36.8	m = 1 b = act	m = 1 b = 1
CRV	0.0130	1.2848	–	–
CSV	0.0059	0.2102	0.2102	–
SRV	0.0850	0.0850	0.0850	0.0850
SSV	0.0695	0.0695	0.0695	0.0695
TV	–	1.6495	0.3647	0.1545

the total variance. The difference between column VI and column V highlights the necessity for proper estimation of sampling variance. Ignoring the effect of the clustering in the sample design would lead to an underestimation of almost 60 per cent. 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 78 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, and demonstrate how this dominance by the interviewer variance can be radically altered. With an average workload size of m = 10, for instance, the interviewer variance – other things being equal – would account for less than a quarter of the total variance.

4.6 SUMMARY MEASURES AND CONFIDENCE INTERVALS

The results in section 4.5 are not typical of all variables in the Lesotho 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 more convenient to use the simple total variance $(\sigma_y^2 + \sigma_\varepsilon^2)/n$ as a base for both.

Thus the correlated sampling variance, which has previously been written as:

$$\frac{\sigma_y^2}{n}\{\text{roh}\ (b - 1)\} \qquad (4.13)$$

can be written as:

$$\frac{\sigma_y^2 + \sigma_\varepsilon^2}{n}\rho_{cl}(b - 1) \qquad (4.23)$$

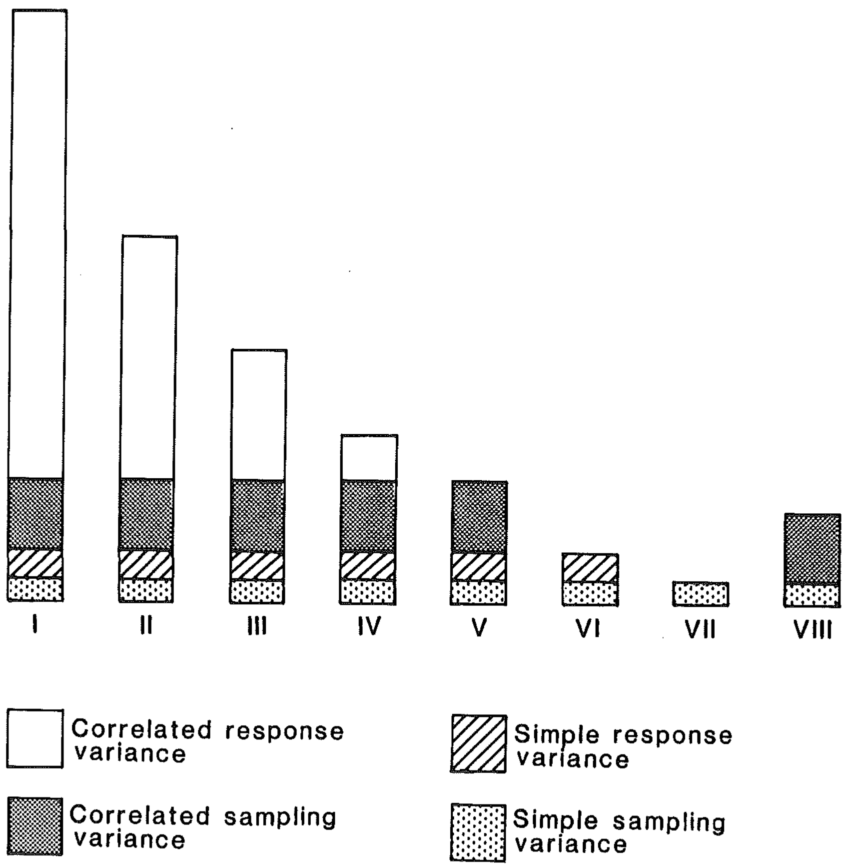

Figure 6 Total variance: ever-use of contraception, Lesotho

Table 17 Summary measures of the variances and standard errors for four variables

Variable	$1/(1-I)$	deff (for b = 36.8)	inteff	$1/(1-\hat{I})$	deft	inteft
Age at marriage	1.44	1.00	1.00	1.20	1.00	1.00
Children ever born	1.09	1.14	1.00	1.04	1.07	1.00
First birth interval	2.41	1.44	4.46	1.55	1.20	2.11
Ever-use of contraception	2.22	2.37	8.32	1.49	1.54	2.88

where ρ_{cl} is a synthetic intra-cluster correlation coefficient which takes into account the presence of the simple response variance. The conventional 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

$$\frac{\sigma_\varepsilon^2}{n}\{\rho_1(m-1)\} \tag{4.17}$$

or as

$$\frac{\sigma_y^2 + \sigma_\varepsilon^2}{n}\rho_{int}(m-1) \tag{4.24}$$

where $\rho_{int} = \rho_1 I$.

The total variance (4.22) can now be written as

$$\frac{\sigma_y^2 + \sigma_\varepsilon^2}{n}\{1 + \rho_{cl}(b-1) + \rho_{int}(m-1)\}. \tag{4.25}$$

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

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

For the results presented here, Deff and Inteff (and consequently Deft and Inteft) are estimated and their estimates will be denoted by deff, inteff. The choice between using variances and standard errors depends on the purpose for which the results are presented. Table 17 provides both for the four variables concerned. The variables are *Age at marriage*, *Children ever born*, *First birth interval* and *Ever-use of contraception*.

In order to make the first and fourth columns of the table comparable to the others, $1/(1 - I)$ is presented instead of I. This quantity 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*. It has a 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 1.24. If the simple total variance is taken as a base, the multiplying factor is 1.14.

Age at marriage is similarly dominated by the simple sampling variance, although the simple response variance in this case accounts for 30 per cent of the simple total variance. There is no increase in the variance due to the clustering of the sample. In other words, the design effect is equal to 1; *Age at marriage* does not differ systematically across clusters. There is also no evidence of any interviewer variance. The overall ratio of the actual total variance to the simple sampling variance is 1.44.

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 11.81 for the *First birth interval* and 21.51 for *Ever-use of contraception*; the ratios of the total variance to the simple total variance are 4.90 and 9.69 respectively.

These ratios are easily calculated from the results given in table 17. The advantage of working with the variances directly is that the variances are additive. The total variance is:

$$\frac{\sigma_y^2}{n}\cdot\frac{1}{1-I}\cdot\{1 + (\text{deff} - 1) + (\text{inteff} - 1)\}. \tag{4.26}$$

If $d^2 = \text{deff} - 1$ and $i^2 = \text{inteff} - 1$, then the total variance is:

$$\frac{\sigma_y^2 + \sigma_\varepsilon^2}{n}(1 + d^2 + i^2). \tag{4.27}$$

It is more relevant in the context of interval estimation to work with the standard errors. The effect on the standard error of each of the components of variance is given in the second half of table 17.

To illustrate the impact of these effects on the four variables, table 18 gives the width of the 95 per cent confidence intervals using the three possible estimation procedures. The sample mean is also given for each

Table 18 Width of 95 per cent confidence interval for four variables using different estimates of the total error (based on n = 3603 in all cases)

Variable	Mean	Simple sampling error (1)	Simple total error (2)	(2) × deft (3)	Correct estimated standard error (4)
Age at marriage	17.90	0.173	0.208	0.208	0.208
Children ever born	3.19	0.162	0.169	0.180	0.180
First birth interval	25.96	1.00	1.55	1.86	3.43
Ever-use of contraception	0.23	0.0172	0.0256	0.0394	0.0797

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 *Age at marriage* illustrates the position when both the correlated components of variance are zero. The last three columns in table 18 are identical for this variable. For *Children ever born* there is a non-negligible design factor, which can be seen from the differences between columns (3) and (4). It should be noted, however, that for some variables not given in table 18 the design factor 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 using s^2/n to estimate the variance (column 2) would be 1.55; using the standard (correct) estimate of sampling variance (column 3) the width would be 1.86. When the interviewer effect is taken into account the interval is seen to be 3.43, a *further* increase of 84 per cent. For *Ever-use of contraception* the disparity is even more striking. Column (2) gives a confidence interval of width 0.0256. Once the design effect is introduced, this increases to 0.0394 (column 3), a rise of 54 per cent. The interviewer effect increases the confidence interval to 0.0797, a *further* rise of over 100 per cent.

The importance of both components of correlated variance can also be illustrated by considering the true confidence level for the estimates constructed using columns (2) and (3). Table 19 gives the results.

The first two variables in tables 17, 18 and 19 are more representative of variables from WFS surveys than are the last two. Furthermore, the results in the tables 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 a later section of this report.

4.7 EFFECTS ON CROSS-CLASSES

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

Table 19 Apparent and true confidence levels for confidence intervals constructed using columns (2) and (3) of table 18

Variable	Estimate	95 per cent confidence level		99 per cent confidence level	
		Apparent confidence level (%)	True confidence level (%)	Apparent confidence level (%)	True confidence level (%)
Age at marriage	Col. (2)	95	95	99	99
	Col. (3)	95	95	99	99
Children ever born	Col. (2)	95	93	99	98
	Col. (3)	95	95	99	99
First birth interval	Col. (2)	95	63	99	76
	Col. (3)	95	71	99	84
Ever-use of contraception	Col. (2)	95	47	99	59
	Col. (3)	95	67	99	80

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: (1) subclasses defined in terms of *demographic* characteristics (age, marriage duration, etc); (2) subclasses defined in terms of *socio-economic* characteristics (woman's literacy, husband's level of education, occupation, etc); and (3) 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 them 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 *cross-classes.* 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 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 cross-classes the intra-cluster correlation coefficient is approximately stable, although it may increase slightly as the relative size of the cross-class decreases.

In this section the analysis is extended to the more complex case of the total variance. The purpose is to investigate the implications of a simple approximate model for the total variance of a subclass. The algebraic model is illustrated by applying it to the total variance found in Lesotho for the four variables discussed in section 4.6. For one of these variables, detailed analysis provides some support for the assumptions of the model.

A model

The model used here is presented in O'Muircheartaigh (1984). In essence it demonstrates that the relationship between the total variance for the whole sample and the total variance for a cross-class can be reduced to a very simple form. If V_s is the variance of the sample mean for a cross-class representing a proportion M_s of the whole sample, and V_t is the corresponding variance for the whole sample, then

$$\frac{V_s}{V_t} = 1 + \frac{1 - M_s}{M_s} \cdot \frac{k_1}{k_1 + k_2} \qquad (4.28)$$

where $k_1 = 1 - \rho_{cl} - \rho_{int}$ and $k_2 = \rho_{cl} b_t + \rho_{int} m_t$, and b_t and m_t are the average cluster take and interviewer workload for the whole sample.

Table 20 Values of ρ_{cl}, ρ_{int} and $k_1/(k_1 + k_2)$

Variable	ρ_{cl}	ρ_{int}	$k_1/(k_1 + k_2)$
Age at marriage	0.000	0.000	1.0000
Children ever born	0.004	0.000	0.8712
First birth interval	0.012	0.045	0.1908
Ever-use of contraception	0.038	0.084	0.0916

An application

The important factors in determining the variance of a cross-class are ρ_{cl}, ρ_{int} and $k_1/(k_1 + k_2)$. Table 20 gives the values of these quantities for the four variables previously considered in section 4.6.

The first variable is an example of the simplest case in which there is no correlated variance. Both ρ_{cl} and ρ_{int} are zero and thus $k_1/(k_1 + k_2)$ is equal to 1. The second variable has just one component of correlated variance – the correlated sampling variance. The value of ρ_{cl} is, however, small, and $k_1/(k_1 + k_2)$ is close to 1.

The third variable is an intermediate case in which both components of correlated variance are present and non-negligible. The value of $k_1/(k_1 + k_2)$ is relatively small.

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 very low value of $k_1/(k_1 + k_2)$ – less than 0.1. The absolute minimum value for this factor is zero.

The implications of the parameters in table 20 can be seen from table 21, 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

Table 21 Relative magnitude of V_s and V_t (values of V_s/V_t)

Variable	Cross-class size (M_s)		
	0.5	0.3	0.1
Age at marriage	2.00	3.33	10.00
Children ever born	1.87	3.03	8.84
First birth interval	1.19	1.45	2.72
Ever-use of contraception	1.09	1.21	1.82

Table 22 Values of *deff*, *inteff* and *toteff* for different values of M_s

Variable	Measure	Total sample	$M_s = 0.5$	$M_s = 0.3$	$M_s = 0.1$
Age at marriage	deff	1.00	1.00	1.00	1.00
	inteff	1.00	1.00	1.00	1.00
	toteff	1.00	1.00	1.00	1.00
Children ever born	deff	1.14	1.07	1.04	1.01
	inteff	1.00	1.00	1.00	1.00
	toteff	1.14	1.07	1.04	1.01
First birth interval	deff	1.44	1.21	1.12	1.04
	inteff	4.46	3.21	2.31	1.41
	toteff	4.90	3.42	2.43	1.45
Ever-use of contraception	deff	2.37	1.66	1.38	1.14
	inteff	8.32	5.12	3.44	1.76
	toteff	9.69	5.78	3.82	1.90

third comprising one-tenth of the sample. Many subclasses used in practice fall in this range, although for multiway classifications even smaller subclasses will frequently be involved.

In evaluating the figures in table 21 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. *Age at marriage* provides an example of such a variable, as can be seen from the first row of the table.

As might be expected, the variable *Children ever born* has values of V_s/V_t close to the upper limit. This is because there is no interviewer variance for this variable and the correlated sampling variance is relatively small. The results for the variable *First birth interval* show how misleading it would be to apply this upper limit to a case where either of the correlated variance components is large. Under the assumptions of the model, using the upper limit for the variance would lead to overestimating the total variance by 66 per cent when $M_s = 0.5$; by more than 100 per cent when $M_s = 0.3$; and by almost 300 per cent when $M_s = 0.1$.

The last variable in the table displays even more dramatic results. This variable is atypical since both the correlated sampling variance and the interviewer variance are extremely large. However, in such a situation the effects are astonishing. For a cross-class 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 to only relatively small increases in the variance. For $M_s = 0.1$ (a cross-class comprising one-tenth of the sample) the ratio of V_s/V_t is only 1.82. For a variable with no correlated variance this ratio would be 10.00.

The results in tables 20 and 21 can also be presented in a form closer to the approach used in discussing sampling variance. Table 22 gives the values of *deff*, *inteff* and *toteff*, where

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

and toteff is the ratio of the *total variance* (4.22) to the *simple total variance* (4.9).

The results in table 22 conform to the pattern observed in the sampling literature for cross-classes. Under the assumptions of the model the effect of the correlated variance components decreases as the proportion of the population in the cross-class decreases. The larger the effect of the correlated variance components, the more dramatic the reduction as M_s decreases. The rates at which *deff* and *inteff* decrease differ since the multipliers $(b - 1)$ and $(m - 1)$ differ.

Finally, to illustrate the practical implications of these results for the evaluation of survey estimates, table 23 gives the width of the 95 per cent confidence intervals for cross-classes 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 cross-classes, the model described on page 34 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 cross-classes both these sizes are greatly reduced, with a consequent reduction in the correlated components.

The final column of table 23 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 cross-classes is a good deal narrower than would be expected if sample size were

Table 23 Width of 95 per cent confidence intervals for cross-classes of different sizes

Variable	Mean	Cross-class	Simple sampling error (1)	Total simple error (2)	(2) × deft (3)	Correct standard error (4)
Age at marriage	17.90	Total sample	0.173	0.208	0.208	0.208
		$M_s = 0.5$	0.245	0.294	0.294	0.294
		$M_s = 0.3$	0.316	0.380	0.380	0.380
		$M_s = 0.1$	0.547	0.658	0.658	0.658
Children ever born	3.19	Total sample	0.162	0.169	0.180	0.180
		$M_s = 0.5$	0.229	0.239	0.246	0.246
		$M_s = 0.3$	0.296	0.309	0.314	0.314
		$M_s = 0.1$	0.512	0.534	0.535	0.535
First birth interval	25.96	Total sample	1.00	1.55	1.86	3.43
		$M_s = 0.5$	1.414	2.192	2.411	3.74
		$M_s = 0.3$	1.826	2.830	3.000	4.13
		$M_s = 0.1$	3.162	4.901	4.999	5.66
Ever-use of contraception	0.23	Total sample	0.0172	0.0256	0.0394	0.0797
		$M_s = 0.5$	0.0243	0.0362	0.0467	0.0832
		$M_s = 0.3$	0.0314	0.0467	0.0546	0.0877
		$M_s = 0.1$	0.0544	0.0809	0.0866	0.1075

the only consideration. For *Ever-use of contraception* the confidence interval for the cross-class with $M_s = 0.1$ (ie based on one-tenth of the total sample) is only 35 per cent wider than the confidence interval based on the total sample. This is because the dominance of the correlated error in the standard error of estimates based on the total sample becomes progressively weaker as the cross-class size decreases.

Discussion

A model is presented (page 35) 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 cross-classes 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 cross-class. A number of important assumptions are made in the model. First, it is assumed that the cross-classes 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 cross-classes. 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 (pages 34–35) 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. In this case the relative size of cross-class variance is determined largely by the cross-class 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. However, for cross-classes this dominance is reduced as the cross-class size decreases. For small cross-classes 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 9.69 for estimates based on the total sample and only 1.90 for estimates based on a cross-class 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 cross-class is reduced.

Although the results (pages 34–36) are not based on direct computations of the variance, values of the parameters on which the calculations are based are obtained from computations carried out on data from the Lesotho study. There is a problem in estimating the correlated variance components for the cross-classes in that as the sample size decreases the estimates themselves become subject to larger variances.

The variable with the largest correlated interviewer variance for Lesotho was *Ever-use of contraception*. Table 24 gives the estimated values of ρ_{int} for five subclasses for this variable, together with their estimated

Table 24 Estimated values of ρ_{int} for cross-classes for *Ever-use of contraception*

	Total sample	Age < 25	Age 25–34	Age 35–44	Educ. 1–5 yr	Educ. 6 yr +
ρ_{int}	0.084	0.061	0.121	0.072	0.088	0.052
se(ρ_{int})	0.018	0.042	0.038	0.044	0.034	0.024
M_s	1.00	0.31	0.33	0.27	0.44	0.48

standard errors. The computations are for the data from the main survey.

From the table it can be seen that the values of ρ_{int} for the subclasses are consistent with the assumption that ρ_{int} remains constant across subclasses. In no case is the value of ρ_{int} more than one standard from the value of 0.084 obtained for the total sample. This evidence provides some support for the model.

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.

5 Further Analysis

5.1 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 putting 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 magnitude 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 609 cases in Lesotho are given in table 25. A clear pattern emerges from the table. The magnitudes of the response deviations are directly related to the interviewer's assessments. There are only three variables for which the differences are not statistically significant. In two of these cases the direction of the differences is in keeping with the general pattern (*Last closed birth interval* and *No of children desired*); the third case (*Ever-use of contracep-*

Table 25 Magnitude of response deviations cross-tabulated by interviewers' assessments

Variable	Reliability			Co-operation			Total
	GOOD	FAIR	POOR	GOOD/ VERY GOOD	FAIR	POOR	
1 Year of last birth	0.47	0.73	2.17	0.47	0.84	0.95	0.54
2 Month of last birth	5.97	9.43	26.83	5.97	11.02	10.79	6.85
3 Year of next to last birth	0.67	0.92	3.00	0.68	0.94	1.98	0.76
4 Month of next to last birth	8.49	11.64	39.33	8.57	12.19	26.64	9.59
5 Year of first birth	0.73	1.62	5.00	0.77	1.48	4.58	0.94
6 Month of first birth	9.09	19.89	62.67	9.49	18.68	57.08	11.72
7 Age	1.04	2.05	1.50	1.07	1.65	4.78	1.24
8 Age in five-year groups	0.20	0.38	0.12	0.21	0.28	0.80	0.23
9 Year of marriage	0.98	2.05	1.87	1.00	1.70	4.81	1.19
10 Marital duration	1.10	2.17	2.00	1.11	1.83	5.00	1.30
11 Children ever born	0.32	0.46	1.00	0.30	0.53	1.32	0.36
12 Years of education	0.49	0.80	1.25	0.50	0.82	1.19	0.56
13 Births in past five years	0.16	0.34	0.37	0.16	0.29	0.68	0.19
14 Age at first marriage	1.16	1.63	2.00	1.18	1.44	3.03	1.26
15 Last closed birth interval	7.50	8.27	15.42	7.80	7.15	10.29	7.74
16 No of children desired	1.38	1.77	1.57	1.42	1.65	1.63	1.46
17 Ever-use of contraception	0.20	0.13	0.25	0.19	0.20	0.00	0.19
18 First birth interval	7.31	14.06	66.96	7.87	10.29	64.20	9.33
Sample size	486	115	8	499	97	12	609

tion) is a special one, being a binary variable. The remaining fifteen variables all show statistically significant differences. The *Reliability* classification is slightly more successful in differentiating between the respondents on the birth history variables, as might be expected.

When the linearity component of the differences is tested separately (with one degree of freedom) the strength of the relationship is confirmed. For twelve of the variables in the case of the *Reliability* classification and ten in the case of the *Co-operation* classification the linearity component is significant at the 0.01 level.

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 2 per cent in each case – but the AVERAGE/FAIR category is also effective in identifying a group with high response variability.

The same analysis was carried out for the age subclasses and for the education subclasses described previously. Since the sample sizes are considerably smaller for the subclasses, the POOR group was amalgamated with the FAIR/AVERAGE group for the analysis. The pattern of results persisted for the subclasses, and the differences were statistically significant for the fertility variables despite the smaller sample sizes.

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; almost 80 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 not particularly successful in differentiating between the two. This suggests that although there may be scope for extending the categorization used in the assessment of reliability, the naming of the categories requires further consideration.

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.

5.2 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 for reducing bias in ratio estimators and now widely used to estimate variances (see, for example, Kish and Frankel 1974, Kalton 1977). The procedure used here is described in O'Muircheartaigh (1984). It can be applied to measures based on the whole sample and also to measures based on subclasses.

The simple response variance

One of the basic measures of response error used in this report is the simple response variance. Table 26 presents

Table 26 $\hat{\sigma}_\varepsilon^2$, se($\hat{\sigma}_\varepsilon^2$) and cv($\hat{\sigma}_\varepsilon^2$) for the 18 variables

Variable	$\hat{\sigma}_\varepsilon^2$	se($\hat{\sigma}_\varepsilon^2$)	cv($\hat{\sigma}_\varepsilon^2$)
Age	5.168	1.300	0.25
Children ever born	0.5580	0.1230	0.22
Year of first birth	4.456	1.188	0.27
Month of first birth	631.7	163.2	0.26
Age in five-year groups	0.2325	0.0520	0.22
Year of last birth	1.118	0.1793	0.16
Month of last birth	171.3	26.58	0.16
Year of marriage	5.973	1.296	0.22
Marital duration	5.964	1.312	0.22
Education in years	0.6460	0.0787	0.12
Year of next to last birth	1.649	0.2866	0.17
Month of next to last birth	244.0	44.69	0.18
Births in past five years	0.1215	0.0121	0.10
Last closed birth interval	185.8	45.43	0.24
Age at marriage	3.102	0.5364	0.17
Ever-use of contraception	0.0850	0.0081	0.10
First birth interval	330.8	80.99	0.24
No of children desired	3.248	0.6683	0.21

Table 27 $\hat{\sigma}_\varepsilon^2$, se($\hat{\sigma}_\varepsilon^2$) and cv($\hat{\sigma}_\varepsilon^2$) for six variables for four

Variable	Children ever born			Year of last birth			Marital duration		
Subclass	$\hat{\sigma}_\varepsilon^2$	se($\hat{\sigma}_\varepsilon^2$)	cv($\hat{\sigma}_\varepsilon^2$)	$\hat{\sigma}_\varepsilon^2$	se($\hat{\sigma}_\varepsilon^2$)	cv($\hat{\sigma}_\varepsilon^2$)	$\hat{\sigma}_\varepsilon^2$	se($\hat{\sigma}_\varepsilon^2$)	cv($\hat{\sigma}_\varepsilon^2$)
Under 25 yr	0.3160	0.1255	0.40	0.351	0.1483	0.42	3.927	1.226	0.31
Over 45 yr	0.5950	0.3834	0.64	1.695	0.8796	0.52	7.021	3.940	0.56
Educ. 0–4 yr	0.8182	0.2506	0.31	2.058	0.4439	0.22	10.400	2.556	0.25
Educ. 7+ yr	0.5760	0.1769	0.31	1.123	0.4143	0.37	4.001	1.413	0.35
All	0.5580	0.1230	0.22	1.118	0.1793	0.16	5.964	1.312	0.22

Variable	Age at marriage			Ever-use of contraception			First birth interval		
Subclass	$\hat{\sigma}_\varepsilon^2$	se($\hat{\sigma}_\varepsilon^2$)	cv($\hat{\sigma}_\varepsilon^2$)	$\hat{\sigma}_\varepsilon^2$	se($\hat{\sigma}_\varepsilon^2$)	cv($\hat{\sigma}_\varepsilon^2$)	$\hat{\sigma}_\varepsilon^2$	se($\hat{\sigma}_\varepsilon^2$)	cv($\hat{\sigma}_\varepsilon^2$)
Under 25 yr	1.416	0.3550	0.25	0.0575	0.0111	0.19	74.27	34.41	0.46
Over 45 yr	2.909	1.0728	0.37	0.0655	0.0296	0.45	592.80	310.66	0.52
Educ. 0–4 yr	4.156	1.2680	0.31	0.0677	0.0131	0.19	628.20	218.14	0.35
Educ. 7+ yr	1.760	0.4060	0.23	0.1075	0.0165	0.15	57.72	28.20	0.49
All	3.102	0.5364	0.17	0.0850	0.0081	0.10	330.80	80.99	0.24

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 18 variables previously considered.

The results in table 26 are reassuring. The coefficient of variation of $\hat{\sigma}_\varepsilon^2$ is remarkably stable across variables, with most values close to 0.20. The level of the values is satisfactory in that it provides reasonable confidence in the estimated values of σ_ε^2. The range of values of the cv($\hat{\sigma}_\varepsilon^2$) is from 0.10 to 0.27, with the lowest values for *Births in the past five years*, *Ever-use of contraception* and *Education*.

The estimates of σ_ε^2 in table 26 are based on the whole sample of n = 609 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 27 gives the results of the jackknife estimation of the variance for four important subclasses: respondents under 25; respondents over 45; respondents with 0–4 years of education; and those with more than seven years of education. The six variables presented are chosen to represent different levels of sensitivity to response errors.

An interesting feature of table 27 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 less educated and over 45 subclasses than for the under 25 and more educated subclasses. This is in keeping with the results previously discussed in section 4.1, and serves 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 27 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 conforms to the theoretical expectation for the variance of a variance estimator of this kind.

On theoretical statistical grounds we would expect the ratio of coefficient of variation of $\hat{\sigma}_\varepsilon^2$ for the total sample to that for a subclass to be approximately inversely proportional to $(n_t/n_s)^{1/2}$, where n_t and n_s are the sample sizes for the total sample and the subclass respectively. For the subclass in table 27 this would imply ratios of 1.9:3.1:1.8:1.9:1 for the coefficients of variation for the under 25, over 45, less educated, more educated and total sample respectively. These are remarkably close to the ratios found in table 27. Equally satisfying is the fact that even for the subclasses, the coefficients of variation are reasonably small. Except for the smallest subclass (respondents over 45), the coefficients of variation are of the order of 0.2–0.3; for the smallest subclass they are of the order of 0.5.

The index of inconsistency, I

The index of inconsistency, I (defined by $\sigma_\varepsilon^2/(\sigma_y^2 + \sigma_\varepsilon^2)$, 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 4.1 to describe the sensitivity of variables to response errors. In figure 1 and table 8 the values of $\hat{I}$ for the total sample are presented, while tables 9–13 and figures 2 and 3 give the value of $\hat{I}$ for major subclasses. The validity of the conclusions drawn from these tables and figures depends on the precision of the estimates of I.

Table 28 presents the results of the jackknife estimation of the variance of $\hat{I}$ for the six variables and four subclasses previously considered. The variables span the range of observed values of $\hat{I}$ and the subclasses represent the extremes of the characteristics considered.

Table 28 $\hat{I}$, se($\hat{I}$) and cv($\hat{I}$) for six variables and four 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($\hat{I}$)
Under 25 yr	0.237	0.0844	0.36	0.282	0.0701	0.25	0.400	0.0820	0.20
Over 45 yr	0.075	0.0422	0.56	0.034	0.0247	0.73	0.423	0.0851	0.20
Educ. 0–4 yr	0.116	0.0319	0.27	0.043	0.0144	0.33	0.117	0.0336	0.29
Educ. 7+ yr	0.094	0.0360	0.38	0.058	0.0283	0.66	0.058	0.0198	0.34
All	0.083	0.0179	0.22	0.037	0.0072	0.19	0.074	0.0177	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($\hat{I}$)
Under 25 yr	0.329	0.0935	0.28	0.535	0.0974	0.18	0.274	0.1065	0.39
Over 45 yr	0.286	0.1737	0.61	0.766	0.2205	0.29	0.780	0.1726	0.22
Educ. 0–4 yr	0.424	0.0755	0.18	0.564	0.0812	0.14	0.856	0.1178	0.14
Educ. 7+ yr	0.185	0.0433	0.23	0.550	0.0719	0.13	0.191	0.0884	0.46
All	0.306	0.0373	0.12	0.550	0.0521	0.09	0.585	0.0919	0.16

The pattern of variation in the values of $\hat{I}$ is similar to that for $\hat{\sigma}_\varepsilon^2$. The values of $\hat{I}$ for the variables presented are: for *Children ever born* I is 0.08; for *Year of last birth*, 0.04; for *Marital duration*, 0.07; for *Age at marriage*, 0.31; for *Ever-use of contraception*, 0.55; and for *first birth interval*, 0.58.

The coefficients of variation for the estimates of I for the total sample are similar to those for the corresponding values for $\hat{\sigma}_\varepsilon^2$. For the subclasses the pattern is also similar, with the smallest subclass (respondents over 45) having the largest coefficient of variation for $\hat{I}$ in the case of four of the six variables. Only five of the 24 coefficients presented exceed 0.4; the average value for the others is about 0.23. These are comparable to the corresponding values for $\hat{\sigma}_\varepsilon^2$, 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 4.1.

In general, the variance of the difference between two random variables x_1 and x_2 is

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

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

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

Table 29A 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 29A 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 $\hat{I}$ between women with little education and those with more than seven years' education for the first birth interval, the confidence interval would be 0.665 $\pm$ 0.294 or 0.371, 0.959.

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 29B.

The difference in the values of $\hat{I}$ is 0.23. 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 greater than the estimated difference, and thus a 95 per cent normal confidence interval would be:

$$0.231 \pm 0.472 \text{ or } (-0.241,\ 0.703).$$

This does not mean that there is no difference between the values of the index of inconsistency for the two subclasses for *Ever-use of contraception*. It does mean, however, that additional evidence would be necessary

Table 29A Standard errors of contrasts of $\hat{I}$ for subclass pairs

Variable	Subclass	$\hat{I}$	se($\hat{I}$)	$\hat{I}_1 - \hat{I}_2$	se($\hat{I}_1 - \hat{I}_2$)
First birth interval	0–4 yr education	0.856	0.1178	0.665	0.147
	7+ yr education	0.191	0.0884		
First birth interval	Age under 25	0.274	0.1065	0.506	0.203
	Age over 45	0.780	0.1726		
Age at marriage	0–4 yr education	0.424	0.0755	0.239	0.087
	7+ yr education	0.185	0.0433		

Table 29B A counter-example to table 29A

Variable	Subclass	$\hat{I}$	se($\hat{I}$)	$\hat{I}_1 - \hat{I}_2$	se($\hat{I}_1 - \hat{I}_2$)
Ever-use of contraception	Age under 25	0.535	0.0974	0.231	0.241
	Age over 45	0.766	0.2205		

before the presence of the difference could be established beyond reasonable doubt.

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 29B provides an example.

Variance estimates for the correlated response variance were also derived using the jackknife. Some examples can be found in section 4.7.

References

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.

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.

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.

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

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.