OCCASIONAL PAPERS 1

DHS Demographic and Health Surveys

An Evaluation of the Pakistan DHS Survey Based on the Reinterview Survey

MACRO INTERNATIONAL INC.

The Demographic and Health Surveys (DHS) is a 13-year project to assist government and private agencies in developing countries to conduct national sample surveys on population and maternal and child health. Funded primarily by the United States Agency for International Development (USAID), DHS is administered by Macro International Inc. in Calverton, Maryland.

The main objectives of the DHS program are to: (1) promote widespread dissemination and utilization of DHS data among policymakers, (2) expand the international population and health database, (3) advance survey methodology, and (4) develop in participating countries the skills and resources necessary to conduct high-quality demographic and health surveys.

For information about the Demographic and Health Surveys program, write to DHS, Macro International Inc., 11785 Beltsville Drive, Calverton, MD 20705-3119, U.S.A. (Telephone 301-572-0200; Telefax 301-572-0999; Telex 198116).

An Evaluation of the Pakistan DHS Survey Based on the Reinterview Survey

Siân L. Curtis

Fred Arnold

Macro International Inc.

Calverton, Maryland

May 1994

The recommended citation for this publication is:

Curtis, Siân L. and Fred Arnold. 1994. *An Evaluation of the Pakistan DHS Survey Based on the Reinterview Survey*. Occasional Papers No. 1. Calverton, Maryland: Macro International Inc.

CONTENTS

TABLES

FIGURES

Preface

Some form of reinterview study was conducted in a number of countries during the second phase of the Demographic and Health Surveys program. In most cases the reinterview surveys served to provide a check on the information collected by the interviewers with a view to improving the quality of subsequent data. Such reinterview surveys were conducted at the same time as the main survey so that any problems that were identified could be addressed immediately. In Pakistan the reinterview survey was expanded into a more detailed study of data reliability. A subsample of women interviewed in the main survey was reinterviewed some time after the completion of the main fieldwork, using a shortened form of the original questionnaire. The purpose of this reinterview survey was to enable a direct comparison to be made between the responses given in the main survey and those given in the reinterview survey for these women, and hence to enable an assessment of the consistency of responses at the national level. This report presents an analysis of the Pakistan reinterview survey.

Chapter 1

Introduction

The aim of this report is to evaluate the reliability of data collected in the Pakistan Demographic and Health Survey (DHS) through an analysis of the reinterview survey. The consistency of reporting of several key variables collected in the Pakistan DHS is examined, in order to identify particular problems that should be addressed in future surveys, as well as to gain an overall insight into the strengths and limitations of the data collected.

Pakistan is a particularly challenging setting in which to attempt to collect high quality demographic data. The level of literacy is very low, particularly for females, and all demographic surveys conducted in Pakistan have experienced problems obtaining reliable data. This is reflected in pronounced heaping in age distributions and other duration variables, such as the duration of breastfeeding, and high levels of missing information on dates of events. These problems are due to a combination of social and cultural factors. Shah (1993) documents several such factors and their likely implications for demographic data collection in Pakistan. For example, three different calendars are used and this may cause confusion in the reporting of events. Birthdays are not widely celebrated so few people actually know their date of birth. In addition, it is common to report one's age at the next birthday rather than in completed years. There is strong digit preference for the number five which is considered blessed by many Muslims, and a tendency toward age exaggeration at older ages due to the respect accorded to old people. Omission of very young children is also likely because traditionally women are expected to stay in bed for 40 days following a birth. During this period visitors are not supposed to enter the room the mother and infant are in. Consequently some recent postpartum women may be omitted from the survey. Estimates of contraceptive use are also affected by "shy users," which is an established phenomenon in Pakistan. In view of this social context it is unrealistic to expect to obtain highly reliable data on most demographic variables, particularly ages and dates, even under the highest quality survey implementation.

Given the problems described above, the consistency of reporting between the Pakistan DHS and the reinterview survey is expected to be low. In many ways it is more important to study reliability of responses in a setting such as Pakistan where reporting errors are a fact of life because it is in precisely such settings that analysts need to know most about the nature of such errors; in particular, whether the errors are systematic or random. While the reinterview survey is subject to many of the same sources of error as the main survey and consequently is no more likely to be correct, comparison of the two surveys can provide unique and invaluable insights into data collection problems in a particularly difficult setting. However, it is essential to interpret the findings of the report in the social and cultural context of Pakistan.

The report begins with a description of the design and implementation of the reinterview survey and an explanation of the measures of consistency used. This is followed by an analysis of the consistency of reporting of background characteristics of women and then of reporting of ages and dates of demographic events. The latter information is essential for many demographic analyses, yet the reporting of ages and dates is known to be poor in many developing countries, including Pakistan. Next, the reporting of births is examined in detail, with particular emphasis on the implications of inconsistencies for the estimation of fertility and infant and child mortality rates. The report then turns to the consistency of reporting of knowledge and use of methods of contraception. Finally, the information collected on fertility desires and preferences is evaluated. These questions are somewhat different from those studied in the earlier sections since they relate to subjective desires rather than to factual information. Such questions are often criticized

for being unreliable due to their hypothetical nature and because many respondents in developing countries may have difficulty understanding them. The last section presents a summary of the findings together with a discussion of their implications.

Chapter 2

Design and Implementation of the Reinterview Survey

The main Pakistan DHS survey was conducted between December 1990 and May 1991. Completed interviews were obtained from 6611 ever-married women age 15-49, using the DHS-II B-core questionnaire. Forty-four of the original clusters were selected for the reinterview survey, which represents about 10 percent of the original clusters selected, and interviews were actually carried out in all but one of these clusters. All women in these clusters who had been interviewed in the main survey were eligible for reinterview. Of 709 women identified for reinterview, completed interviews were obtained from 528. Interviewers were given the name and address of the respondent, the name of her husband, and her relationship to the head of the household, to aid in correct identification of the women. All but 45 of the 528 women who were successfully reinterviewed reported that they remembered being interviewed before, suggesting that identification was reasonably successful.

The fieldwork for the reinterview was carried out between September and November 1991, and the interval between the two surveys for individual women included in this analysis ranged from 5 months to 11 months. One advantage of the relatively long time interval between the two surveys is that women are less likely to recall their responses in the main survey, and consequently the results of the reinterview survey are less likely to be contaminated by such recall effects. However, one disadvantage is that genuine changes in the woman's situation may have occurred between the two surveys, and these need to be distinguished from response errors.

In general, the best interviewers from the main survey were chosen to implement the reinterview survey and no interviewer was permitted to reinterview a woman whom she had already interviewed in the main survey. If such a situation arose, an alternative interviewer was substituted. A much shorter version of the original questionnaire was used for the reinterview. It focused on fertility and contraceptive use, with a reduced number of questions on marriage and background characteristics of the woman and her husband. In particular, the reinterview survey did not include any questions from the health section of the main questionnaire, nor were anthropometric data collected. The responses from the two surveys are matched on the basis of the woman's identification number, which is constructed from the region, area of residence, district number, cluster number, household number, and line number of the woman. A total of 505 women with complete interviews in both surveys were matched by ID number in the two surveys. Of the remaining 23 women with complete reinterviews, 15 did not have a complete interview in the main survey, and no ID matches were found in the main survey for the remaining 8. This latter problem is probably due to errors in entering the ID information in either the main survey or the reinterview survey, or errors in transcribing it onto the reinterview questionnaires.

For practical reasons, a highly clustered sample design was used for the selection of the sample for the reinterview survey. Consequently, the sample weights for the main survey had to be adjusted for the reinterview survey. The adjustment is based on the distribution of the reinterview clusters by region and urban/rural residence. These adjusted reinterview sample weights are normalized so that the weighted number of matched women with complete interviews in both surveys equals the unweighted number (505). In the rest of this report, all numbers reported for the reinterview sample are weighted using these adjusted sample weights. The weighted distributions of these matched women by selected characteristics taken from the main survey, together with the corresponding distributions for the full Pakistan DHS sample, are seen in the first two columns of Table 2.1. The distribution of reinterviewed women by age, marital status, and

education is very similar to that for the full Pakistan DHS sample. There is a slightly higher representation of older women and highly educated women in the reinterview sample than in the full sample, but the differences are very small. However, some differences are apparent for the distributions by province and urban/rural residence. For example, when compared to the main survey the reinterview sample contains some under-representation of women from Balochistan, and some over-representation of urban women. This is because very few clusters from Balochistan and urban North West Frontier Province (NWFP) were included in the reinterview survey. Indeed, no urban clusters from Balochistan were included in the reinterview survey. Nevertheless, the differences between the reinterview sample and the full Pakistan DHS sample appear to be relatively small for these variables, and no substantial sample bias is apparent in the reinterview sample.

Table 2.1 Percent distribution of full PDHS sample and matched reinterview sample by background characteristics from the main survey, Pakistan DHS, 1990-91

Background characteristic	Full PDHS sample	Full matched sample	Matched sample excluding mismatches
Age			
15-19	6.5	4.4	4.6
20-24	16.0	14.5	14.8
25-29	22.6	21.2	22.3
30-34	18.0	17.4	17.3
35-39	14.8	18.2	17.4
40-44	12.8	14.0	14.2
45-49	9.3	10.3	9.4
Marital status			
Married	96.3	96.9	96.8
Widowed	2.4	1.6	1.8
Divorced	0.3	0.3	0.4
Separated	1.0	1.1	1.1
Educational level attended			
No education	79.2	75.7	75.1
Primary	9.1	11.7	11.9
Middle	4.4	3.8	4.1
Secondary	6.2	7.2	7.2
Higher	1.1	1.6	1.7
Province			
Punjab	59.7	59.6	59.5
Sindh	23.1	26.7	27.2
NWFP	13.3	12.4	12.1
Balochistan	3.9	1.4	1.2
Residence			
Urban	30.5	38.9	39.8
Rural	69.5	61.1	60.2
Total	100.0	100.0	100.0
Number	6611	505	474

Initial examination of the data for the 505 matched women revealed that some of the women were matched incorrectly. For example, one matched woman appeared to be age 24 with no children in the main survey, but age 45 with 11 children in the reinterview survey. These incorrect matches are caused by either incorrect identification of women by interviewers at the time of the reinterview or data entry errors that assign women the wrong ID number. Such incorrect matches will bias the analysis by making inconsistencies appear to be much worse than they actually are. However, identification of incorrect matches is problematic since identification is based on the degree of discrepancy between the responses in the two surveys, which is the very issue that this analysis aims to investigate.

If too broad a set of criteria for identifying incorrect matches is adopted, several incorrect matches will be classified as genuine matches, which will upwardly bias the estimates of inconsistency. On the other hand, the results could be biased in the opposite direction if too narrow a set of criteria is used. Whatever criteria are used, there is likely to be some misclassification of incorrect matches so some degree of compromise is necessary in defining the criteria. After careful examination of the data the following criteria were used to define an incorrect match:

- Discrepant by four or more on children ever born, or
- Discrepant by more than 10 years on woman's age, or
- Discrepant by more than 10 years on year of first birth.

These represent a relatively conservative set of criteria, but it was decided to err on the cautious side and risk not identifying all incorrect matches rather than risk identifying too many incorrect matches. However, this does mean that the data are likely to contain additional, less extreme mismatches that have not been identified.[1] Based on these criteria, 31 women were identified as incorrect matches, leaving 474 women for the analysis. Of the 31 women identified as probable mismatches, 23 did not remember being interviewed before, suggesting that these women probably were incorrectly identified for the reinterview survey. The distribution of the reinterview sample, excluding the incorrect matches, by various background characteristics is presented in the final column of Table 2.1. The distribution for this sample for all variables is virtually identical to that for the full reinterview sample.

[1] Several tabulations were also run using a much narrower set of criteria for excluding mismatches. The set of criteria used was: discrepant by three or more on children ever born or discrepant by five years or more on current age or discrepant by five years or more on year of first birth. This set of criteria resulted in far more women being classified as incorrect matches, leaving a sample of 381 women. In addition, using this set of criteria, rural women with no education were more likely to be excluded than other women. However, the general findings did not change substantially for any of the variables examined.

Chapter 3

Measures of Reliability

The primary aim of the reinterview survey is to assess the reliability of reporting of key variables in the main Pakistan DHS. In general terms, reliability can be defined as "*the extent to which a measurement remains constant as it is repeated under conditions taken to be constant*" (Kaplan, 1964, p. 200). The degree of consistency in the responses of the same individuals to the same question in the main survey and the reinterview survey can thus be used as an indicator of the reliability of the responses to that question. However, it is important to be aware that consistency does not guarantee validity of the responses since consistent biases may be repeated or errors may exist in the responses in both surveys. In addition, respondents in the reinterview survey might recall the answer they gave in the main survey and repeat it, and consequently a consistent response might indicate good recall on the part of the respondent rather than true reliability of the response.

Inconsistencies in responses may arise for a number of reasons. First, there may be an error in one or both of the responses. The error may be on the part of the respondent, for example, because she does not know the answer and guesses or because she chooses to give an incorrect answer. Interviewers may also make errors, errors may occur in data entry, or errors may in some way be due to the design of the survey. However, not all inconsistencies imply that one or the other of the responses are incorrect. There may also have been a genuine change in the situation of the respondent, for example, she may have had another child or experienced a child death between the two surveys. In many situations, this can be adjusted for to some extent. In addition, the respondent may have changed her attitude or opinion about an issue between the surveys. This is particularly relevant to attitudinal questions such as those on fertility preferences. Such genuine changes in the responses only apply to some variables though, and some of the factual information collected should not change between the surveys. For example, the woman's date of birth or childhood place of residence should always be the same.

A final important point that affects the use of the consistency of responses in the two surveys as an indicator of response reliability is that it is virtually impossible to repeat the survey conditions exactly in the two surveys. Indeed, the conditions in the reinterview survey are known to differ from those in the main survey in that the reinterview survey used better interviewers and a shortened form of the questionnaire, which excluded the health and anthropometric sections. Consequently, differences in the responses in the two surveys may also be due to such differences in the survey implementation, and hence the consistency of the responses is influenced by both the reliability of the responses and their sensitivity to the details of survey implementation. However, the impact of survey practices on the responses is also of interest, and inconsistencies may provide useful indications about areas of weakness in survey implementation that need to be addressed.

The simplest measure of consistency is the crude index of disagreement, which measures the percentage of cases with different responses in the two surveys. For metric variables, similar indices can be used to describe the distribution of the discrepancies, such as the percentage of women discrepant by more than a particular amount, e.g., by more than five years on their year of birth. These crude indices of discrepancy are extremely useful for descriptive purposes, but they suffer from a number of weaknesses. The first is that they do not allow for the fact that, even if the responses recorded in the two surveys were totally independent, some women would, by chance, give the same response. One index that overcomes this problem is kappa, which represents a measure of the observed level of agreement compared to the expected level of agreement if responses in the two surveys were totally random.

Kappa is defined as:

$$\hat{K} = \frac{\sum_i p_{ii} - \sum_i p_{i.}\, p_{.i}}{1 - \sum_i p_{i.}\, p_{.i}} \qquad (1)$$

where p_{ii} = proportion of women in category i on a particular variable in both the reinterview survey and the main survey

$p_{i.}$ = proportion of women in the i^{th} row, i.e., recording category i in the reinterview survey

$p_{.i}$ = proportion of women in the i^{th} column, i.e., recording category i in the main survey.

The second problem is that, for ordinal or metric variables, the crude indices do not allow for the degree of discrepancy in an individual woman's reported responses. For example, a woman who reports two births in the main survey and three births in the reinterview survey will not be considered any differently from a woman who reports one birth in the main survey and four in the reinterview survey. Clearly, the second case is more discrepant than the first one, and it would be useful to consider this in the index of discrepancy used. Kappa also fails to take the degree of discrepancy into account and is very sensitive to any grouping of continuous or ordinal variables (Maclure and Willett, 1987). One index that is often used to solve these problems is the weighted kappa index, which is defined as:

$$\hat{K}_w = \frac{P_o^* - P_e^*}{1 - P_e^*} \qquad (2)$$

where

$$P_o^* = \sum_i \sum_j w_{ij}\, p_{ij} \qquad (3)$$

$$P_e^* = \sum_i \sum_j w_{ij}\, p_{i.}\, p_{.j} \qquad (4)$$

and p_{ij} = proportion of women in category i on a particular variable in the reinterview survey and category j in the main survey

$p_{.j}$ = proportion of women in the j^{th} column, i.e., in category j in the main survey

w_{ij} = the weight associated with cell ij of the cross-classification of the responses to a particular question in the main survey and the reinterview survey.

A number of possible weights can be used (O'Muircheartaigh, 1982), but in practice it is advisable to use a standard set of weights to enable meaningful interpretation and comparison of weighted kappa across studies (Maclure and Willett, 1987). One that is suitable for metric variables, such as the number of births reported by women, is

$$w_{ij} = 1 - (i-j)^2. \qquad (5)$$

These weights were used for calculation of weighted kappa for metric variables in the World Fertility Survey (WFS) reinterview studies and, for comparability, are used in this report for metric variables. This form of weighted kappa also has the advantage that it is approximately equal to the intraclass correlation coefficient if the marginal distribution of the variable is approximately the same in the two surveys.

For ordinal variables, such as level of education, the weight used is

$$w_{ij} = 1 - |i-j|/(L-1), \tag{6}$$

where L is the number of categories of the variable. For other categorical variables which do not have ordered categories, including dichotomous variables, unweighted kappa is presented. In fact, for dichotomous variables, unweighted kappa is identical to weighted kappa using either of the weights defined above. Note that the various forms of kappa are not directly comparable because they are based on different sets of weights.

The interpretation of kappa is not as obvious as the interpretation of the crude indices of discrepancy. If the responses in the two surveys are in complete agreement, kappa will equal one. If the observed agreement is equal to that expected by chance, kappa will equal zero. Thus, values of kappa between zero and one indicate better agreement than expected by chance, and values of kappa below zero indicate poorer agreement than expected by chance. For most purposes, values of kappa and weighted kappa above 0.75 indicate excellent agreement beyond chance, values below 0.40 represent poor agreement beyond chance, and values in the range 0.40-0.75 represent fair to good agreement beyond chance (Fleiss, 1981, p. 218).

Chapter 4

Consistency of Reporting of Background Characteristics

Information on a number of background characteristics of the woman was collected in the reinterview survey. However, not all women were asked all the background questions in both surveys, due to the nature of the skip pattern used. For example, women who report that they have never attended school are not asked what level of school they reached. Consequently, if a woman reports inconsistently on whether she ever attended school, she will automatically be inconsistent on the level of school she reached. To isolate inconsistencies due purely to inconsistent responses to a particular question from inconsistencies due to "carry-over" effects from earlier inconsistent responses, only women who were actually asked the particular question in both surveys are included in the analysis for that variable.

Table 4.1 presents the percentage of women who did not give the same response in the two surveys for each background variable, together with the value of kappa and the number of cases on which the analysis is based. The form of kappa used varies slightly according to the form of the variable being analyzed. Level of education and reading ability are ordinal variables and consequently kappa is calculated using the weights for ordinal variables defined in equation (6). Unweighted kappa is used for all the remaining variables as they are either dichotomous or categorical with unordered categories. A small number of cases have missing values for some of the variables in the reinterview survey, and these are excluded from the analysis.

Table 4.1 Indices of disagreement for reliability of reporting of background characteristics, Pakistan DHS, 1990-91

Background characteristic	Percent discrepant	Kappa	Number
Marital status	1.2	0.82	474
Number of times married	2.2	0.45	473
Blood relationship between spouses	17.4	0.64	474
Nature of blood relationship	37.0	0.40	242
Writing ability	22.5	0.53	52
Reading ability	11.4	0.71*	407
Attended school	6.4	0.83	474
Educational level	12.1	0.89*	100
Residence under age 12	11.1	0.74	472
Native language	10.4	0.86	469

Note: Figures are based on women who answered each question in both surveys.
* These variables use weighted kappa (see text).

The most consistently reported variables are those relating to marital status, with only one percent of women discrepant on their current marital status and only two percent discrepant on the number of times they had been married. One of the factors that contributes to this high level of consistency is the relative homogeneity of the sample. In both surveys around 96 percent of women report that they are currently married and that they have been married only once. This is reflected in the relatively low value of kappa for the number of times married, which suggests only moderate reliability of this variable despite its very high consistency. However, the value of kappa for marital status is quite high, suggesting that the reporting of this characteristic in the two surveys is both reliable and consistent.

In contrast, the responses to the questions concerning blood relationships between spouses are less consistent. Table 4.1 shows that 17 percent of women did not give the same response in the two surveys when asked whether there was any blood relationship between themselves and their husband and of those who reported a relationship in both surveys, 37 percent were discrepant on its nature. Examination of the individual tables suggests that the majority of discrepancies on the existence of a blood relationship between

the woman and her most recent husband are due to women who reported no blood relationship in the main survey but reported that there was a blood relationship in the reinterview survey. No clear pattern emerges in the discrepancies in the responses to the nature of the blood relationship. This may suggest some confusion among women concerning the precise classification of the blood relationship between themselves and their spouse. Women often report a *bradari* relationship with their husband which could be any type of relationship along fraternal or paternal lines and the exact nature of the relationship may not be clear in many cases (Shami, Grant and Bittles, 1994).

Writing ability is also reported less consistently than most of the other background characteristics, with 22 percent of women who were asked this question in both surveys discrepant in their response. However, because this question was only asked of women with less than middle school education who reported that they could read, either easily or with difficulty, only 52 women actually responded to it in both surveys and only 12 women gave discrepant answers.

Reading ability is reported more reliably and more consistently, with 11 percent of women discrepant in their response to this question. However, 83 percent of the discrepancies are movements to and from the "with difficulty" category, with slightly more women moving between the "with difficulty" and "can't read" categories than moving between the "with difficulty" and "easily" categories. This suggests that the definition of "with difficulty" is a little vague for many women, and their interpretation varied in the two surveys.

The consistency of whether the woman ever attended school is very high—only six percent of women are discrepant on this variable—and reliability is relatively high (kappa = 0.83). The discrepancies that do occur are fairly random in direction. The level of education reached is also reported reliably, although consistency is slightly lower, with 12 percent of women discrepant. This level of consistency and reliability is similar to that obtained in the Indonesia WFS reinterview study (MacDonald, Simpson and Whitfield, 1978). The relatively high reliability is partly due to the fact that all discrepancies that do occur in the reported level of education reached are discrepancies of only one level, most commonly between middle and secondary. It is possible that women who attended the higher level for only a short amount of time classified that attendance in different ways in the two surveys.

In the WFS reinterview study in Indonesia, place of residence before age 12 was found to be reported quite unreliably, with 31 percent of women discrepant, and kappa having a value of 0.45 (MacDonald, Simpson and Whitfield, 1978). In Pakistan, only 11 percent of women are discrepant on their childhood place of residence, and the reliability of this response is much higher, with kappa equal to 0.74. The main explanation proposed for the relatively low consistency of the reporting of childhood place of residence in Indonesia is that it is essentially an opinion question, and women's interpretation of city versus town or village may vary, especially in situations in which the level of urbanization of the place had changed markedly since the woman's childhood. The higher consistency in the Pakistan reinterview survey suggests less confusion in interpreting the question than in the Indonesia WFS, which in turn may be due to differences in the wording of the question, but may also reflect the slower pace of urbanization in Pakistan compared to Indonesia.

The final background characteristic examined is the native language of the respondent. Somewhat surprisingly, 10 percent of women are discrepant on their native language, although overall reliability is relatively high, with kappa equal to 0.86. Examination of the full cross-tabulation showed that the most common discrepancies are women who report Urdu as their native language in the main survey but Punjabi in the reinterview survey. Most of the remaining discrepancies are also shifts between Urdu and other languages, especially from other languages in the main survey to Urdu in the reinterview survey. The reason for these discrepancies is unclear but may result from the fact that Urdu is the national language of Pakistan. Women may speak both the national language and their own regional language, and consequently may report

either as their native language. Discrepancies involving shifts between regional languages probably represent mismatches.

Finally, it is important to note that with the exception of residence under age 12 and native language of respondent all the background variables discussed could change between the two surveys, and consequently some of the discrepancies reported could represent genuine changes. However, the magnitude of change would be expected to be small and is unlikely to explain much of the observed discrepancies. The majority of discrepancies are likely to be due to errors of some form in the responses, differences in the interpretation of some questions by respondents, or mismatches.

Chapter 5

Age and Date Reporting

5.1 Comparison of Basic Indices of Discrepancy

Perhaps the most crucial information collected for any demographic analysis is the information on ages and dates of demographic events. However, in many societies, age and date reporting are known to be very poor. Pakistan is a very good example of such a society, and problems in the collection of age data have been documented in every demographic survey and census ever conducted in the country. The poor quality of age data in Pakistan is seen in severe age heaping, particularly on ages ending in zero, and in a tendency to exaggerate ages, particularly among older women and high-parity women (Retherford and Mirza, 1982). Inconsistencies in the age distribution of the population have been observed between surveys conducted close together; for example, the age distribution of respondents in the 1973 Housing, Economic and Demographic (HED) Survey was found to differ substantially from that obtained in the 1972 census (Retherford and Mirza, 1982). These problems stem from the fact that very few women know their age or date of birth. In the 1975 Pakistan Fertility Survey, for example, only 6 percent of women were able to report both the year and month of their birth (Retherford and Mirza, 1982), although in the Pakistan DHS this figure increased to 23 percent. Given this long history of problematic age and date reporting, the consistency of age and date reporting between the Pakistan DHS and the reinterview survey is expected to be poor. It is important, however, to investigate the degree of discrepancy between the two surveys and, in particular, to investigate whether there are any systematic differences in age and date reporting between the two surveys.

In this section the reporting of ages and dates of four events are analyzed: the date of birth and current age of the woman, the date and age of the woman at first marriage, the date and age of the woman at her first live birth, and the date and age of the woman at her most recent live birth. For each event the consistency and reliability of the reported year and month of the event are analyzed separately, and the analysis is restricted to women who reported the year (or month) of the particular event in *both* surveys. Hence no imputed information is used in the analysis of date reporting. However, the examination of current age and age at which each event occurred does use imputed information (for the methodology of imputation, see Croft, 1991). The main reason for this approach is to allow all matched women to be included in the analysis of age reporting. Although 24 percent of the reinterviewed women gave both the year and the month of their birth in each survey, only 44 women (9 percent) gave both the year and month of their birth in both surveys, so an analysis of the reporting of current age that does not include imputed information is not feasible. This also applies to the analysis of the reliability of age at each of the other events being considered, since the age of the woman at the time of each event is calculated from the difference between her reported date of birth and the reported date of the event. In addition, most analyses of Pakistan DHS data use the imputed ages of women, so some assessment of the reliability of this information is useful.

Table 5.1 presents crude indices of disagreement and weighted kappas for reporting of year, month and age of each of the events of interest. In the calculation of weighted kappa, age, month, and year are all treated as metric variables, so the weight defined in equation (5) is used. Thus, if a woman reports the month of the event as December in the main survey, but January in the reinterview survey, the discrepancy will be 11 months. This approach was also adopted in the WFS analyses, although there is a case for treating such a discrepancy as a discrepancy of only one month. In addition, it is important to note that the values of weighted kappa calculated for current age and the age at each event are not comparable with those obtained

in the WFS study in Indonesia, since a different set of weights was used for age variables in that analysis.

From Table 5.1 it is clear that age and date reporting are quite poor in general. Age at most recent birth is reported least consistently, followed by age at first birth, current age, age at first marriage, and year of birth of the woman. Not surprisingly, the calendar years in which more recent events occurred are reported more consistently than the years in which events further back in time occurred. In particular, women have a much better idea of the year of birth of their children than of their own year of birth. In contrast, the month in which more recent events occurred is reported less consistently than the month in which events further back in time occurred. However, this may be due to selection effects because women are much more likely to report a month for recent events than for events further back in time.

Table 5.1 Indices of disagreement for reporting of year, month and age of the woman's birth, first marriage, first live birth and most recent live birth, Pakistan DHS, 1990-91

Date/Age of event	Percent discrepant	Percent discrepant by 5+ years	Weighted kappa	Number
Year of birth	81.2	23.6	0.88	467
Month of birth	48.1	NA	0.69	44
Current Age	83.4	24.0	0.88	474
Year of 1st marriage	72.8	9.9	0.94	458
Month of 1st marriage	56.5	NA	0.54	177
Age at 1st marriage	81.4	21.1	0.37	474
Year of 1st birth	67.2	6.4	0.95	433
Month of 1st birth	66.8	NA	0.33	335
Age at 1st birth	84.7	24.6	0.36	436
Year of last birth	48.5	3.6	0.90	432
Month of last birth	70.3	NA	0.38	380
Age at last birth	86.2	29.9	0.77	434

Note: Figures for year and month of each event are based only on women who gave a year (or month) for that event in both surveys. Figures for age at each event are based on all matched reinterviewed women who reported the event in both surveys. Imputed ages are used when necessary.
NA = Not applicable.

One of the reasons for the poor consistency of the ages at which events occurred is that age is calculated from the date of birth of the woman and the reported date of the event, and is therefore affected by discrepancies in the reporting of both of these dates. In particular, the woman's year of birth, which is a crucial factor in the calculation of ages at other events, is reported very inconsistently. Consequently, since the majority of women are inconsistent on their year of birth, they would also have to be inconsistent on the year of subsequent events by the same amount in order to have the same age at the event in the two surveys. The fact that age at last birth is the least consistently recorded age measure reflects the fact that year of last birth is reported more consistently than other dates, and hence the discrepancy in the woman's age cannot be "canceled out" for so many women.

The values of weighted kappa presented in Table 5.1 suggest that the year in which events occurred is reported more reliably than the month or age at the event. The current age of the woman and the age at most recent birth are reported with moderate reliability, but age at first birth and age at first marriage are reported much less reliably. This finding appears to contradict the earlier discussion, which found that age at last birth was recorded less consistently than either age at marriage or age at first birth, but it is due, at least in part, to the fact that these three variables have very different distributions, which affects the value of weighted kappa. Both age at first birth and age at first marriage are concentrated over a relatively small range of values around the late teens and early twenties, whereas the distribution of age at most recent birth is over the entire reproductive age range with no dominant values. In this latter situation, weighted kappa is not a very sensitive indicator of reliability because the discrepancies are small compared to the range. Consequently, the age at most recent birth appears to be much more reliable than age at first birth and age

at first marriage. This is one of the disadvantages of using weighted kappa for comparison of the reliability of variables with very different distributions and ranges.

Weighted kappa is generally lower for month of the event than for either year of or age at the event, the exception being for first marriage, where weighted kappa is lower for age than for month. This suggests that the reliability of reporting is lower for the month of the event than for the other variables. However, one of the reasons for the lower values of weighted kappa is that the month of the event is distributed over a much smaller range of values than other variables, as discussed above. The woman's month of birth and month of marriage are reported more reliably than the month of birth of the children. Again, this probably reflects selection effects, since very few women actually report month of birth or month of marriage. Thus, those who do are likely to represent a group of women with superior knowledge of the dates of these events.

Table 5.2 presents the crude index of disagreement for each age and date variable for Pakistan, together with those obtained from the WFS reinterview studies conducted in Indonesia, Peru, and Lesotho. The table demonstrates that the percentage of discrepant cases is consistently higher in Pakistan than in the WFS studies. For example, 24 percent of women in Indonesia, 49 percent of women in Peru, and 29 percent of women in Lesotho, were discrepant on year of first marriage, compared to 73 percent in Pakistan. However, some caution should be exercised in interpreting these results, as the completeness of reporting of events varies between these surveys. For example, in the Indonesia study the analysis of year of first marriage was based on the 65 percent of the reinterviewed sample who reported year of first marriage in both surveys, whereas in Pakistan the analysis was based on 97 percent of the reinterviewed sample. Hence the results for Indonesia probably represent a select group who are likely to report dates more consistently than the general population. It is not clear how many women were missing information on ages and dates in the Peru and Lesotho studies or how such cases were handled. In addition, poorer date reporting would be expected in Pakistan, since the level of female education and literacy is much lower than in Indonesia, Peru or Lesotho.

Table 5.2 Percentage of women with discrepant responses on age and date of selected events for Pakistan, Indonesia, Peru, and Lesotho

	Percent discrepant			
Date/Age of event	Pakistan	Indonesia	Peru	Lesotho
Year of birth	81	21	-	-
Month of birth	48	16	-	-
Current Age	83	57	46	40
Year of 1st marriage	73	24	49	29
Month of 1st marriage	57	30	-	-
Age at 1st marriage	81	63	54	51
Year of 1st birth	67	21	29	25
Month of 1st birth	67	26	36	-
Age at 1st birth	85	63	-	-
Year of last birth	49	-	25	23

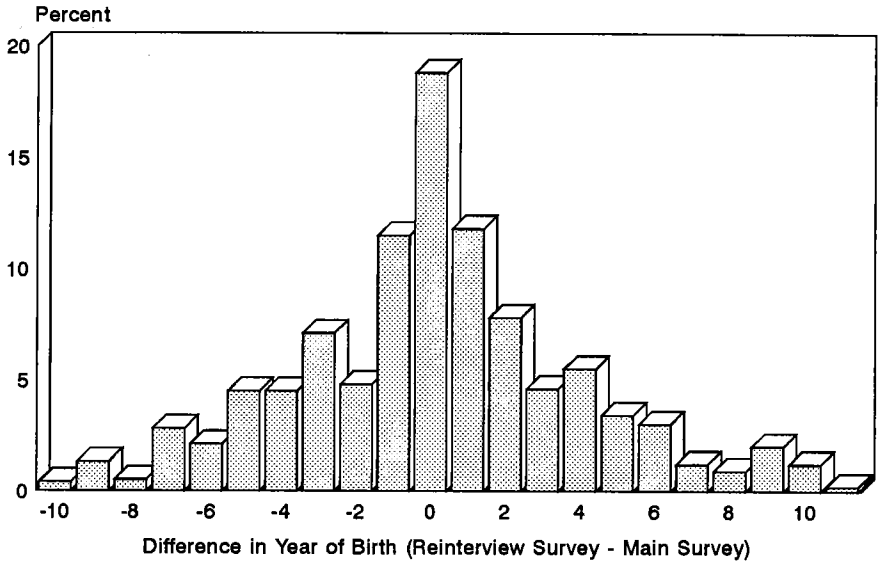
Source: Indonesia (MacDonald, Simpson & Whitfield, 1978); Peru (O'Muircheartaigh, 1984a); Lesotho (O'Muircheartaigh, 1984b).
- Not available.

The comparisons of these indices of age and date reporting provide useful insights into the relative consistency and reliability of the reporting of the ages and dates of various events. However, in order to gain a deeper understanding of the nature and implications of these discrepancies, it is necessary to examine the reporting of each age and date in detail. This is done in the following sections.

5.2 Date of Birth and Age of Respondent

Year of birth was reported in both surveys by 467 of the 474 matched women, representing 98 percent of the matched sample. However, only 19 percent of these women reported the same year of birth in the two surveys, and 24 percent were discrepant by five years or more. Figure 5.1 shows the distribution of the difference between the year of birth reported in the reinterview survey, and that reported in the main survey. One of the most striking features of this figure is that it is very symmetrical around zero, where the majority of cases are clustered. This suggests that the errors in the reported year of birth are fairly random, and there is no tendency by women to systematically report a later or earlier year of birth in the reinterview study. Only nine percent of women reported month of birth in both the main and the reinterview survey, so the examination of the consistency of the reported month of birth is based on a very small number of women. Despite the fact that women in this group are likely to be very select and to have superior knowledge of their date of birth, 48 percent did not report the same month of birth in the two surveys. The most commonly mentioned months in the main survey were October, January, and June, whereas in the reinterview survey June, July, and August were mentioned most frequently.

Figure 5.1
Percent distribution of matched women by difference in year of birth between reinterview survey and main survey, Pakistan DHS, 1990-91



Note: Based on women who gave a year of birth in both surveys (N=467).

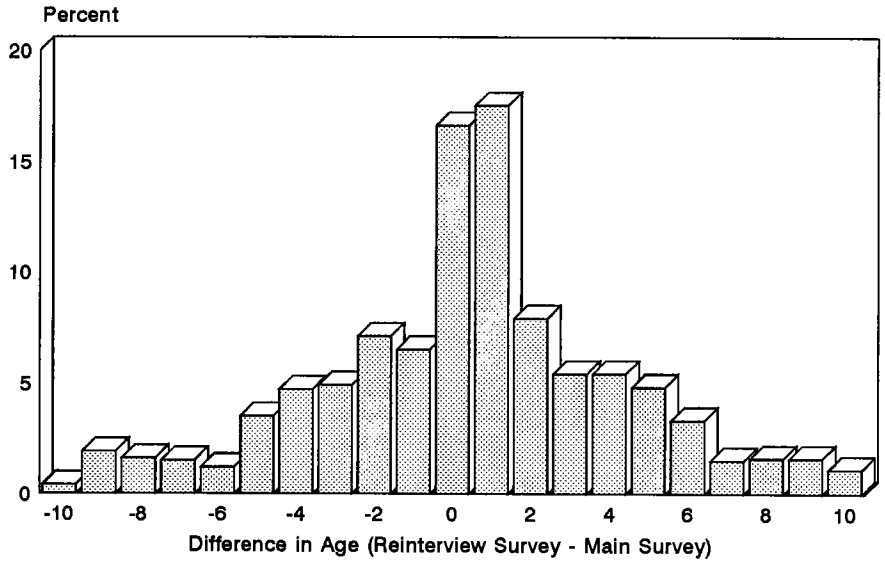
An examination of the difference in reported age in the two surveys confirms that consistency of age reporting is low. Only 17 percent of women report the same age in the two surveys, although another 18 percent report themselves to be one year older in the reinterview survey, which could be a genuine change in age (Table 5.3). No attempt was made to adjust for genuine increases in age at the individual level since so few women actually report a month of birth. Figure 5.2 presents the distribution of the difference in age reported in the reinterview survey and the main survey. In fact, 55 percent of women in the reinterview

Table 5.3 Percent distribution of women by difference in reported age in the reinterview survey and the main survey, by age group in the main survey, Pakistan DHS, 1990-91

Difference in reported age	Age group reported in main survey								
	15-19	20-24	25-29	30-34	35-39	40-44	45-49	Total	Number
6-10 years younger	0.0	0.0	0.0	11.8	8.4	12.3	13.9	6.6	31
2-5 years younger	1.6	6.8	9.0	22.9	28.9	33.1	36.1	20.2	96
1 year younger	9.4	4.6	10.8	4.2	7.5	6.4	0.0	6.5	31
Same age	14.3	22.5	13.1	17.0	20.7	16.9	7.6	16.6	79
1 year older	16.8	22.8	19.6	17.6	17.7	6.2	21.4	17.5	83
2-5 years older	36.8	26.3	34.9	17.2	10.9	23.1	20.9	23.5	112
6-10 years older	21.1	17.0	12.5	9.2	5.8	2.0	0.0	9.2	43
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	
Number	22	70	106	82	82	68	45		474

survey reported their age within two years of that given in the main survey, and 76 percent reported it within four years of that given in the main survey. This degree of consistency in age reporting is still somewhat lower than was found in the Indonesia WFS reinterview survey, in which 88 percent of women reported their age in the reinterview within four years of that in the main survey (MacDonald, Simpson and Whitfield, 1978). However, this is to be expected given the low level of knowledge of date of birth among women in Pakistan.

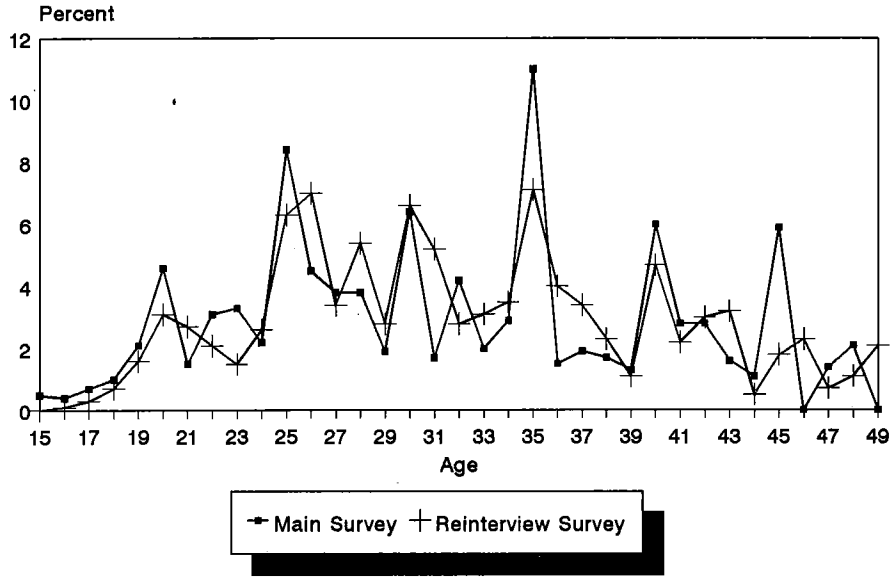
Figure 5.2
Percent distribution of matched women by difference in age between reinterview survey and main survey, Pakistan DHS, 1990-91



Note: Based on all women (N=474).

Again, Figure 5.2 is very symmetrical around the central peak, which suggests that misreporting of age is random, and there is no evidence of any tendency to report systematically higher ages in either survey other than would be expected due to genuine increases in age. This conclusion is supported further by Figure 5.3, which presents the age distribution of women in single years in each survey. The two distributions are very similar despite the inconsistencies in the age reporting of individual women. The mean age of the sample in the main survey is 31.8 compared to 32.2 in the reinterview survey. In contrast, the median age in the main survey (32.0) is slightly higher than in the reinterview survey (31.7). The index of dissimilarity can be used to measure the overall agreement between the two age distributions (Shryock and Siegel, 1971, pp. 232-233). For these two distributions the index of dissimilarity is 22.7 percent, which can be interpreted as indicating that 22.7 percent of women would have to change their reported age for the two distributions to match exactly. However, these two distributions would not be expected to match exactly, since on average the two surveys are 7.8 months apart, so many women would be expected to be one year older in the reinterview survey. The age distribution in the main survey can be adjusted to the time of the reinterview survey by assuming that 65 percent (100x7.8/12) of women at every age would have had a birthday during the period between the surveys. The index of dissimilarity calculated after this adjustment is 16.1 percent, suggesting that the two age distributions are in quite close agreement. This in turn confirms that, overall, upward discrepancies in the age reported in the reinterview survey are balanced by corresponding downward displacements.

Figure 5.3
Percent distribution of matched women by age in single years in main survey and reinterview survey, Pakistan DHS, 1990-91



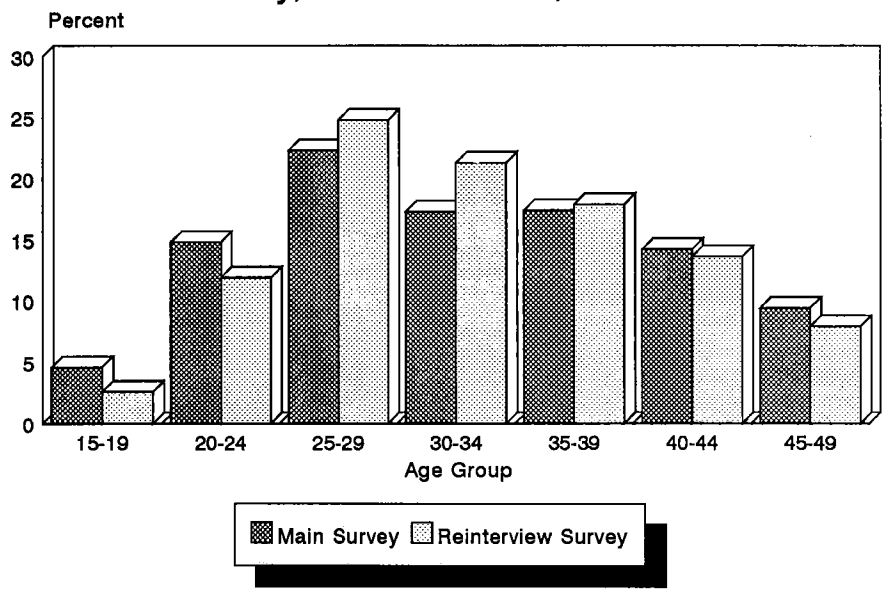
Note: Based on all women (N=474).

Both distributions display considerable age heaping, particularly on ages ending with a zero or a five, which again reflects the fact that women in Pakistan have only an approximate idea of their age. This heaping is more pronounced in the main survey than in the reinterview survey, which may in part reflect the fact that better interviewers were used in the reinterview survey. An alternative explanation is that interviewers had much more time in the reinterview surveys since the questionnaire was so much shorter than in the main

survey, and consequently they were able to put more emphasis on obtaining better quality age data. However, this does not mean that the age reported in the reinterview survey is correct.

Although the two distributions in Figure 5.3 are very similar overall, there are fewer women under 21 in the reinterview survey than in the main survey. This is seen more clearly in Figure 5.4, which presents the age distribution of women in five-year age groups for the two surveys. In the main survey there are more women in the 15-19 and 20-24 year age groups and, correspondingly, fewer women in the 25-29 and 30-34 year age groups, than in the reinterview survey. The percentage of women age 35-39 and 40-44 is very similar in the two surveys, but in the 45-49 year age group there are again more women in the main survey than in the reinterview survey. This is due to the tendency for the ages of women in the reinterview survey to be shifted towards the mean, particularly for younger women.

Figure 5.4
Percent distribution of matched women by age in five-year age groups in main survey and reinterview survey, Pakistan DHS, 1990-91



Note: Based on all women (N=474).

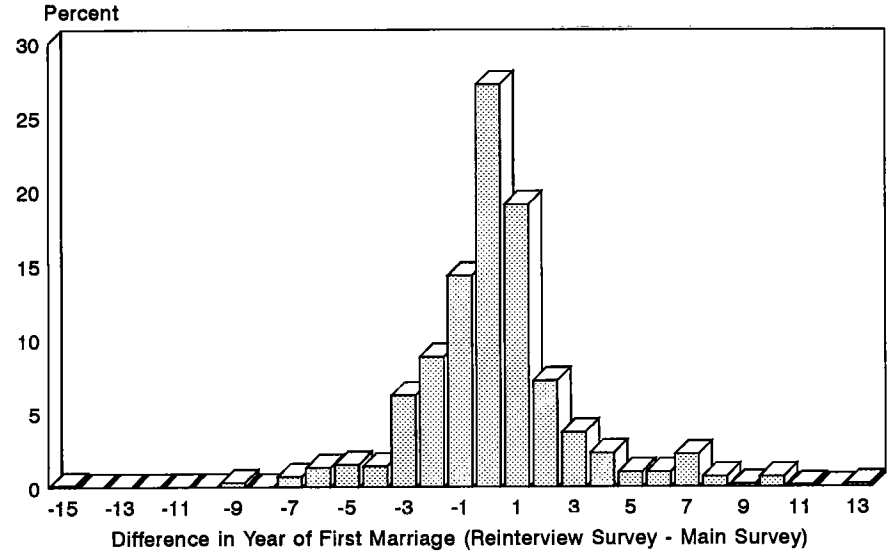
This is clearly seen in Table 5.3, which presents the discrepancy in reported age between the reinterview survey and the main survey, by age reported in the main survey. Younger women, particularly those under 30, are more likely to report an older age in the reinterview than a younger one. After age 30, the discrepancy in age reporting between the surveys is much more random, although as the age at main survey increases, there is an increasing tendency to report a younger age in the reinterview survey. This shift in age reporting towards the mean in the reinterview survey may be partly due to the fact that interviewers expect women in the reinterview survey to be age 15-49, and consequently may be reluctant to accept an age reported outside of this range. Indeed, there are no women in the reinterview survey who report an age outside of the range 15-49. This inevitably restricts the degree of discrepancy possible in one direction for younger and older women. In addition, biological and social indicators may also help place bounds on age reporting, especially for younger women. For example, if a young woman has started menstruating and is married, she may know she is older than 15, but she may not know exactly how much older.

5.3 Age and Date of First Marriage

In both the main survey and the reinterview survey ever-married women were asked to report the date and age when they started cohabiting with their husband. Probing was used to differentiate the *nikah* (the marriage contract ceremony) from the actual *rukhsati* (the departure for the husband's household, i.e., consummation of the marriage). Hence, the age and date of first marriage reported should refer to the age and date of consummation of the first marriage.

Year of first marriage was reported in both surveys by 458 of the matched women, slightly fewer than reported their year of birth in both surveys. However, year of first marriage appears to be reported slightly more consistently than year of birth, with 73 percent of women not reporting the same year of marriage in the two surveys, and only 10 percent of women discrepant by more than five years. Again the errors in year of first marriage appear to be random, as seen by the fairly symmetrical shape of Figure 5.5, which shows the distribution of the difference in the reported year of first marriage between the reinterview and the main survey. There is a slight tendency to report one year later in the reinterview survey more often than one year earlier. However, this is compensated for by a slight tendency to report two or three years earlier in the reinterview survey more often than two or three years later.

Figure 5.5
Percent distribution of matched women by difference in year of first marriage between reinterview survey and main survey, Pakistan DHS, 1990-91

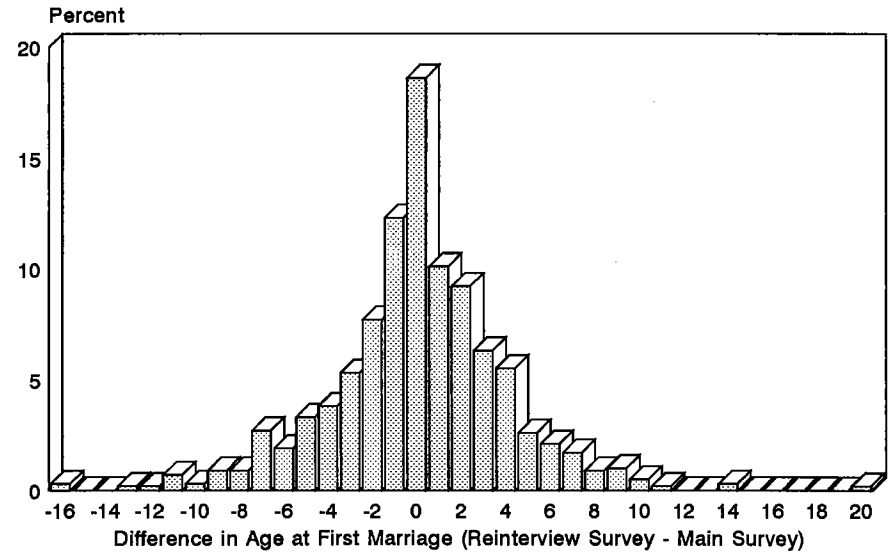


Note: Based on women who gave year of first marriage in both surveys (N=458).

Of the 474 matched women, 177 (37 percent) reported a month of first marriage in both surveys. However, slightly more than half of these women did not report the same month of first marriage. In the main survey, April, March, and July were the most commonly mentioned months of first marriage, whereas in the reinterview survey August, October, and April were the most frequently cited.

Age at first marriage was calculated from the date of birth and date of first marriage for all women, using imputed information when necessary. Overall, 81 percent of the matched women were discrepant on the age at first marriage calculated in the two surveys, and 21 percent were discrepant by more than five years. As discussed earlier, the higher level of discrepancy in age at first marriage compared to year of first marriage is due, at least in part, to the combined effect of errors in the reporting of both date of birth and date of first marriage. However, the errors in the reported age at first marriage appear to be random, as seen in Figure 5.6.

Figure 5.6
Percent distribution of matched women by difference in age at first marriage between reinterview survey and main survey, Pakistan DHS, 1990-91



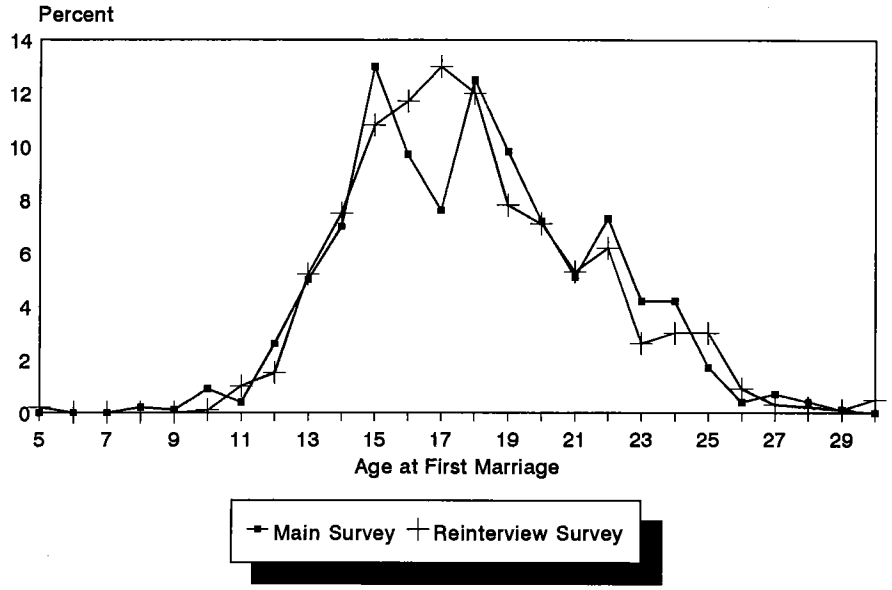
Note: Based on all women (N=474).

Consequently the distribution of age at first marriage is very similar in the two surveys, as seen in Figure 5.7, where it is presented in single years for the two surveys. The index of dissimilarity for these two distributions is 11.5 percent, suggesting that only a small percentage of women would have to change their reported age at first marriage for the two distributions to coincide exactly. The distribution is smoother in the reinterview survey than in the main survey. In particular, in the main survey there appears to be some heaping on ages 15 and 18 that is not present in the reinterview survey. The median age at first marriage for matched women age 20-49 (as reported in the main survey) is 19.1 in the main survey but is slightly lower (18.7) in the reinterview survey. However, both estimates are close to the value of 18.9 obtained for the full Pakistan DHS sample, suggesting that errors in the reporting of age at first marriage have relatively little impact on this figure.

5.4 Age and Date of First Live Birth

A total of 436 matched women had at least one birth reported in the birth history in both surveys, and 433 of these reported the year of the first birth in both surveys. However, it is important to note that there is an additional source of discrepancy when comparing the dates of births from within the birth history. If

Figure 5.7
Percentage distribution of matched women by age at first marriage in main survey and reinterview survey, Pakistan DHS, 1990-91



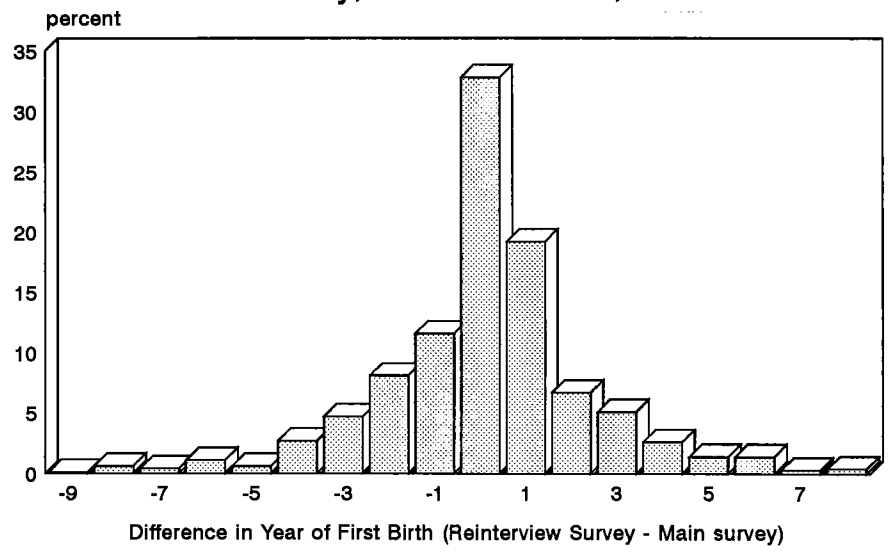
Note: Based on all women (N=474).

births are omitted at the start of the birth history in one or the other of the surveys, the date of birth identified for the first birth in this analysis may not correspond to the same child, and consequently discrepancies in the reported year of first birth may be due to omissions of births as well as discrepant reporting. Despite this potential problem, women appear to recall the year of first birth more consistently than either their own year of birth or their year of first marriage. Just under one-third of women report the same year of first birth in the two surveys, and only six percent are discrepant by more than five years. One possible explanation for the better reporting of year of first birth is that the woman may know the age of her oldest child, at least within a year or two, and consequently has some guide as to the year of her first birth. This is particularly true if the oldest child survived.

Figure 5.8 shows the distribution of the difference between the year of first birth reported in the reinterview survey and the main survey. Again the graph is fairly symmetrical, suggesting random errors in the reporting of year of first birth. The general shape is very similar to that seen in Figure 5.5 for year of first marriage, which suggests that women or interviewers may use the year of birth of the oldest child as a guide to the year of marriage, or vice versa. This seems plausible given that contraception is not widely used, especially before the first birth, so most women are likely to conceive fairly soon after marriage.

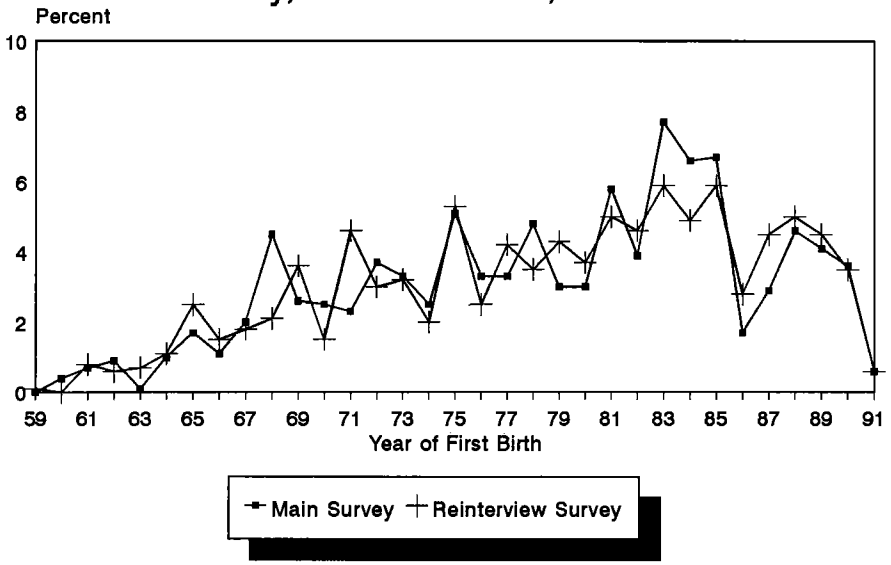
The distribution of year of first birth is very similar in the two surveys, as shown in Figure 5.9. However, there is some irregularity in the distribution around 1986 and 1987, when the percentage of first births drops dramatically and then picks up again. This is seen in both surveys, but is more pronounced in the main survey and is probably due to shifting of births that occurred in 1986 and 1987 to an earlier period, most notably 1983-85, in order to avoid the long health and anthropometric sections of the questionnaire in the main survey. However, the fact that a similar pattern is observed in the reinterview survey suggests that heaping of the age or year of birth of the oldest child may also be part of the explanation. There is evidence of heaping of first births in other years too, and again this is a little more pronounced in the main survey than in the reinterview survey. This is discussed further in Section 6.2.

Figure 5.8
Percent distribution of matched women by difference in reported year of first birth between reinterview survey and main survey, Pakistan DHS, 1990-91



Note: Based on women with at least one birth in both surveys who gave the year of 1st birth in both surveys (N=433).

Figure 5.9
Percent distribution of matched women by year of first birth reported in main survey and reinterview survey, Pakistan DHS, 1990-91

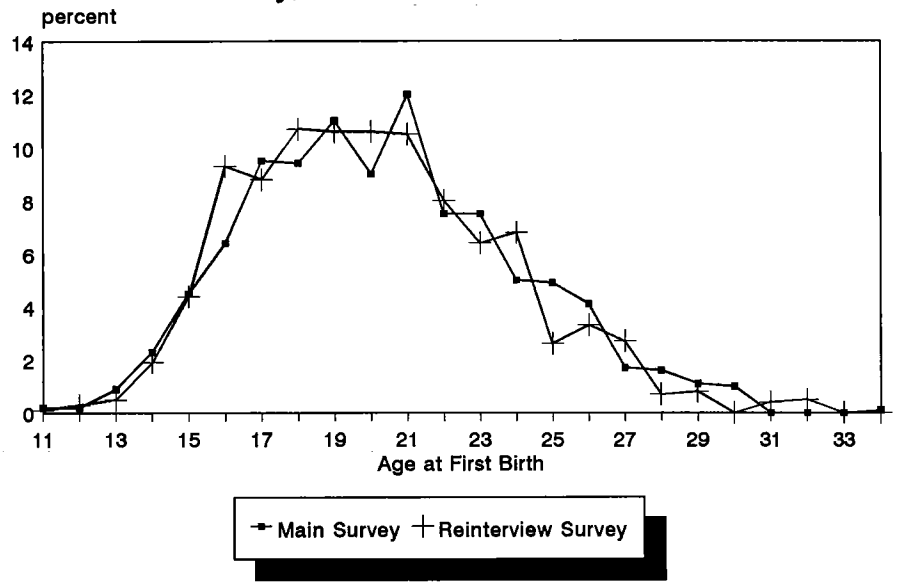


Note: Based on women with at least one birth in both surveys who gave the year of 1st birth in both surveys (N=433).

Month of first birth is also reported by the majority of women who report a birth in both surveys (77 percent), and one-third report the same month of birth in the two surveys. Births are reasonably evenly spread throughout the year in both surveys, although fewer births are reported in February in the main survey and in May in the reinterview survey.

Age at first birth is recorded much less consistently than the year of first birth, with 85 percent of women discrepant, and 25 percent discrepant by more than five years. Again this is partly due to the poor reporting of the woman's own age, which carries over to the reporting of age at first birth. The distribution of age at first birth is quite similar in the two surveys (Figure 5.10) because the errors in the age at first birth appear to be fairly random (Figure 5.11). The index of dissimilarity for the two age distributions is only 10.1 percent.

Figure 5.10
Percent distribution of matched women by age at first birth in main survey and reinterview survey, Pakistan DHS, 1990-91



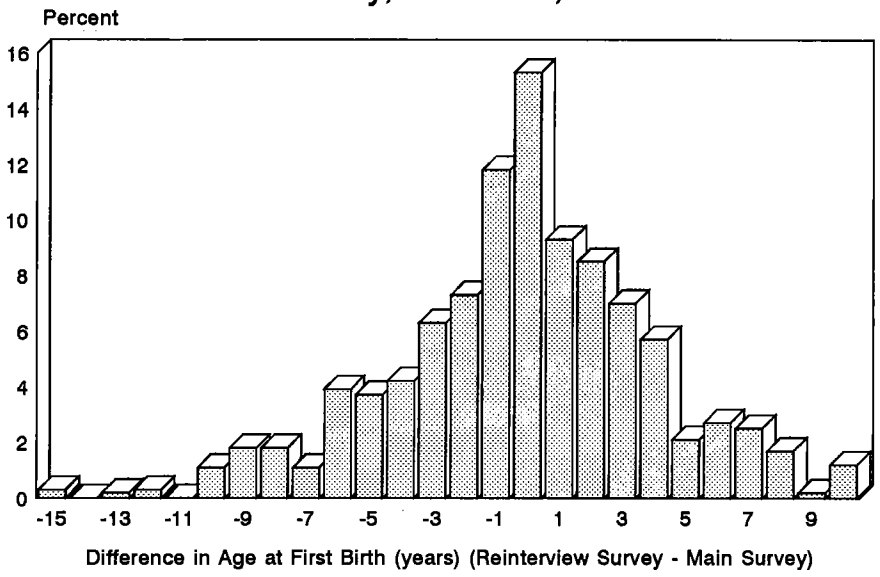
Note: Based on all women with at least one birth in both surveys (N=436).

A few women are discrepant by more than 10 years on their age at first birth. These probably represent mismatches but they may be women who report poorly on both their year of first birth and their own date of birth with the combined result being a highly discrepant age at first birth. The median age at first birth for matched women age 25-49 (as reported in the main survey) is slightly higher in the main survey (21.6) than in the reinterview survey (21.0). This is similar to the pattern observed for median age at first marriage, again suggesting that the reporting of the two is closely related, and both values are close to the value of 21.3 obtained in the full Pakistan DHS.

5.5 Age and Date of Most Recent Live Birth

The date of the most recent live birth was taken from the birth history in each survey. However, some adjustment needs to be made for women who have had a live birth since the main survey so that the date of

Figure 5.11
Percent distribution of matched women by difference in age at first birth between reinterview survey and main survey, Pakistan, 1990-91



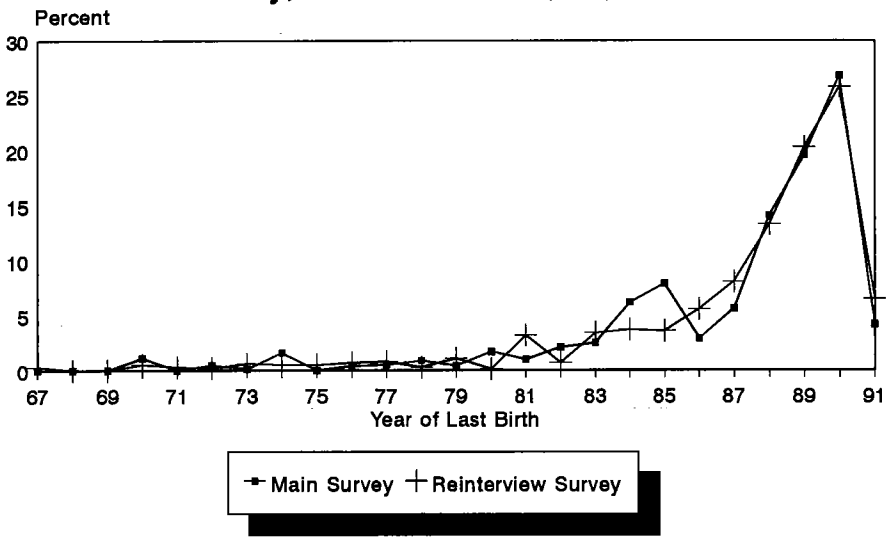
Note: Based on all women with at least one birth in both surveys (N=436).

the most recent live birth refers to the period before the main survey. This adjustment for new births is complicated by the very misreporting of dates of birth of children in the two surveys that are of interest. Some children who were actually born just before the main survey are reported in the reinterview survey as being born just after the main survey. Obviously these should not be counted as new births. In contrast, some births that actually occurred after the main survey are reported in the reinterview survey as having occurred before the main survey. These *should* be classified as new births. The decision as to whether to classify a birth as a genuine new birth or not was based on two variables: the reported interval between the birth of the child as recorded in the reinterview survey and the date of the main survey, and the recorded pregnancy status of the mother at the time of the main survey. All births reported after the main survey that occurred to women who reported that they were pregnant at the time of the main survey, or who were unsure whether they were pregnant, were classified as genuine new births. If the woman reported that she was pregnant at the time of the main survey, but in the reinterview reported that the most recent birth was in the three months prior to the main survey, the birth was also classified as a new birth. Finally, births that occurred at least five months after the main survey to women who reported that they were not pregnant at the time of the main survey were classified as new births, since it is likely that the pregnancy was not recognized at the time of the main survey. This approach will inevitably lead to some misclassification of new births, but this should be relatively insignificant, and given the inaccuracy of the reporting of dates of birth in Pakistan, these criteria represent a reasonable compromise.

Only women with at least one birth that is classified as having occurred before the main survey are included in this analysis (434 women). Of these, 432 reported the year of the most recent live birth (before the main survey) in both surveys, and just over half of them reported the same year in the two surveys. Only four percent of these women are discrepant by more than five years, and 80 percent report the year of last birth within one year in the two surveys. One of the reasons for the much higher consistency of reporting of year of last birth, compared to the other events studied in this report, can be seen in Figure 5.12, which shows

the distribution of the year of last birth in the two surveys. Both distributions are very heavily concentrated on the years immediately prior to the survey. Indeed, around 65 percent of women in both surveys report their last live birth since January 1988, which would be expected, given that contraception is not widely used. Consequently, the date of the most recent live birth is defined within a much narrower range for most women. For those women whose most recent live birth was around 1986, there is some evidence of shifting of the date of that birth to before 1986 in the main survey. However, this is not as pronounced as for first births because the majority of women have had another birth since this cut-off point.

Figure 5.12
Percent distribution of matched women by year of most recent birth in main survey and reinterview survey, Pakistan DHS, 1990-91

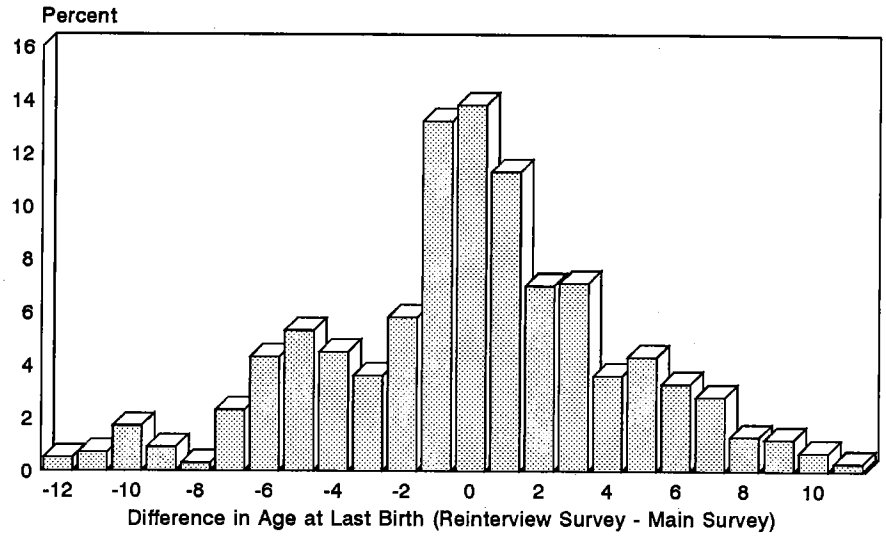


Note: Based on women with at least one birth in both surveys who gave the year of last birth in both surveys (N=432).
Births believed to have occurred since the main survey are excluded.

Despite the relatively good reporting of the year of the most recent birth, month of birth is still reported much less consistently. Of the 434 women included in this analysis, 380 reported the month of the most recent birth in both surveys, slightly more than for any of the other events examined here. However, 70 percent did not report the same month in the two surveys. Again, births are fairly evenly spread throughout the year in both surveys, although there is a slight tendency to report more births in the summer months in the reinterview survey.

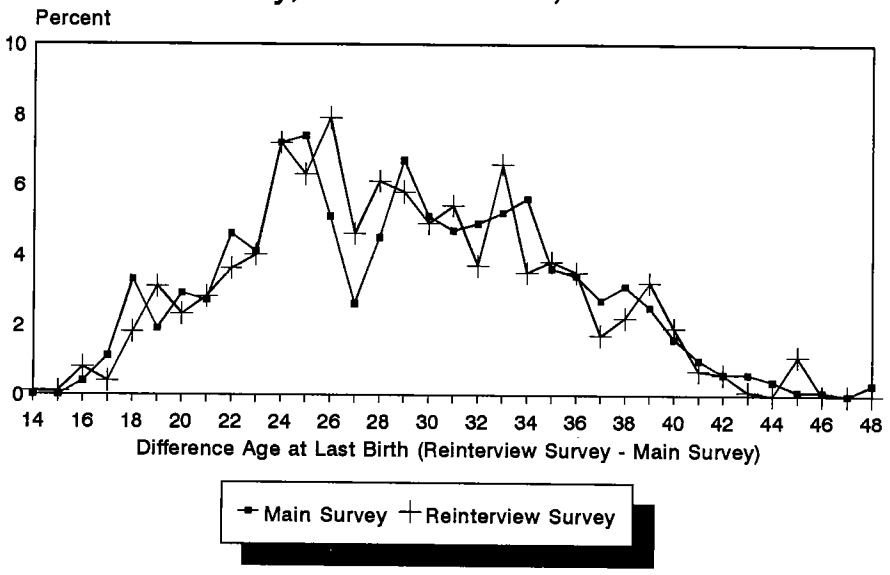
Age at most recent live birth is also reported very inconsistently, with 86 percent of women not providing the same age in the two surveys, and 30 percent discrepant by more than five years. As already discussed in Section 5.1, this is partly due to the poor reporting of the woman's age. Again the distribution of the difference between the age at most recent birth in the reinterview survey and in the main survey is fairly symmetrical (Figure 5.13), suggesting that there is no systematic difference between the two surveys. The distribution of age at most recent birth is very similar in the two surveys, as shown in Figure 5.14.

Figure 5.13
Percent distribution of matched women by difference in age at most recent birth between reinterview survey and main survey, Pakistan DHS, 1990-91



Note: Based on all women with at least one birth in both surveys (N=434).
Births estimated to have occurred since the main survey are excluded.

Figure 5.14
Percent distribution of matched women by age at most recent birth in main survey and reinterview survey, Pakistan DHS, 1990-91



Note: Based on all women with at least one birth in both surveys (N=434).
Births estimated to have occurred since the main survey are excluded.

5.6 Summary

The reliability of the reporting of ages and dates of four key demographic events has been examined in this section. In general, age and date reporting in the Pakistan DHS is very poor, as has been found in other surveys and censuses conducted in Pakistan. There are high levels of inconsistency for all the events examined, and the degree of discrepancy is also high, with many women discrepant by more than five years on age or year of each event. Some of these large discrepancies may be due to mismatches or data entry errors. However, it is difficult to be sure how many are mismatches and how many are genuine cases that report poorly, and eliminating these extreme cases from the analysis would not change the general picture painted by the results. The generally poor quality of date and age reporting is also reflected in the high degree of heaping present in the data, and in the very incomplete reporting of the month in which events occurred, although the year of the event is reported by the majority of women for all events. Although a high degree of discrepancy is observed in the reported ages and dates for individual women, the discrepancies for all events examined appear to be fairly random, and there is no evidence of systematic shifting of dates and ages between the surveys to any great extent. Consequently, the distributions of the age and year of events generally are very similar in the two surveys, as are summary figures such as the median age at first marriage and first birth. Hence it appears that inconsistencies in the reporting of ages and dates are primarily due to errors by the respondent that occur because she does not know the correct age or date, and therefore estimates it, rather than to any specific features of the survey. This is supported by the fact that similar problems have been found in other studies in Pakistan.

Chapter 6

Fertility Data

6.1 Analysis of Basic Indices of Discrepancy

One concern that came out of the analysis of the main survey in Pakistan was the displacement of births out of the most recent five-year period prior to the survey. Some evidence of this phenomenon has already been observed in the reporting of year of first birth and, to a much smaller extent, year of most recent birth. The detailed health sections of the questionnaire are only asked for children born in the most recent five-year period, since January 1986 in the case of Pakistan, and it is believed that some interviewers displace some births to an earlier date in order to reduce their workload. Further incentives for displacement arise from the fact that the anthropometric data are also only collected for surviving children born since January 1986. The collection of anthropometric data, in particular, can present additional problems for interviewers, since the measuring equipment has to be transported to each household at the appropriate time, which can be difficult. Such displacement of births has been a problem in a number of DHS surveys (Arnold, 1990) and there is strong evidence of this type of displacement in Pakistan, with 1038 births reported in 1986 compared to 2005 in 1985 (National Institute of Population Studies and IRD/Macro International Inc., 1992). In view of these concerns about displacement, a six-year period prior to the survey was used in the calculation of recent fertility rates in the final report, rather than the usual three-year period. Even with this adjustment, the estimate suggests that fertility has declined in the most recent period. However, there is some uncertainty about how much of this decline is genuine and how much is due to omission and displacement of births. Therefore, one of the key elements of this analysis is to investigate whether similar patterns of fertility are observed in the reinterview survey, in which the health questions and anthropometric measurements were omitted and better interviewers were used, and in particular, to try to evaluate how much of the fertility decline observed in the main survey is genuine.

Table 6.1 presents indices of discrepancy for the total number of children ever born, the total number of daughters ever born, the total number of sons ever born, the total number of children that died, and the total number of surviving children. The first three indices are crude measures of discrepancy, indicating the percentage of women who report a different number of children in the reinterview survey than in the main survey, the percentage for whom the difference between the number of children in the two surveys is more than one, and the percentage of discrepant women who report more children in the reinterview than in the main survey. In all cases the number of births reported at the

Table 6.1 Indices of disagreement for reporting of fertility variables, Pakistan DHS, 1990-91

Fertility variables	Percent discrepant	Percent discrepant by more than 1	Percent of the discrepant cases reporting more births in re-interview	Weighted kappa
Children ever born	25.2	9.5	66.0	0.96
Daughters ever born	16.8	4.0	69.3	0.94
Sons ever born	20.1	4.6	63.4	0.94
Dead children	21.7	6.7	70.7	0.71
Living children	6.8	2.8	47.1	0.98

reinterview survey is adjusted to the time of the main survey by discounting new births that occurred in the interval between the two surveys, as described in Section 5.5.

Overall, 25 percent of women did not report the same number of births in the two surveys, after adjustment for new births. This is higher than that reported in the World Fertility Survey reinterview studies. In Indonesia, 10 percent of women were discrepant on the number of live births they reported in the two surveys (Macdonald, Simpson and Whitfield, 1978), in Peru the figure was 12 percent (O'Muircheartaigh, 1984a), and in Lesotho it was 19 percent (O'Muircheartaigh, 1984b). The magnitude of the discrepancies is also quite large, with 10 percent of women discrepant by more than one birth. There is a strong pattern in the discrepancies in the number of births reported in the two surveys. Of the women who are discrepant on the number of children born, 66 percent report more births in the reinterview than in the main survey. This provides evidence that births were probably underreported in the main survey. However, there is no evidence of strong sex differentials in the consistency of reporting of births.

The number of dead children reported in the reinterview survey was also adjusted to the time of the main survey. This was done by classifying as a survivor any child whose date of death in the reinterview survey, as calculated from the reported date of birth and age at death, was after the main survey. Inevitably, this approach will misclassify some deaths due to misreporting of date of birth and/or age at death, but matching within the birth histories to check the survival status recorded in the main survey is not feasible given the degree of omission and misreporting of dates of births that appear to be present. The analysis was also done without reclassifying deaths of children born before the main survey who appeared to have died after the main survey; doing so made very little difference in the results. The differential in consistency of reporting by survival status of the child is very pronounced: 22 percent of women are discrepant on the number of dead children they report in the two surveys compared to only 7 percent on the number of surviving children they report. In addition, 71 percent of women who are discrepant on the reported number of deaths report more deaths in the reinterview than in the main survey, whereas only 47 percent of women who are discrepant on the reported number of surviving children report a higher number in the reinterview survey. These findings suggest that dead children are underreported in the main survey, whereas the number of surviving children is reported reasonably consistently, and the inconsistencies that do occur are fairly random in direction.

Overall, weighted kappa for the total number of births reported by women is 0.96, which demonstrates a high degree of reliability in the reporting of the number of births in the two surveys (Table 6.1). However, it would be very worrying if this were not the case. This value of weighted kappa is slightly below that reported by MacDonald, Simpson and Whitfield (1978) in Indonesia (0.98), and by O'Muircheartaigh (1984a) in Peru (0.98), although it is higher than reported by O'Muircheartaigh (1984b) in Lesotho (0.92). The number of daughters born and the number of sons born are equally reliably reported in the Pakistan DHS, with weighted kappa equal to 0.94 in each case. Not surprisingly, given the earlier discussion, the number of children that died is the least reliably reported figure in the table, and the number of surviving children is the most reliably reported.

6.2 Reporting of Births by Year of Birth

The above analysis suggests that the number of births reported by women is higher in the reinterview survey than in the main survey, and this primarily appears to be due to the reporting of more children that died in the reinterview survey. Whether there is any pattern in the reporting of births by period of birth is investigated in the next stage of the analysis. Figures 6.1 and 6.2 present the number of births in the reinterviewed sample by reported year of birth and survival status, for each survey. Children who were reported to have died at the time of the reinterview survey had their survival status changed to alive if their

Figure 6.1
Number of living children by calendar year of birth reported in main survey and reinterview survey by the matched sample, Pakistan DHS, 1990-91

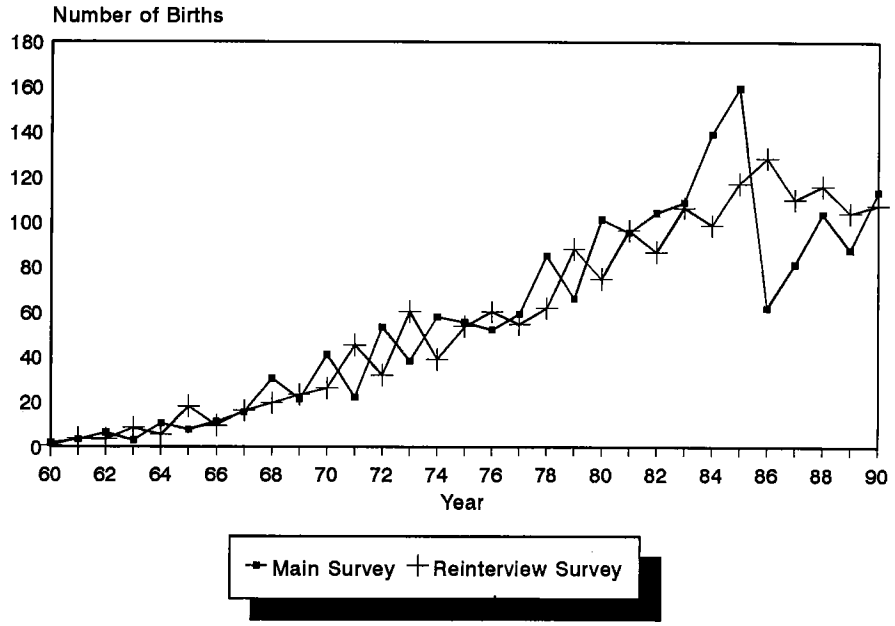
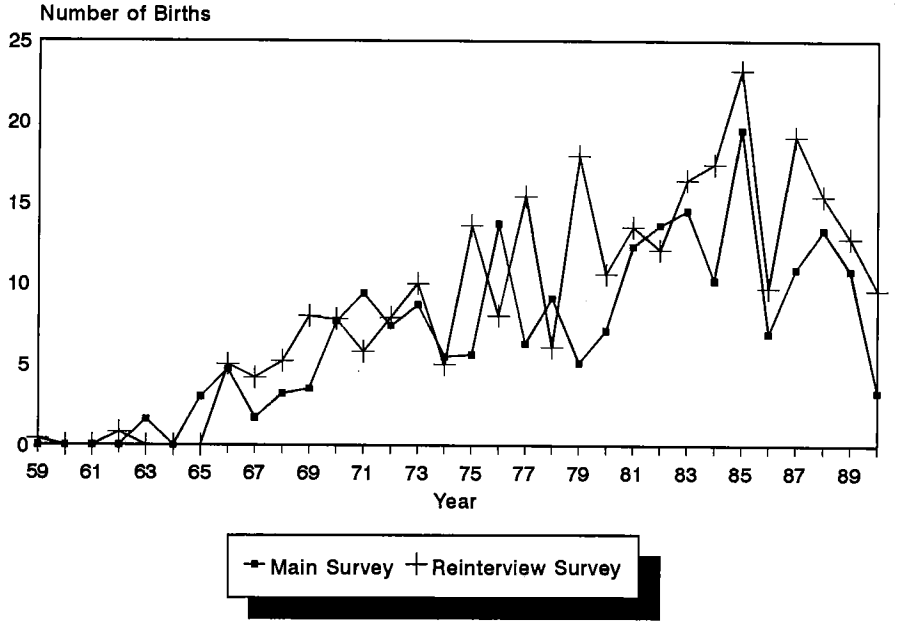


Figure 6.2
Number of dead children by calendar year of birth reported in main survey and reinterview survey by the matched sample, Pakistan DHS, 1990-91



reported age at death indicated that they died after the date of the main survey. Births in 1991 were excluded because the main survey only covered the very beginning of the year.

For living children, the figure shows that the differences in the reporting of births in the two surveys are most pronounced in the period immediately prior to the main survey. In the main survey there is a dramatic peak in the number of births reported in 1984 and 1985 followed by an equally dramatic trough in 1986. No such pattern is observed in the reinterview survey, providing evidence to support the argument that this particular feature of the main survey data is an artifact of displacement of births by interviewers, probably to avoid the health and anthropometric sections of the questionnaire for the displaced births. The number of births reported remains higher in the reinterview survey from 1986 to 1989, but by 1990 the two surveys are again in reasonable agreement. A total of 1774 surviving births were reported from 1960 to 1990 in the reinterview survey, compared to 1797 in the main survey for the same period, a difference of only one percent. From 1986 to 1989, however, there were 459 births reported in the reinterview survey, compared to 335 in the main survey, i.e., there were 27 percent fewer births reported in this period in the main survey than in the reinterview survey. Hence, it appears that there was substantial underreporting of surviving children in the period since January 1986 but the majority of these births were displaced to an earlier period rather than being omitted entirely.

For dead children, the picture is quite different. The pattern of births by year is somewhat erratic due to the small numbers involved and considerable heaping in several years. However, in general it is clear that there are fewer non-surviving births reported in the main survey in almost every year. Given the small numbers and the severe heaping, it is difficult to draw any firm conclusions regarding the pattern of underreporting of deaths in the main survey compared to the reinterview survey by time period, but there does not appear to be any strong tendency for the degree of underreporting to increase further back in time. Indeed, there appears to be relatively more underreporting in the period since 1986. It is also interesting to note that the pattern of a peak in births in 1985 followed by a trough in 1986 is also seen for dead children, but in *both* surveys. It is not clear why dead children appear to have been displaced to before 1985 in the reinterview survey, whereas surviving children were not, but it could be associated with the general poorer reporting of dead children than surviving ones and, in particular, with heaping of the year of birth of children that died or of the number of years since the birth of a child that died.

Excluding births in 1991, a total of 219 children who died were reported in the main survey compared to 281 in the reinterview survey, i.e., 28 percent more deaths were reported in the reinterview survey than in the main survey. For the period from 1986 to 1990 there were 49 percent more deaths reported in the reinterview survey than in the main survey, which suggests slightly more displacement of births in the main survey than in the reinterview survey or slightly more omission of children who died in the most recent period. Yet another explanation could be that some deaths of children born close to the main survey, but who died after the main survey, may be misplaced to before the main survey in the reinterview. Such misclassifications of the survival status of these children at the time of the main survey would lead to an excess of deaths in the reinterview survey, and a corresponding deficit of surviving children. There is some evidence for this in that there are slightly fewer surviving births reported overall in the reinterview survey. In addition, Figure 6.2 shows that the differential between the two surveys is particularly marked for children born in 1990. Only 3 deaths of children born in 1990 are reported in the main survey, compared to 10 in the reinterview survey, but sample sizes are too small to draw firm conclusions on this.

In summary, it appears that fewer deaths were reported fairly consistently over the whole birth history in the main survey, which probably indicates underreporting of deaths in the main survey. Of course, the reinterview survey is not necessarily correct. It is possible, although unlikely, that deaths were overreported in the reinterview survey. It is also possible that additional deaths were omitted in both surveys, particularly further back in time. Such omissions will not be detected by comparisons between the two surveys.

The reason why the reinterview survey was able to identify more deaths than the main survey is unclear, but there are several possible explanations. First, the reinterview survey employed better interviewers in general, who may have been better at obtaining information on dead children. In addition, the fact that the reinterview survey covered a much smaller sample than the original Pakistan DHS and used a much shorter questionnaire might have given the interviewers more time to probe for omitted deaths. Another factor is that the health section in the Pakistan DHS included a large number of questions on dead children. It is known that interviewers in the main survey felt uncomfortable discussing dead children in so much detail with the mothers, and hence were reluctant to probe for deaths knowing that they would then have to ask the additional questions. However, this would not explain the fact that deaths were omitted throughout the birth history, and not just in the period covered by the health section.

Another notable feature of Figure 6.1 and, to a lesser extent, Figure 6.2 is the sawtooth pattern that is observed in both surveys for most of the period from around 1970 to 1983. This is due to a common pattern of heaping on ages ending in even numbers that has been observed in surveys and censuses in Pakistan. For example, Robert Retherford and his colleagues examined five recent national surveys and censuses in Pakistan and found severe heaping on ages 8, 10, and 12, corresponding to births in the 9th, 11th, and 13th years before the survey (Retherford and Alam, 1985; Retherford and Mirza, 1982; Retherford et al., 1987). Heaping on even numbered ages is evident throughout the age distribution in the Pakistan DHS (National Institute of Population Studies and IRD/Macro International Inc., 1992), but it is most prominent at ages 8, 10, and 12. The fact that the pattern in the reinterview survey is almost the mirror-image of that observed in the main survey is due to the fact that the reinterview survey took place almost a year after the main survey.

In the same surveys and censuses, Retherford and Alam (1985) also found a pattern of age exaggeration due to the upward rounding of children's ages. This pattern consistently led to implausibly low fertility estimates in the years immediately preceding the surveys. To the extent that the same pattern prevails in the Pakistan DHS, some of the distortion in the ages of young children may be due to inaccurate age reporting on the part of respondents rather than to deliberate age displacement on the part of interviewers. In addition, in Figure 6.1 the number of surviving births rises steadily in both surveys from 1960 to the mid-1980s and then declines slightly. The steady increase in the number of births is due to the effect of truncation and, under a constant fertility regime, this increase would be expected to continue until the year of the survey. The deviation from this expected pattern seen in both the main survey and the reinterview survey seems a little too large to be due entirely to a recent fertility decline, and may suggest some evidence of omission of recent births in both surveys. However, it is also likely that exaggeration of the ages of children, as described above, combined with some degree of recent fertility decline, also contribute to the decline in the number of births in the period immediately before the survey.

6.3 Impact on Summary Measures of Recent Fertility

The obvious question that arises from the analysis of Figures 6.1 and 6.2 is, "What impact does the omission and displacement of recent births in the main survey have on the estimated Total Fertility Rate (TFR) for the period before the main survey?" Although this is a very important question that needs to be investigated, some caution should be exercised in generalizing the findings from the reinterview sample to the full Pakistan DHS sample. First, the sample size in the reinterview is very small for the calculation of fertility rates. While sampling error will not affect the comparison of the fertility rates calculated for the reinterview sample from the main survey and the reinterview survey, it will affect the generalization to the full sample. In addition, it has already been noted in Chapter 2 that there are some compositional differences between the reinterview sample and the full Pakistan DHS sample, particularly with respect to the distribution

by province and urban/rural residence. Hence, the analysis in this section should be viewed as exploratory, and the findings viewed as suggestive of the possible impact of birth displacement on fertility rates.

Table 6.2 presents TFRs for the period 0-5 years before the main survey as estimated from the main survey and the reinterview survey for the reinterviewed sample. Since women who have never been married are not included in the Pakistan DHS, in the calculation of the TFR it is necessary to assume that never-married women have not had any births. The TFR is consequently calculated by multiplying the denominator of the age-specific fertility rates (which is the exposure of ever-married women) by the inverse of the proportion of all women at each age who have ever been married. The calculation of these ever-married sample inflation factors is based on the proportion of women who have ever been married by single years of age as reported on the household questionnaire in the Pakistan DHS survey. The age on the household questionnaire is used rather than the age in the individual survey because the ages of never-married women are only available from the household questionnaire. If there is differential reliability of reporting in the household and individual questionnaires, using the age of ever-married women as given on the individual questionnaire but the age of never-married women from the household questionnaire to calculate the ever-married sample inflation factors could bias the results. Consequently, ever-married women maintain the inflation factor associated with the age reported on the household questionnaire no matter what age they report in the individual questionnaire. Similarly, women in the reinterview survey maintain the inflation factor associated with the age reported on the original household questionnaire no matter what age they report in the reinterview survey. In addition, the calculation of these factors for use with the reinterview sample is based on the household questionnaires from the clusters selected for the reinterview only, since it is possible that marriage patterns, and hence the inflation factors, for these clusters differ slightly from the full Pakistan DHS sample.

Table 6.2 Age-specific fertility rates and total fertility rates for the six years preceding the main survey and mean number of children ever born to women age 40-49 for main survey and reinterview survey, and "standardized" fertility estimates for the reinterview survey, Pakistan DHS, 1990-91

		Reinterview		
Mother's age	Main survey	Reinterview survey	Standard-ized for age	Standard-ized for birth history
15-19	88.1	77.7	86.3	83.9
20-24	259.5	244.2	287.8	205.7
25-29	285.4	325.3	341.3	282.5
30-34	268.7	253.6	290.8	214.5
35-39	144.3	195.0	227.1	159.9
40-44	[61.3]	[66.0]	[66.2]	[53.2]
45-49	[42.6]	[72.4]	[53.9]	[37.7]
TFR 15-49	5.7	6.2	6.8	5.2
TFR 15-44	5.5	5.8	6.5	5.0
Mean CEB (40-49)	6.2	6.4	6.3	6.2

Note: Rates are calculated for all women 15-49, using information on women's marital status by age from the household questionnaire, and information on the number of births from the individual questionnaire in the main survey and in the reinterview survey. Bracketed figures are partially truncated rates. Age-specific rates are per 1000 women.
TFR: Total fertility rate expressed per woman
CEB: Children ever born

Differences in the TFRs calculated in the two surveys reflect misreporting of the woman's date of birth as well as differences in the birth histories. The exposures in the denominator of the age-specific fertility rates can be considered as the sum of the exposures of individual ever-married women in each age group, weighted by the ever-married sample inflation factor associated with each woman, which in turn is based on her reported age in the household questionnaire. The amount of exposure individual women

contribute to an age group will differ between the main survey and the reinterview survey due to misreporting of the woman's date of birth. This means that the total impact of the ever-married inflation factors on the exposure in a particular age group will also vary between the surveys. In addition, the age of a woman at the time of even an accurately reported birth will vary due to misreporting of her current age. Hence, differences in the TFR between the main survey and the reinterview survey will be due to the combined effect of differences in the reporting of the number and dates of births and differences in the assignment of births and exposure to age groups due to misreporting of the woman's age. In a sense this problem is somewhat akin to that of standardization of mortality rates. One way to "standardize" the TFR is to use the age of the woman reported in the main survey in the calculation of the TFR in both the main survey and the reinterview survey. Any differences in the TFR are then due only to differences in the birth histories. Differences in the TFR that can be attributed solely to the combined effects of age misreporting can also be estimated by calculating the TFR for each survey based on the woman's age reported in the survey but on the birth history from the main survey. Both "standardized" and "unstandardized" TFRs are presented for the reinterview survey.

Births reported in the reinterview survey were included only if they were reported as having occurred before the date of the main survey. This approach means that some genuine new births, which are misreported as having occurred before the main survey, will be included in the numerators of the age-specific fertility rates. However, the number of cases involved is very small and will be compensated for largely by displacements in the opposite direction. The estimated TFR for the matched sample based on the information in the main survey is 5.7 compared to 6.2 based on the information from the reinterview survey[2]. Hence, the TFR estimated in the main survey is 8.1 percent lower than in the reinterview survey. However, if the TFR estimated from the reinterview survey is "standardized" using the ages reported in the main survey, the TFR increases to 6.8. This suggests that, *if* the birth histories in the reinterview are a reasonable representation of the true situation, underreporting of births in the main survey alone results in an underestimate of the TFR for the period 0-5 years before the survey of approximately 16.2 percent. The impact of age misreporting alone can be seen by comparing columns (2) and (5) of Table 6.2. The TFR is reduced from 5.7 in the main survey to 5.2 in the reinterview survey due to differences in the reporting of the ages of the women. Hence, the effect on the TFR of differences in the birth histories in the two surveys is offset by the effect of differences in age reporting by women.

This finding has important implications for the estimates based on the full Pakistan DHS survey. If the national estimate of the TFR of 5.4, based on the full Pakistan DHS sample, is underestimated by a similar amount, i.e., 8.1%, the TFR for the period 0-5 years before the survey would be around 5.9 births per woman. This is still lower than any of the TFRs estimated from earlier studies, which would suggest that there has been some degree of recent fertility decline in Pakistan, but that the magnitude of the decline is lower than originally suggested by the Pakistan DHS. However, this estimate of the TFR is very sensitive to misreporting of women's ages, partly because of the need to use ever-married sample inflation factors. For example, if the "age-standardized" TFR estimated for the reinterview survey using the ages reported in the main survey is taken as a representation of the degree of underreporting of the TFR in the full Pakistan DHS, the estimated TFR for the full sample would increase to 6.4.

The last row of Table 6.2 shows the mean number of children ever born to women age 40-49 from the two surveys. Again the figure is higher in the reinterview survey than in the main survey, although the difference is less than that observed for the TFR. It also appears that at least part of the difference observed can be attributed to differences in age reporting by women in the two surveys, which results in the calculation being based on a slightly different group of women.

---

[2] Note that the estimate of the TFR for the reinterview sample based on the main survey (5.7) is higher than the estimate of 5.4 obtained for the full Pakistan DHS sample. This reflects the impact of sampling error.

Another issue that affects the comparison of the TFR estimated for the reinterview sample in each survey is the period chosen for the calculation. The analysis presented in Section 6.2 showed that the discrepancies in the birth histories for surviving children appear to be due mainly to displacement of births, particularly in the period since 1983. If the period chosen for the calculation of the TFR excludes some of the births that were displaced in the main survey, the difference in the estimated TFR will be exaggerated since birth displacement is less common in the reinterview survey. This is illustrated by Table 6.3, which presents age-specific fertility rates estimated from each survey for the reinterview sample for three four-year periods before the main survey, together with the TFR calculated for women age 15-39.

For the period 0-3 years before the survey, the TFR estimated from the reinterview survey for women age 15-39 is higher than that estimated from the main survey, but for earlier periods the TFR estimated from the main survey is higher than that estimated from the reinterview survey. The use of a six-year period before the survey for the calculation of the TFR in the final report of the Pakistan DHS was aimed at reducing the bias in the TFR due to displacement of births, but it is clear from Figure 6.1 that some births in the main survey were displaced back as far as the seventh year before the survey. Hence, the use of a six-year window still excludes some of the displaced births and may exaggerate the difference between the estimated TFR from the main survey and the reinterview survey. When an eight-year period before the survey is used to ensure that all the displaced births are included, the TFR for the reinterviewed sample for women age 15-44 is 6.1 estimated from the main survey and 6.0 estimated from the reinterview survey. For women age 15-49, the TFR estimated from both surveys for the eight-year period is 6.3, although these estimates are somewhat affected by truncation. The corresponding estimates for the full Pakistan DHS sample are 5.7 for women age 15-44 and 5.9 for women age 15-49. These findings suggest that the two surveys actually produce quite similar estimates of the TFR once all the births displaced to before 1986 are included in the calculation. However, note that these results represent the combined and compensatory effects of differences in the birth histories and differences in age reporting between the two surveys, and that differences in the birth history alone result in higher estimates of the TFR in the reinterview survey.

Table 6.3 Age-specific fertility rates for women age 15-49 and total fertility rates for women age 15-39 for four-year periods before the main survey, main survey and reinterview survey, Pakistan DHS, 1990-91

	Period before main survey		
Mother's age	0-3 years	4-7 years	8-11 years
Main survey			
15-19	75	159	111
20-24	247	311	272
25-29	257	373	390
30-34	247	322	322
35-39	136	197	[238]
40-44	53	[104]	-
45-49	[42]	-	-
TFR 15-39	4.8	6.8	6.7
Reinterview survey			
15-19	80	95	95
20-24	228	289	309
25-29	291	401	308
30-34	253	277	273
35-39	183	171	[282]
40-44	49	[99]	-
45-49	[73]	-	-
TFR 15-39	5.2	6.2	6.3

Note: Rates are calculated for all women 15-49, using information on women's marital status by age from the household questionnaire, and information on the number of births from the individual questionnaire in the main survey and in the reinterview survey. Bracketed figures are partially truncated rates and most are based on less than 125 woman-years of exposure. Age-specific rates are per 1000 women.
TFR: Total fertility rate expressed per woman
- No information

6.4 Impact on Infant and Child Mortality Rates

The analyses in Sections 6.1 and 6.2 suggest that the higher number of births reported in the reinterview survey is due almost entirely to the reporting of more children that died. Obviously this is likely to have an impact on the infant and child mortality rates in the two surveys. Table 6.4 presents the neonatal, postneonatal, infant, child and under-five mortality rates estimated for the reinterviewed sample from the main survey and the reinterview survey, for the periods 0-9 and 0-24 years before the main survey. The 25-year period is much wider than would normally be used for the estimation of mortality rates, but virtually all children that died should be included in this window, so any differences between the two surveys will primarily reflect omissions of children that died rather than differential placement of births in the two surveys. Again, it is important to interpret this analysis cautiously, especially when generalizing from the results because the sample size is very small, so all estimates have large sampling errors. Although this will not affect the comparisons between the two surveys for the reinterview sample, it will affect the generalizability of the findings to the full sample. This is illustrated by the fact that the infant mortality rate for the period 0-9 years before the survey estimated from the main survey for the reinterview sample is considerably lower than that estimated for the full Pakistan DHS sample (80.5 per 1000 compared to 94.0 per 1000). This difference might be due to sampling error, but it might also indicate relatively poorer coverage of deaths in the reinterview sample than in the full Pakistan DHS sample. In addition, there is no guarantee that the figures from the reinterview survey are more accurate than those obtained in the main survey.

Table 6.4 Neonatal, postneonatal, infant, and childhood mortality for 0-9 and 0-24 years before the main survey for reinterviewed women, estimated from the main survey and the reinterview survey, Pakistan DHS, 1990-91

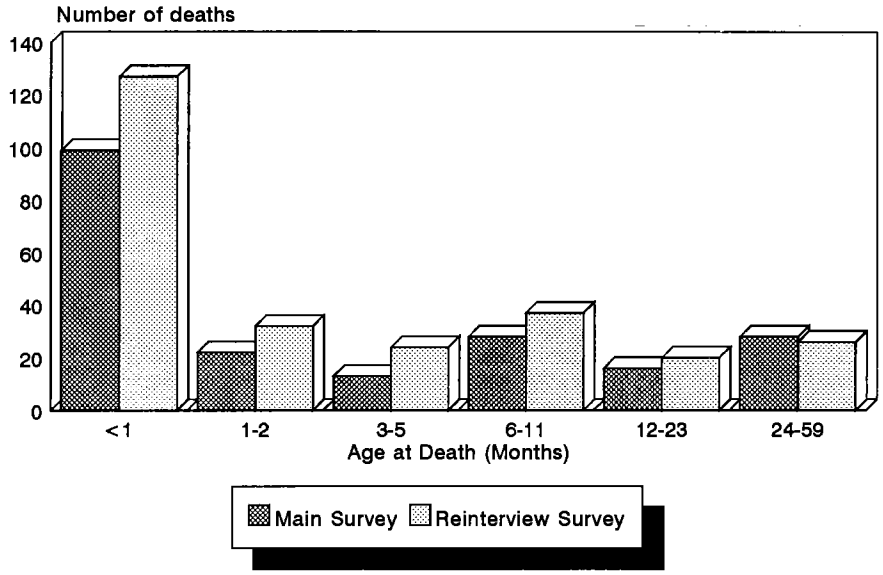
Period	Neonatal mortality (NN)	Post-neonatal mortality (PNN)	Infant mortality ($_1q_0$)	Childhood mortality ($_4q_0$)	Under-5 mortality ($_5q_0$)
Main survey					
0-9 years	48.5	32.0	80.5	25.4	103.8
0-24 years	50.0	32.6	82.6	27.9	108.2
Reinterview survey					
0-9 years	61.0	46.3	107.3	22.4	127.4
0-24 years	62.7	46.9	109.6	29.3	135.7

With the exception of child mortality 0-9 years before the main survey, the mortality rates are consistently higher in the reinterview survey than in the main survey. For example, the infant mortality rate estimated from the main survey is 80.5 per 1000 for the period 0-9 years before the survey, compared to 107.3 per 1000 in the reinterview survey, and the under-five mortality rate for the same period is 103.8 per 1000 in the main survey compared to 127.4 per 1000 in the reinterview survey. *If* the mortality rates estimated from the reinterview survey are a better representation of the child mortality situation for the reinterviewed sample, and *if* the results hold for the full sample, the infant mortality rate estimated for the Pakistan DHS sample for the period 0-9 years before the survey would be increased from 94.0 per 1000 to 125.3 per 1000, and the under-five mortality rate for the same period would be increased from 120.4 per 1000 to 147.8 per 1000.

It is interesting to note that the absolute differential in the mortality rates in the two surveys is similar for both neonatal and postneonatal mortality, so it does not appear that the reinterview survey is picking up additional early neonatal deaths only. Rather, it is picking up more deaths throughout infancy. However, the differential for child mortality is much lower. Indeed, for the period 0-9 years before the survey, the child mortality rate estimated from the main survey is slightly higher than that estimated from the reinterview

survey, although the reverse is true for the full 0-24 year period. This is also seen in Figure 6.3, which presents the number of child deaths by age in each survey for the period 0-24 years before the main survey. More deaths are observed in the reinterview survey in every age group except 24-59 months, where the number of deaths is very similar in the two surveys. In absolute terms the number of additional deaths obtained in the reinterview survey is highest for the first month of life, where mortality is highest. However, in relative terms the differential is highest for the 3-5 month age group in which 84 percent more deaths are observed in the reinterview survey than in the main survey, although part of this may be due to discrepancies in the reported age at death of children between the two surveys.

Figure 6.3
Number of child deaths 0-24 years before main survey, by age at death, reported in main survey and reinterview survey, Pakistan DHS, 1990-91



Note: Based on matched sample.

Another important point is that the differential between the two surveys in neonatal and postneonatal mortality is similar for the 0-9 and 0-24 year periods before the survey. This would seem to support the suggestion from Figure 6.2 that the reinterview survey picks up additional deaths throughout the birth history, not just deaths further back in time. Hence, it would appear that *if* the differentials in the reporting of infant deaths in the two surveys are due to omissions of deaths in the main survey then these omissions are not concentrated among deaths in the more distant past. One consequence of this finding, together with the finding in the previous paragraph that omissions of deaths do not appear to be concentrated among early neonatal deaths, is that standard data quality tests aimed at picking up such omissions will not be successful in identifying the omissions that seem to have occurred in Pakistan. As mentioned earlier, it is very important to remember that the data from the reinterview survey are not necessarily correct. Underreporting of early neonatal deaths or deaths in the distant past could still be present, but to a similar degree as in the main survey. Comparisons of the two surveys will not pick this up. It is also possible, although less likely, that deaths might have been overreported in the reinterview survey for some reason or that the degree of underreporting in the main survey was higher in the clusters selected for the reinterview than in the full sample. However, there does appear to be strong evidence that deaths were omitted to some extent in the

main Pakistan DHS survey, and consequently the mortality rates in Pakistan are likely to be higher than suggested by the Pakistan DHS, although exactly how much higher is uncertain.

6.5 Characteristics Associated with Poor Reporting

One final question that needs to be addressed is whether the degree of discrepancy in the number of births reported in the two surveys varies according to the characteristics of the woman. An initial problem encountered is how to define the characteristics of the woman given that they often differ in the two surveys. In this analysis, the characteristics reported in the main survey are used.

Table 6.5 presents the percentage of women who report the same number of children in each survey, the percentage who report more children in the main survey than in the reinterview, and the percentage who report more children in the reinterview survey than in the main survey, by urban/rural residence, whether the woman had ever attended school, literacy, province, age, and parity. In all cases adjustment is made for genuine new births that occurred between the two surveys.

Generally, the patterns shown in the tables are as expected. Older women and less educated women are less likely to report their number of births consistently than younger women and more educated women. The percentage of women reporting more births in the reinterview survey sharply increases after age 35, which suggests that births are more likely to be underreported in the main survey by older women than by younger women.

Table 6.5 Percent distribution of women by discrepancy in the number of children ever born, according to background characteristics from the main survey, Pakistan DHS, 1990-91

Background characteristic	Not discrepant	More births in main survey	More births in re-interview	Total	Number
Residence					
Urban	73.3	12.1	14.6	100.0	189
Rural	75.8	6.2	17.9	100.0	285
Education					
Attended school	81.4	4.5	14.1	100.0	118
Did not attend school	72.7	9.9	17.4	100.0	356
Literacy					
Reads easily	86.5	3.2	10.3	100.0	103
Does not read easily	71.6	10.1	18.3	100.0	372
Province					
Punjab	72.2	10.3	17.5	100.0	282
Sindh	79.5	6.2	14.4	100.0	129
NWFP	80.5	6.2	13.4	100.0	57
Balochistan	-	-	-	-	-
Age					
15-19	100.0	0.0	0.0	100.0	22
20-24	83.9	6.2	9.9	100.0	70
25-29	82.0	5.6	12.4	100.0	106
30-34	85.5	5.4	9.0	100.0	82
35-39	59.0	9.2	31.8	100.0	82
40-44	64.0	15.3	20.7	100.0	68
45-49	57.2	17.9	24.9	100.0	45
Parity					
0	96.2	0.0	3.8	100.0	39
1-2	83.4	6.4	10.1	100.0	103
3-4	76.7	4.8	18.5	100.0	131
5-6	68.7	10.4	20.9	100.0	94
7-8	55.9	18.2	25.9	100.0	64
9+	70.8	14.7	14.5	100.0	43
Number	355	41	79	474	474

- Too few cases

As expected, women who had attended school were more likely to report consistently in the two surveys. This effect is also seen for literacy; 87 percent of women who read easily report the same number of births in the two surveys, compared to 72 percent of women who do not. However, discrepant women are more likely to report more births in the reinterview survey, regardless of educational or literacy status.

Overall, there is little difference between urban and rural women in the percentage reporting consistently in the two surveys. However, the pattern of differential reporting does vary by area of residence: rural women are more likely than urban women to report more births in the reinterview survey than in the main survey. Provincial differentials in the consistency of reporting of births in the two surveys are relatively small. Women in Punjab report the number of births they have had slightly less consistently than women in Sindh and NWFP. However, discrepant women are more likely to report more births in the reinterview survey in all regions. Discrepancies in the other direction are slightly more common in Punjab than elsewhere, but the difference is not large. Finally, the percentage of women reporting the number of births consistently in the two surveys declines with increasing parity in the main survey, reaching a minimum at parity 7-8.

Chapter 7

Contraceptive Knowledge and Use

The reinterview questionnaire included a shortened form of the section on knowledge and use of contraception from the main survey. In particular, the battery of questions on knowledge and use of individual methods of contraception was used. In this section the consistency of responses to these questions is examined, together with information on ever use of contraception, ever use of modern and traditional methods, first method used, and number of living children at first use.

In both the main survey and the reinterview survey, women were asked what contraceptive methods they had heard of. The responses for knowledge of any method that the women themselves named were coded as "yes, spontaneous." The interviewer then proceeded to describe all the methods listed that were not mentioned by the woman. If the woman said she had heard of the method when it was described to her, the response was coded as "yes, probed." If the woman did not recognize the method, the response was coded as "no." Women were then asked if they had ever used any of the methods that they reported knowledge of, either spontaneously or after probing. Women can then be classified according to whether they have ever used any method of contraception, any traditional method, and any modern method.

Table 7.1 presents the percentage of women discrepant on their response to knowledge of each method of contraception, knowledge of any method, knowledge of any modern method, and knowledge of any traditional method. For the purposes of this analysis the categories of "yes, spontaneous" and "yes, probed" for knowledge of method are combined into a single "yes" category. In addition, the very act of conducting the main survey could affect the responses in the reinterview survey, since women may remember the method after hearing it described in the main survey. Even if the woman does not remember the method from the first survey, she may have heard of it subsequently from another source. Consequently, shifts in response from "no" in the main survey to "yes" in the reinterview survey may represent genuine changes in the knowledge of the woman with respect to a particular method. However, shifts from "yes" in the main survey to "no" in the reinterview survey do represent inconsistencies. Data on the percentage of women that report a "yes" response in the main survey but a "no" response in the reinterview survey are presented in Table 7.1. The percentage of women reporting that they know each method is also reported for each survey, as is kappa.

For knowledge of individual methods of contraception, the percentage discrepant ranges from 6 percent for other methods to 34 percent for injections and the IUD. However, the percentage that shift from having heard of a method in the main survey to not having heard of it in the reinterview survey ranges from only 5 percent for the pill to 12 percent for withdrawal and the IUD. This suggests that the majority of discrepancies occur because women's knowledge of contraceptive methods has increased between the two surveys, even if this is because they remember hearing about the method in the first survey. This is supported by the fact that reported knowledge was higher in the reinterview survey for every method except withdrawal. Some of the most highly discrepant methods are the most striking in this respect. In total, 34 percent of women were discrepant on knowledge of injections, yet only 8 percent shifted from having heard of them in the main survey to not having heard of them in the reinterview survey, and the majority of these shifted from the "yes, probed" category to the "no" category. For female sterilization, 28 percent of women were discrepant on knowledge of the method, but only 7 percent reported no knowledge of the method in the reinterview survey but knowledge in the main survey, and again the majority of these were shifts from "yes, probed" to "no." Knowledge of the pill seems to be much higher in the reinterview survey than in the main

Table 7.1 Indices of disagreement for knowledge of contraceptive methods, and level of knowledge of specific methods, main survey and reinterview survey, Pakistan DHS, 1990-91

Method	Percent discrepant	Percent moving from "yes" to "no" response	Percent knowing method in main survey	Percent knowing method in re-interview survey	Kappa	Number
Pill	33.1	5.0	65.4	88.4	0.14	474
IUD	33.7	11.6	59.9	70.3	0.27	472
Injection	33.8	7.6	62.8	81.6	0.19	474
Diaphragm/Foam/Jelly	31.4	8.6	15.7	29.9	0.13	474
Condom	29.8	8.5	42.0	54.8	0.41	474
Female sterilization	27.5	7.0	75.9	89.4	0.07	474
Male sterilization	27.2	8.2	22.7	33.6	0.34	474
Periodic abstinence	26.1	7.0	19.1	31.3	0.32	474
Withdrawal	19.0	12.1	23.5	18.3	0.43	473
Other	6.4	2.9	2.9	3.5	-0.03	466
Any method	17.2	3.6	84.5	94.6	0.10	474
Any modern method	17.4	3.4	83.7	94.4	0.13	474
Any traditional method	27.5	11.0	31.0	36.5	0.39	474

survey, with 65 percent of the matched women reporting that they had heard of it in the main survey, compared to 88 percent in the reinterview survey.

Information on knowledge of any method of contraception, and knowledge of any modern and any traditional method, is obtained by combining the responses for individual methods. Overall, 17 percent of women were discrepant on knowledge of any method of contraception between the two surveys, however only 4 percent reported that they did not know any methods in the reinterview survey but that they did know a method in the main survey. Therefore, overall knowledge of any method increased from 85 percent in the main survey to 95 percent in the reinterview survey. A very similar picture is seen for knowledge of any modern method. However, knowledge of any traditional method is less consistently reported, with 28 percent of women inconsistent and 11 percent shifting from knowledge in the main survey to no knowledge in the reinterview survey. This is primarily due to women shifting from having heard of withdrawal in the main survey after probing, to not having heard of it in the reinterview survey. It is also worth noting that knowledge of contraceptive methods in the main survey is generally higher in the reinterview sample than in the full Pakistan DHS sample. For example, in the full Pakistan DHS sample 77 percent of women reported knowing a modern method of contraception compared to 84 percent in the reinterview sample.

Table 7.2 presents the percentage of women who are discrepant on ever use of each method of contraception, and on ever use of any method, ever use of a modern method, and ever use of a traditional method. For individual methods, this analysis is restricted to women who reported that they had heard of the particular method in both surveys, since women who report that they have never heard of the method are not asked if they have ever used it. For ever use of any method, any modern method, and any traditional method, all women are included in the analysis, and women who report that they have not heard of a method in a particular survey are coded as never having used the method for that survey.

Table 7.2 Indices of disagreement for ever use of contraceptive methods, and level of ever use of specific methods, main survey and reinterview survey, Pakistan DHS, 1990-91

Method	Percent dis-crepant	Percent moving from "yes" to "no" response	Percent ever used in main survey	Percent ever used in re-interview survey	Kappa	Number
Pill	9.2	4.4	7.0	7.3	0.30	284
IUD	7.9	4.7	12.3	10.9	0.61	225
Injection	6.6	3.3	6.6	6.6	0.46	262
Diaphragm/Foam/Jelly	14.2	6.1	6.1	8.1	-0.07	32
Condom	21.0	15.4	34.9	25.0	0.51	159
Female sterilization	2.9	1.1	5.5	6.2	0.74	324
Male sterilization	0.0	0.0	0.0	0.0	-	69
Periodic abstinence	30.8	22.0	37.3	24.1	0.29	56
Withdrawal	39.2	33.3	49.7	22.2	0.21	49
Any method	16.0	8.8	24.8	23.2	0.56	474
Any modern method	13.5	7.9	22.4	20.0	0.60	474
Any traditional method	11.0	7.2	11.4	8.1	0.38	474

Note: Ever use of individual methods is based only on women who reported having heard of the method in both surveys. Ever use of any method, any modern method, and any traditional method is based on all women.
- Not defined

Among women who had heard of the method in both surveys the percentage who were discrepant on ever use of the method ranges from 0 percent for male sterilization, which no women reported in either survey, to 31 percent for periodic abstinence, and 39 percent for withdrawal. For most methods the discrepancies appear to be fairly random, in that around half are women shifting from reporting having used the method in the main survey to not having used it in the reinterview survey. Hence there does not appear to be any evidence of a systematic shift to reporting more use in either survey for most methods. Consequently, among women who had heard of the method in both surveys the percentage who reported ever use of the method is generally similar in the two surveys. For the traditional methods there does appear to be less reporting of ever use in the reinterview survey than in the main survey, and generally ever use of periodic abstinence and withdrawal seems to be reported less consistently than ever use of individual modern methods. One exception is ever use of condoms, which is also reported very inconsistently and again there is a tendency to report less use in the reinterview survey. The reinterview study conducted in the Dominican Republic experimental DHS survey also found that the consistency of reporting of ever use of periodic abstinence and withdrawal was lower than for modern methods with the exception of condoms (Westoff, Goldman and Moreno, 1990). Similar findings concerning the reporting of use of traditional methods have been observed in Sri Lanka and Haiti where differences between the contraceptive prevalence rates estimated from surveys conducted close together can be attributed to differential reporting of traditional method use (Mauldin and Segal, 1988). However, it is important to be aware of the fact that the estimates of consistency for use of periodic abstinence and withdrawal are based on very small numbers of women, since very few women reported knowing of either method in both surveys. This caveat also applies to the results for male sterilization and vaginal methods.

When knowledge and use of all methods are considered together, 16 percent of women were discrepant on ever use of any method of contraception, and 9 percent shifted from reporting use in the main survey to reporting no use in the reinterview survey. The percentage discrepant is only 11 percent for ever use of traditional methods, and 14 percent for ever use of modern methods. About half of the discrepancies for both groups of methods are again women who report use in the main survey but no use in the reinterview survey, and overall ever use of any method, any traditional method, and any modern method is slightly lower in the reinterview survey. However, ever use of modern and traditional methods reported in the main survey is higher for the reinterview sample than for the full Pakistan DHS sample.

The value of kappa suggests that ever use of traditional methods is reported less reliably than ever use of modern or any methods. This is partly because around 90 percent of women report that they have never used a traditional method in each survey. Since the never used response dominates in both surveys, the majority of women would be placed in the never used category in both surveys even under a completely random distribution of women across the cells of the table, given the observed distribution of ever use in each survey. Hence the reliability of ever use of a traditional method of contraception is low. This also applies to ever use of any method and any modern method, but to a lesser extent, and also to individual methods, since so few women have ever used a method of contraception. This contributes to the low reliability, as measured by kappa, for all the ever use variables, despite reasonably high levels of consistency.

Only 75 women reported that they had ever used a method of contraception in both surveys. For these women, the consistency of reporting of the first method used was examined. Somewhat surprisingly, 40 percent of these women were discrepant on the first method they reported using. This represents only fair agreement beyond chance with unweighted kappa equal to 0.51. The most commonly mentioned method in both surveys was the condom, but 35 women in the main survey reported using the condom as their first method, compared to only 21 in the reinterview survey. Proportionally more women changed from reporting male methods (condom, withdrawal) than female methods as their first method in the main survey, but discrepancies occurred for all methods. For injections and periodic abstinence, all women who reported these two methods as their first method in the main survey also reported them as their first method in the reinterview survey. However, only seven women reported using these methods as their first method in the main survey, and additional women reported using them first in the reinterview survey. No pattern is seen in the methods reported in the reinterview survey by discrepant women.

In addition to the first method reported, it is also possible to examine the consistency of reporting of the number of living children at first use in the two surveys for the 75 women who reported use in both surveys. Again, the discrepancy is relatively high, with 46 percent of women discrepant on this variable. Reliability is reasonable though, with weighted kappa equal to 0.70. This is again partly due to the relatively wide range of values for the number of living children at first use. The majority of discrepancies are of only one or two living children, but there are a few women who are discrepant by more than two living children. This discrepancy may reflect omission of an isolated early spell of use in one survey or it may be mismatches. No clear pattern can be seen in the discrepancies, but the numbers are very small.

Chapter 8

Fertility Preferences

A number of the questions on fertility desires and intentions from the main Pakistan DHS were included in the reinterview survey. The consistency of these questions is of particular interest because they are questions concerning attitudes and opinions, rather than facts. As such, discrepancies in these questions do not necessarily indicate incorrect responses or confusion over the interpretation of questions, but may represent genuine changes in attitudes or opinions between the two survey dates. In general, women may change their fertility intentions and desires in response to changes in their individual circumstances or in response to changes in social norms. Alternatively, women's fertility preferences may still be in a state of development and may not be very clearly defined, resulting in variation in the responses an individual woman gives. In addition, the concepts represented by these questions may be alien to many women, and consequently they may find the questions difficult to understand, particularly if control over individual fertility is not well established. In this situation a high proportion of fatalistic non-numeric responses, such as "up to God/Allah," might be expected together with a relatively large amount of shifting between numeric and non-numeric responses in the two surveys. The consequence of all this is that a higher degree of discrepancy would be expected for these questions than for the earlier ones. However, these questions are widely used for many purposes, so their reliability in the face of these issues is important.

Table 8.1 presents the percentage of women who are discrepant on each of the fertility preference questions together with unweighted kappas and the number of cases on which these figures are based. Note that unweighted kappa is used here because of the presence of non-numeric categories.

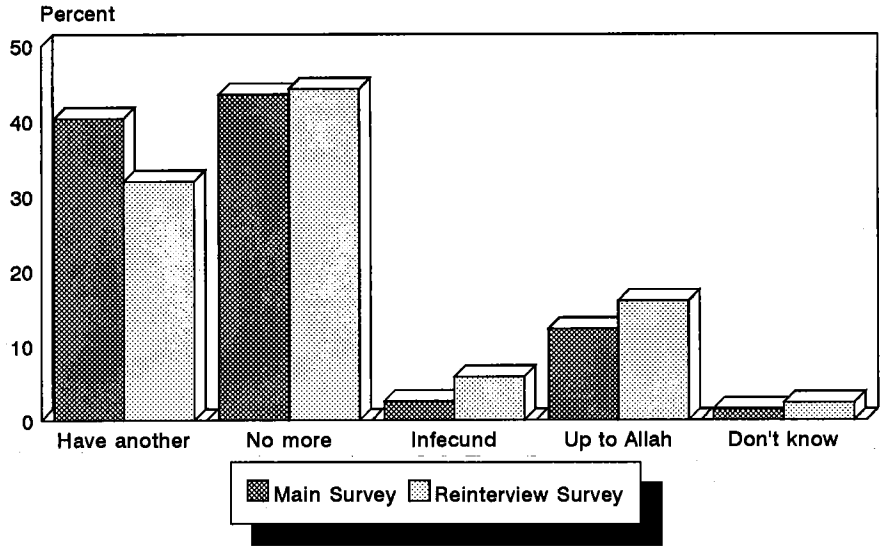
Table 8.1 Indices of disagreement for reliability of reporting of fertility preferences, Pakistan DHS, 1990-91

Fertility preferences	Percent discrepant	Percent numeric to non-numeric	Percent non-numeric to numeric	Kappa	Number
Want more children	37.0	NA	NA	0.44	427
Sex preference of next birth	30.2	NA	NA	0.45	107
Time to next birth	60.6	16.6	4.0	0.15	107
Ideal family size	47.1	22.7	10.7	0.22	474
Ideal number of boys	49.7	6.8	2.4	0.14	109
Ideal number of girls	54.3	6.8	2.4	0.14	109
Ideal number of either sex	11.1	6.8	2.4	0.15	109

Note: Figures are based on women who answered each question in both surveys.
NA = Not applicable

As in earlier analyses, only women who answered a particular question in both surveys are included in the analysis of that question. In total, 37 percent of women are discrepant between the two surveys in their response to whether they would like any more children, resulting in only moderate reliability for this question (kappa = 0.44). The pattern in these discrepancies is not very clear, but there is some indication of shifting towards the more fatalistic "up to Allah" response in the reinterview survey. This is seen more clearly in Figure 8.1 which presents the distribution of women by their desire for more children in the two surveys.

Figure 8.1
Percent distribution of matched women by desire for more children in main survey and reinterview survey, Pakistan DHS, 1990-91



Note: Based on women who were currently married and non-sterilized in both surveys (N=427).

There are fewer women reporting that they want another child in the reinterview survey than in the main survey, but more report that it is up to Allah or that they don't know. Of course, some women who reported that they wanted another child at the time of the main survey may have had another child between the main survey and the reinterview survey. Consequently, they may be discrepant in their response due to this change in their circumstances. The proportion reporting that they do not want another child is about the same in the two surveys. The higher proportion reporting that they cannot get pregnant in the reinterview survey may represent genuine changes in some women's perception of their fecundity status since the main survey as a result of the time interval between the surveys.

Women who reported that they wanted another child were then asked what sex they would like that child to be. Of the women who answered this question in both surveys, 30 percent were discrepant in their response. Virtually all of the discrepancies were due to women reporting that they wanted a boy in one survey, but that the sex did not matter in the other survey. However, the numbers involved are relatively small. One source of discrepancy in this question that is worth noting is that women may not be talking about the same child if they had had a child between the two surveys.

Women who wanted another child were also asked how long they would like to wait until their next birth. In the survey, responses were recorded in months or years and in a number of non-numeric categories. For the purposes of this analysis the numeric responses were grouped into the following categories (in years): less than two, two, three, four, and five or more years. Responses of "soon/now" were included in the less than two years category, but all other non-numeric responses were left as they were originally coded. Despite the grouping, consistency between the two surveys is relatively low, with 61 percent of women giving different responses in the two surveys. This low level of reliability is confirmed by the low value of kappa but some caution should be exercised in the interpretation of kappa in this situation because it is sensitive to

the choice of grouping of the responses (Maclure and Willett, 1987). There are no clear patterns in the discrepancies, but the overall distribution of women in the two surveys suggests that more women gave non-numeric answers in the reinterview survey than in the main survey. This is also seen in columns three and four in Table 8.1, which show that 17 percent of women gave a numeric response in the main survey but a non-numeric response in the reinterview survey, whereas only 4 percent moved from a non-numeric to a numeric response.

Table 8.2 presents the percentage discrepant and weighted kappa (using the weights for an ordinal variable given in equation 6) for women who gave a numeric response for their desired time to next birth in both surveys. The fact that these women gave a numeric response in both surveys may suggest that they are better able to understand the questions and relate to the abstract concepts of family formation preferences. Consequently, these women might be expected to have more clearly defined fertility goals, and as such, they might also report more consistently in the two surveys. However, reliability is still relatively low, with half of the women discrepant between the two surveys, and kappa still indicating relatively poor agreement beyond chance.

Table 8.2 Indices of disagreement for reliability of numeric responses on fertility preferences, Pakistan DHS, 1990-91

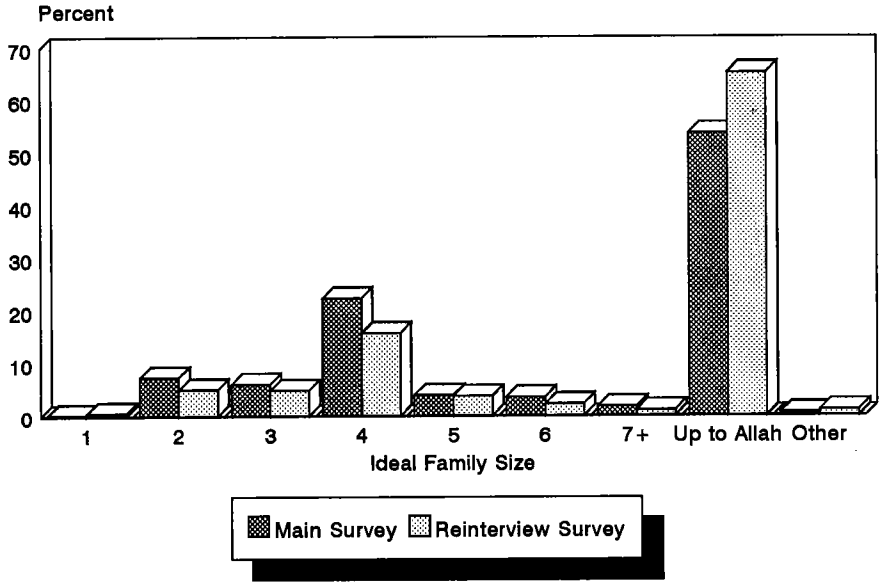
Fertility preferences	Percent discrepant	Weighted kappa	Number
Time to next birth	50.0	0.30	84
Ideal family size	55.0	0.24	109
Ideal number of boys	45.2	0.12	97
Ideal number of girls	50.5	0.27	97
Ideal number of either sex	2.1	-0.01	97

Note: Based only on women who gave a numeric response to each question in both surveys.

All women were asked their ideal family size in the two surveys. Overall, 47 percent of women did not give the same response in the two surveys, and reliability as measured by kappa is low (Table 8.1). Again, there is evidence of an increased tendency to give non-numeric answers in the reinterview survey compared to the main survey. This is reflected by the fact that 23 percent of women shifted from a numeric response in the main survey to a non-numeric response in the reinterview survey, whereas only 11 percent shifted from a non-numeric response to a numeric one. This can also be seen in Figure 8.2 which presents the distribution of women by ideal family size in each survey. The percentage of women reporting "up to Allah" is considerably higher in the reinterview survey, and correspondingly, the percentage of women reporting almost every numerical response is lower in the reinterview survey. Consistency is actually lower among women who gave numeric answers in both surveys, with 55 percent not giving the same response (Table 8.2). The discrepancies in these numeric responses are fairly random in direction, but there is a slight tendency to give a higher desired family size in the reinterview survey than in the main survey. The difference is usually only one or two though, and may reflect genuine changes in opinions. In particular, it may reflect upward revisions in ideal family size in response to a new pregnancy. However, the numbers are too small to investigate this possibility further.

The final question in the section on fertility preferences refers to the sex preferences that women have for their family. This question was only asked of women who state a numeric ideal family size. Table 8.1 shows that the consistency of reporting was slightly higher for the desired number of boys than for the desired number of girls. Only 11 percent of women were discrepant on the number of children of either sex they wanted, but this is partly due to the fact that the majority of women report the exact number of boys and girls they would like in both surveys, so the number of either sex is zero. Again, agreement between the two surveys, beyond that expected by chance, is low. There are very few non-numeric responses to these questions compared with the earlier questions on fertility desires, but there are more non-numeric responses in the reinterview survey than in the main survey. Sample sizes are small and there are no obvious patterns

Figure 8.2
Percent distribution of matched women by ideal family size in main survey and reinterview survey, Pakistan DHS, 1990-91



Note: Based on all women (N=474).

in discrepancies, but there is a very slight tendency for women to report that they want fewer boys in the reinterview survey than they reported in the main survey.

Chapter 9

Discussion and Recommendations

The consistency of a wide range of variables collected in the Pakistan DHS has been examined in this analysis of the Pakistan reinterview survey. Overall, the results are somewhat disappointing in that the majority of variables are not reported consistently, and there is evidence that both recent fertility rates and infant and child mortality rates are underestimated in the Pakistan DHS. However, given the cultural context of Pakistan, low reliability was anticipated.

In general the reporting of background characteristics is reasonably consistent and reliable. The one exception is the type of blood relationship between spouses, which seems to suggest some confusion among women or interviewers about the classification of such relationships. In contrast, age and date reporting is very inconsistent between the two surveys. Such inconsistency was expected since ages and dates are not widely known in Pakistan, and this has been reflected in poor quality age data in all censuses and surveys in the country. The poor quality of the age data in the Pakistan DHS is apparent from the low percentage of women able to report both their month and year of birth and also in the degree of heaping present in the age distribution of women. Therefore, the reinterview survey merely confirms what was already known. However, what the reinterview survey does demonstrate is that the discrepancies between the two surveys are fairly random, and consequently they do not affect the overall age distribution to any great extent. This, in turn, means that simple aggregate-level age measures, such as the median age at first birth and median age at first union, are fairly robust to the age reporting errors. In contrast, measures that rely heavily on the individual age data, such as the TFR, are very sensitive to age and date reporting errors, and this must be kept in mind when using these measures in countries with poor age data. In addition, the fact that the discrepancies in reported age between the two surveys appear random does not preclude the possibility that the age data in both surveys are subject to similar systematic biases which would not be detected by comparisons between the two surveys.

Comparison of the birth history data collected in each survey suggests that the reporting of the total number of surviving children is very consistent. However, the analysis confirms that there was substantial displacement of the births of surviving children to before January 1986 in the main survey, probably because of attempts by interviewers to reduce the workload associated with the health and the anthropometric sections of the questionnaire. Again, this was detected in the original analysis of the main survey, although the reinterview survey does clarify the fact that the deficit of surviving births since January 1986 appears to be due primarily to displacement of surviving births rather than to omission. However, the possibility of omissions in both surveys cannot be ruled out entirely. In contrast, substantially more dead children were reported in the reinterview survey than in the main survey, providing evidence that dead children were omitted in the main survey. In addition, the omissions appear to be spread across all ages at death under one year and throughout the birth history. Consequently, these omissions were not picked up by the standard data quality tests for the omission of deaths that were applied to the Pakistan DHS, since these tests assume that omissions will be concentrated among deaths at very early ages and deaths further back in time. This displacement of surviving births and omission of deaths in the main survey has serious implications for the estimates of fertility and child mortality rates obtained from the Pakistan DHS, which are almost certainly substantially below the true values. For example, adjustment of the TFR calculated from the full Pakistan DHS based on the finding of the reinterview survey suggests that the TFR in Pakistan for the six-year period prior to the survey is probably about 5.9 rather than 5.4 as originally suggested. However, it is difficult to

be certain how accurate the adjustments based on the reinterview survey are, due to the small sample size and highly clustered sample design used for the reinterview sample.

Knowledge of contraceptive methods appears to be higher in the reinterview survey than in the main survey. This may reflect a genuine increase in knowledge between the two survey dates, particularly since in the reinterview survey women may remember methods from the main survey. Use of contraception is reported reasonably consistently in the two surveys. This is partly due to the fact that the majority of women report that they have never used a method in both surveys. Consistency of reporting of ever use of contraception among women who report use of contraception in at least one survey is lower. This is also reflected in relatively low consistency in the reporting of the first method used and the number of living children at first use. This inconsistency in the reporting of use may be due to the general low level of contraceptive use in Pakistan. For example, a woman may be reluctant to report use if she feels that contraceptive use is not widely accepted, and her propensity to report use may be sensitive to her perceptions of the interviewer.

Finally, the analysis of the reinterview survey confirms that the data on fertility preferences are not particularly reliable. Again, this is not surprising given the large percentage of women in the full Pakistan DHS that give the fatalistic response "up to Allah" for ideal family size, which demonstrates that the concept of an ideal family size is not well-established. Consequently, many women do not really have an ideal family size as such.

This analysis of the reinterview survey has been instructive in highlighting problems in the Pakistan DHS data, and, more generally, in collecting demographic data through surveys in countries with low levels of literacy. The next question that needs to be asked is, "What should be done in the future to address these issues?" Certainly many of the inconsistencies identified, for example, in age and date reporting or fertility preferences, are due to the cultural environment in which the survey is conducted. This limits the role that further improvements in survey design and implementation can have in obtaining better quality data for these variables. For example, while every effort can and should be made to obtain the best possible age data, if the respondent really does not know her age, some degree of error will be introduced into the data. Obviously this will be much more severe in societies in which knowledge of ages and dates is low than in societies in which knowledge of ages and dates is quite good. In addition, it is important to recognize that increasing the completeness of age and date information will often lead to decreased reliability of those data. However, it is preferable to have approximate information on ages and dates than no information at all.

The discrepancies in the birth histories in the two surveys are particularly serious, and further emphasis needs to be placed on obtaining high quality birth history data since these data are so fundamental for demographic analysis. The fact that respondents reported child deaths in the reinterview survey that they did not report in the main survey suggests that the omissions of deaths in the main survey were unlikely to be due simply to respondents not wanting to mention deaths. Certainly, the heavy emphasis placed on high quality training that is stressed in all DHS surveys needs to be maintained. However, it is important to recognize that although high quality training for interviewers is absolutely vital, it is not the only factor that affects their performance and motivation. Indeed, interviewer training in Pakistan was deemed to have been successful, and the interviewers appeared both to have a thorough grasp of the survey and to be enthusiastic. However, the final report of the Pakistan DHS documents a number of problems encountered during the implementation of the survey, including the poor quality of work by some supervisors, landslides and flooding in some regions which made access to these areas extremely difficult, delays in paying interviewers, security problems in some areas, and low cooperation in some areas where interviewers were mistaken for members of the narcotics board or television license examiners (National Institute of Population Studies and IRD/Macro International Inc., 1992). Such problems are certain to have an impact on the work of interviewers. It is important that equal emphasis is placed on providing high quality training for all personnel

involved in the survey, and that every effort is made to maintain the initial enthusiasm of staff in the face of the practical difficulties that will inevitably arise in most countries.

The displacement of births in the main survey to before 1986 does appear to be linked to the presence of the health section since much less displacement was found in the reinterview survey. This type of displacement has been observed in a number of DHS surveys and was particularly severe in the Pakistan DHS. This suggests that the inclusion of a health section restricted to children under a given age can compromise the quality of the demographic data collected, introducing a conflict between the demographic and health elements of some DHS surveys. In DHS-III surveys the health section covers a three-year period rather than the five-year period covered in DHS-I and DHS-II surveys. This should considerably reduce the burden of the health section on the interviewers, thereby reducing the incentive to displace births. If interviewers continue to displace births, it could present additional problems for the calculation of the TFR since a three-year period before the survey also is used, which would coincide with the cut-off for inclusion in the health section. If displacement does occur it may be necessary to use a five-year period for the calculation of the TFR to ensure that all displaced births are included. This would still be preferable to the longer period that may be required with a five-year health section.

It is important to note that many of the problems identified by this analysis of the reinterview survey, such as displacement of births to before 1986 and poor age reporting, can also be identified by examination of the original Pakistan DHS data, and the addition of data quality tables to the final reports of DHS surveys is a useful development in this respect. The use of field check tables during the data entry process in DHS-II and DHS-III surveys also enables many data quality problems to be identified at an early stage so that they can be addressed while the survey is still in the field. This should have a positive impact on the quality of data obtained. In Pakistan the field check tables were not run until very late in the fieldwork due to the fact that the Persian Gulf War interfered with the amount of technical support that could be provided during the survey itself. Hence, although problems that the reinterview highlights were identified by these tables, it was too late in the fieldwork for much to be done about them.

Although standard data quality checks can pick up many data quality problems, the omission of deaths in the Pakistan DHS identified by the reinterview survey would not be picked up by such checks. Thus, this raises the issue of whether reinterview surveys should be conducted in all DHS surveys to check consistency of reporting. Certainly in the case of Pakistan the reinterview survey appears to be of better quality, in general, than the main survey and therefore has been useful in evaluating the quality of the original survey. However, this is by no means certain to be the case and it is very likely that in some situations a reinterview survey may be of significantly poorer quality than the main survey. In such a situation the reinterview survey will be of limited use in evaluating the quality of the original survey. In addition, consistency and reliability do not imply validity, so even if the reinterview survey suggests high reliability of responses this does not necessarily mean that the responses are correct. The danger of implementing reinterview surveys in all countries is that they divert both staff time and funds away from the main survey, which could have an adverse effect on the quality of data collected. Consequently, it would seem that a more efficient approach would be to ensure a thorough evaluation of the data quality for all DHS surveys based on demographic analyses, internal consistency checks, field check tables, and spot-checking in the field. This approach has limitations in that the types of data quality checks currently in use do not pick up all data quality problems, as has been demonstrated in this report, but if a large number of data quality problems are identified by standard techniques, it should alert the analyst to the fact that there are likely to be other hidden problems. It is also possible that some of the inconsistencies in responses in the two surveys in Pakistan identified by this report could be due to data entry errors, so expansion of the current data verification procedures could also have a positive impact on data quality, particularly for surveys in which existing verification procedures indicate a relatively large number of data entry errors.

Despite their limitations, reinterview studies do provide some form of external validation of the reliability of responses and it has been argued that reinterview surveys should become standard practice (Moser and Kalton, 1971). Reinterview surveys can also serve as a deterrent to poor quality work among interviewers, supervisors, and project managers if they know that their work will be subject to intensive scrutiny. Hence, any reinterview survey should be designed at the time of the main survey and its existence should be made known to the survey staff. If reinterviews are conducted we recommend that they be conducted closer to the date of the original survey than was the case in Pakistan to reduce the problem of adjusting for changes between the two surveys. However, this may increase contamination of the reinterview responses by recall of responses from the main survey or it may reduce data quality in the reinterview survey due to interviewer and/or respondent fatigue. Consequently some degree of compromise in the timing of the reinterview survey is required. Correct identification and matching of women is crucial and reducing the interval between the two surveys may also reduce identification problems. In addition, the sample design for the reinterview survey should be aimed at reducing matching problems. For example, in Pakistan clusters were selected for the reinterview and all women in these clusters who were originally interviewed were reinterviewed. This approach seems to have been relatively effective in identifying women correctly and may be better than selecting women at random within clusters for the reinterview survey. The questionnaire used for the reinterview survey needs to be as carefully designed as the original questionnaire, focusing on the assessment of the reliability of responses in the original survey for key questions for which analysis of reliability is feasible. For example, current status data are likely to change between surveys unless the surveys are very close together. Consequently, reinterview surveys are of limited use in assessing reliability of responses to these questions unless the reinterview survey specifically asks about the status at the time of the main survey. However, this then introduces an additional source of discrepancy since the form of the question is different in the two surveys; in particular, the response in the reinterview survey would be subject to recall error.

To conclude, the analysis presented in this report demonstrates that response reliability can be very low on some questions in some surveys. Low response reliability may have little impact on aggregate measures if the discrepancies are random in direction but it can have considerable impact if systematic differences exist between responses. Some analyses are more sensitive to low reliability than others and analysts need to be very aware of the implications of measurement error for their analyses. If more is known about the nature of data quality problems, measures and analyses can be designed to minimize the impact of the errors. One example is lengthening the period over which demographic estimates are calculated to reduce the bias caused by birth displacement. However, it is important to note that these results concerning the level of reliability of the Pakistan DHS data are not generalizable to all DHS surveys since there is considerable variation between surveys in the cultural context in which the survey is conducted and in the ease of implementation of the survey. As outlined above, the Pakistan DHS encountered a number of problems in implementation which, in conjunction with the difficulties in obtaining high quality demographic data imposed by the cultural and social context of the country, are reflected in the data quality tables in the final report. In view of these circumstances the reliability of the Pakistan DHS data would be expected to be lower than that obtained in most other surveys. However, studying response reliability in such a difficult setting is an important contribution to our understanding of the nature of data collection problems in societies such as Pakistan where inaccuracy in responses is inevitable.

REFERENCES

Arnold, F. 1990. Assessment of the Quality of Birth History Data in the Demographic and Health Surveys. In *An Assessment of DHS-I Data Quality*. DHS Methodological Reports, No. 1. Columbia, Maryland: IRD/Macro Systems, Inc.

Croft, T. 1991. Data Editing and Imputation. In *Proceedings of the 1991 Demographic and Health Surveys World Conference, Vol. II*. Columbia, Maryland: IRD/Macro International Inc.

Fleiss, J.L. 1981. *Statistical Methods for Rates and Proportions*. 2nd ed. New York: John Wiley and Sons.

Kaplan, A. 1964. *The Conduct of Inquiry, Methodology for Behavioural Science*. San Francisco: Chandler.

MacDonald, A.L., P.M. Simpson and A.M. Whitfield. 1978. *An Assessment of the Reliability of the Indonesia Fertility Survey Data.* WFS Scientific Reports No. 3. Voorburg, Netherlands: International Statistical Institute.

Maclure, M. and W.C. Willett. 1987. Misinterpretation and Misuse of the Kappa Statistic. *American Journal of Epidemiology* 126(2):161-169.

Mauldin, W.P. and S.J. Segal. 1988. Prevalence of Contraceptive Use: Trends and Issues. *Studies in Family Planning* 19(6):335-353.

Moser, C.A. and G. Kalton. 1971. *Survey Methods in Social Investigations.* London: Heinemann.

National Institute of Population Studies and IRD/Macro International Inc. 1992. *Pakistan Demographic and Health Survey 1990/91*. Columbia, Maryland: IRD/Macro International.

O'Muircheartaigh, C.A. 1982. *Methodology of the Response Errors Project.* WFS Scientific Reports No. 28. Voorburg, Netherlands: International Statistical Institute.

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

O'Muircheartaigh, C.A. 1984b. *The Magnitude and Patterns of Response Variance in the Lesotho Fertility Survey*. WFS Scientific Reports No. 70. Voorburg, Netherlands: International Statistical Institute.

Retherford, R.D. and I. Alam. 1985. *Comparison of Fertility Trends Estimated Alternatively from Birth Histories and Own Children.* Papers of the East-West Population Institute, No. 94. Honolulu: East-West Center.

Retherford, R.D. and G.M. Mirza. 1982. Evidence of Age Exaggeration in Demographic Estimates for Pakistan. *Population Studies* 36(2):257-70.

Retherford, R.D., G.M. Mirza, M. Irfan and I. Alam. 1987. Fertility Trends in Pakistan - the Decline That Wasn't. *Asian and Pacific Population Forum* 1(2):1-10.

Shah, Gulzar. 1993. Expected Patterns of Misreporting in Surveys of Pakistan. Unpublished paper.

Shami, S.A., J.C. Grant and A.H. Bittles. 1994. Consanguineous Marriages within Racial/Occupational Class Boundaries in Pakistan. *Journal of Biosocial Science* 26(1):91-96.

Shryock, H.S. and J.S. Siegel. 1971. *The Methods and Materials of Demography, Vol. 1.* Washington, D.C.: U.S. Government Printing Office.

Westoff, C.F., N. Goldman and L. Moreno. 1990. *Dominican Republic Experimental Study: An Evaluation of Fertility and Child Health Information.* Columbia, Maryland: IRD/Macro Systems, Inc.