Chapter III
METHODOLOGY

The methods and materials used for the study like objectives, theoretical framework, hypotheses, data requirement and source of data, different approaches to analyse the data and quality of data with limitations are discussed in detail in this chapter. As noted earlier, the present study seeks to examine how the family building process is influenced particularly by the educational levels of the spouses, specifically to examine the relationship between educational attainment of the spouses and fertility at individual level in Uttar Pradesh state, India. To this end, the following specific objectives have been formulated.

Objectives

1) To examine the differentials in age at marriage, total fertility rate, age at first birth, desired fertility, and contraceptive use across the educational classes of the spouses.

2) To assess the net effects of education of women and husbands on children ever born, age at first birth, desired fertility and contraceptive use controlling for other variables.

3) To examine the differentials in age at marriage, parity progression ratios and closed birth intervals by education.

4) To examine the trends in age at marriage, parity progression ratios and closed birth intervals by education.

5) To assess the net effects of education of woman and husband on the risk of subsequent births.

6) To examine the trends in educational differentials in age at marriage, parity progression ratios and birth intervals in order to see whether the effects of education on family building variables have changed over time.

Theoretical Framework

The direct effect of education in demographic literature on fertility has been widely studied (Cochrane, 1979; Kasarda et al., 1986; Chamratrithirong et al., 1992; Castro Martin and Juarez, 1995; U.N., 1995; Jejeebhoy, 1995). According to Easterline’s (1978) framework, the supply of and demand for children, and the cost of fertility regulation can determine fertility outcome. Education may reduce the supply of children, mainly through the postponement of marriage, thus shortening a woman’s overall period of exposure to pregnancy (Smith, 1983). Some of the earlier studies have documented that women with
lower education have a large number of closely spaced children, whereas those with higher education have smaller number of widely spaced children (Rao and Balakrishnan, 1989; Wineberg and McCarthy, 1989).

The demand for children refers to family size preferences (the number of children a couple wants to have). Education shapes women’s values, orientations and attitudes and tends to favour a small family norm. Education is associated with a higher perceived ability to afford children but also with higher perceived costs of children. The net effect of these factors generates an inverse relation between education and desired family size that tends to reduce the demand for children (Cochrane, 1979). The cost of fertility regulation refers to factors that constrain the use of contraception intended to reduce the actual supply of children in order to match demand (desired family size). Education tends to reduce such costs, since it increases awareness, access, affordability and acceptance of birth control (U.N., 1995).

A clear understanding of the effects of education on fertility would be possible if the effects on various stages of the family building process are examined. The family building process begins with the first birth and hence age at first birth is a key variable to be investigated. However, for a population like Uttar Pradesh, in which fertility takes place primarily within marriage, marriage is a pre-requisite to the initiation of the family building process. Hence, the effects of education on first age at marriage and then the age at first birth need to be examined. It is to be noted that after marriage and first birth, further family building process is represented by progression to second and subsequent parities. Hence, an analysis of the effects of education on parity progression ratios and birth intervals would aid an understanding of the effect on fertility. If fertility is within the calculus of conscious choice, the timing of progression to subsequent births and the event of progression itself would depend on whether the couple wanted an additional child (additional children). This calls for an assessment of the effect of education on desire for additional children. A desire for no more children will influence fertility if it leads to contraceptive use. Hence, contraceptive use is an important factor to be examined. Thus, assessment of the effects of education on age at marriage, age at first birth, birth intervals, parity progression ratios, desire for children and contraceptive use will provide a better understanding of how education influences fertility.
Besides education, the type of occupation and the level of income, place of residence (rural or urban), cultural factors such as religion and caste are also likely to affect fertility both directly and indirectly. For instance, women engaged in modern sector employment and earning higher income would marry late and also exhibit lower desire for children (Cochrane, 1979; U.N., 1987) because of the higher aspirations as well as higher cost of rearing children (quantity-quality trade off). These in turn will lead to lesser proportion of women to higher order births since these women are likely to use contraceptive methods extensively and thereby, maintain larger spacing between the births. Similarly residence, religion and caste can also influence the demand and supply of children and thus family size desire, contraceptive use and fertility. These variables can also affect age at marriage. Since many of these socio-economic variables are likely to be associated with education, it is essential that, in order to see the effect of education as such (net effect), the other socio-economic variables are controlled statistically.

It has been generally observed that fertility differentials emerge during the early phase of transition since some sections of society begin to control fertility but become narrow during the late phase of transition as the fertility of various sections converges to a low level. Therefore, an assessment of changes in differentials over time gives an indication of the nature and stages of fertility transition.

**Hypotheses**

Some of the following major hypotheses are proposed for empirical examination.

1. There exists a positive relationship between education and age at marriage.
2. As the educational level of the women increases, probability of transition to the next parity (as measured by parity progression ratio) declines first for higher order births and later for the lower order births.
3. As the educational level of woman increases, the spacing between births becomes longer.
4. An increase in the level of education brings down the demand for children (desire^family size) and raises contraceptive use.
Data Source and Sampling

Uttar Pradesh, one of the largest states in India, forms the universe of the present study. In India, the National Family Health Survey (NFHS), 1992-93 was carried out in 24 states and the national territory of Delhi. The major objective of the survey was to collect reliable and up-to-date information on fertility, family planning and maternal and child health. In Uttar Pradesh, the NFHS gathered information from a representative sample of 11,438 ever married women of age 13-49 from 10,110 households. The survey was conducted between 10th October 1992 and 22nd February 1993.

The NFHS has used three types of Questionnaires - Woman's Questionnaire, Household Questionnaire and Village Questionnaire. The Woman's Questionnaire is subdivided into seven sections where information pertaining to respondent's background, reproduction, contraception, health of children, fertility preferences, husband's background, woman's work and height and weight of living children born since 1st January 1986 have been covered. The Household Questionnaire contains information regarding the household, such as source of water, type of toilet facilities, materials used in the construction of the house, source of lighting, cooking fuel, ownership of agriculture land and livestock, ownership of various consumer durable goods and characteristics of the head of the household such as religion, caste or tribe. The Village Questionnaire collected information such as electricity, water, transport, and educational and health facilities available in the sample villages. Since most of the information pertaining to fertility behaviour and socio-economic status was available in the Woman's and Household Questionnaires, the data had been availed from the same.

The sample design adopted for the NFHS is a systematic; two-stage stratified sample of households. All the districts in Uttar Pradesh were subdivided into six contiguous regions according to their geophysical characteristics. Within each geophysical region, districts were classified into backward and non-backward districts. Districts in these six regions were combined into one or more of eight sampling domains. However, the sample is not self-weighted in the case of either the state as a whole or its urban and rural areas.
Eight different sample weights were used to adjust for non-response, and normalised so that total number of weighted cases is equal to the total number of unweighted cases (for details see PRC, Lucknow and IIPS, 1995, pp. 11-24).

**Unit of Analysis**

Since the focus is on the relationship between education of the spouses and fertility, *ever-married women* in the age group of 13-49 are the unit of analysis. For the analysis of some of the variables, additional constraints are imposed or relaxation allowed; these are mentioned at appropriate places.

**Variables Used for Analysis**

**Dependent Variables**

The following dependent variables have been used for different analyses.

1) **Children Ever Born**: Number of children ever born to women till the survey date (Question No.208 of the NFHS)

2) **Age at Effective Marriage**: Age at first cohabitation with the husbands in completed years (Q.116C of the NFHS).

3) **Age at First Birth**: This variable is computed based on the difference between woman’s date of birth (Q.105C of the NFHS) and date of birth of the first child (Q.219C of the NFHS).

4) **Desire for Additional Children**: In the NFHS the sterilised women were not asked any question on desire for additional children and were presumed to desire no more. The currently married and not sterilised non-pregnant women were asked (Q.503) “Would you like to have (a/another) child or would you prefer not have any (more) children”? The pre-coded responses were Yes (would like to have another child), No, Can not get pregnant, Up to God, and Don’t know. For pregnant women the question was “After the child you are expecting would you like to have another child or would you prefer not to have any more children?” The pregnant women were also asked (Q.229), “At the time you became pregnant, did you want to become pregnant then, did you want to wait until later, or did you not want to become pregnant at all?” Thus, while Q.503 relates to desire at the time of
survey, Q.229 relates to desire prior to the current pregnancy. While analysing the desire for additional children we prefer to use the responses to Q.229 for pregnant women as the sex of the current pregnancy is not known to incorporate it in the explanatory variable namely sex composition of the living children.

The (non-pregnant) women who said that they wanted no more children and those who are sterilised are assumed to be, not wanting additional children. The pregnant women who stated that at the time they became pregnant they did not want to be pregnant at all were treated as women not wanting additional child. The others include non-pregnant women who desired another child, those who did not explicitly say so, but left it to God, those who were undecided / did not know about it, and pregnant women who at the time they became pregnant wanted to be pregnant then or wanted to be pregnant later. These women are deemed not desiring no more children, since they did not explicitly say that they did not want additional children. Women who thought that they could not get pregnant (528 women) were dropped from the analysis, as were the women who did not respond to the question (25 women).

5) Contraceptive Use: In order to elicit the information about the use/non-use of contraception in the NFHS question nos. 312 and 313 were posed to the respondents. Based on the responses to these questions contraceptive use is measured as 1 (i.e., when woman or couple currently using any method to delay or avoid getting pregnancy) and 0 when she (they) do not.

6) Parity Progression Ratios: Progression ratios to the successive parities have been computed from the birth histories of the women (for details see Chapter VII).

7) Birth Intervals: In the NFHS date of birth of each child was given in birth histories, i.e., starting from the first birth to the last birth, at the time of the survey. From this information birth intervals between successive births (i.e., first to second, second to third etc.) have been computed (for more details see Chapter VIII).

*The possibility of using more than two categories (Do not desire for additional children, Desire for additional children, Up to God, Undecided) was considered but the frequencies in the last two categories were too small for a meaningful analysis. Hence, these were pooled with category Desire for additional children.
**Explanatory Variables**

Most of the explanatory variables used in the present research work have been classified (without distorting the original nature as given in the NFHS) as shown in Table 3.1.

**Table 3.1. Description of the Explanatory Variables**

<table>
<thead>
<tr>
<th>Explanatory Variables</th>
<th>Codes with Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Woman’s education</td>
<td>0 = Illiterate</td>
</tr>
<tr>
<td></td>
<td>1 = Primary school</td>
</tr>
<tr>
<td></td>
<td>2 = Middle school and above</td>
</tr>
<tr>
<td>Woman’s age (in years)</td>
<td>0 = 13-24</td>
</tr>
<tr>
<td></td>
<td>1 = 25-34</td>
</tr>
<tr>
<td></td>
<td>2 = 35-49</td>
</tr>
<tr>
<td>Caste</td>
<td>0 = Others</td>
</tr>
<tr>
<td></td>
<td>1 = SC/ST</td>
</tr>
<tr>
<td>Religion</td>
<td>0 = Hindu</td>
</tr>
<tr>
<td></td>
<td>1 = Muslim</td>
</tr>
<tr>
<td>Residence</td>
<td>0 = Urban</td>
</tr>
<tr>
<td></td>
<td>1 = Rural</td>
</tr>
<tr>
<td>Childhood place of residence</td>
<td>1 = City/Town</td>
</tr>
<tr>
<td></td>
<td>2 = Village</td>
</tr>
<tr>
<td>Husband’s education</td>
<td>0 = Illiterate</td>
</tr>
<tr>
<td></td>
<td>1 = Primary school</td>
</tr>
<tr>
<td></td>
<td>2 = Middle school and above</td>
</tr>
<tr>
<td></td>
<td>3 = High school and above</td>
</tr>
<tr>
<td>Husband’s occupation</td>
<td>0 = Agriculture</td>
</tr>
<tr>
<td></td>
<td>1 = Lower Non-Agriculture</td>
</tr>
<tr>
<td></td>
<td>2 = Higher Non-Agriculture</td>
</tr>
<tr>
<td>Standard of living index (SLI)</td>
<td>0 = Low (scores 0-9)</td>
</tr>
<tr>
<td></td>
<td>1 = Medium (scores 10-19)</td>
</tr>
<tr>
<td></td>
<td>2 = High (scores 20 &amp; above)</td>
</tr>
</tbody>
</table>

Cont...
<table>
<thead>
<tr>
<th>Explanatory Variables</th>
<th>Codes with Categories</th>
</tr>
</thead>
</table>
| Sex composition of living children| 0 = No child  
  One child  
  1 = One daughter  
  2 = One son  
  Two children  
  3 = Two daughters  
  4 = Two sons  
  5 = One daughter and one son  
  Three or more children  
  6 = All daughters  
  7 = All sons  
  8 = Son<daughter  
  9 = Son>daughter  
  10= Son equal to daughter |
| Exposure to mass media            | 0 = Neither  
  1 = Radio only  
  2 = Television and Radio |
| Discussion of FP with husband     | 0 = Never  
  1 = Once or Twice  
  2 = More Often |

The inclusion of these explanatory variables have been done based on their importance in influencing the education–fertility relationship (for details see the section on Theoretical Framework in this Chapter)

**Explanatory Variables Generated for Analysis**

For a meaningful analysis of data, some of the following variables have been generated with the relevant data available in the NFHS.

1) **Husband's Education**: The information on husband’s education was available in completed years of schooling (Q.604). From this, husbands who have not received any education are coded as illiterate, 1-5 years of education as primary, 6-9 years as middle and 10 and above years of education as high school and above (Table 3.1).

2) **Husband's Occupation**: Information on husband's occupation was coded (Q.608) in the NHFS data files as:
From these individual occupations, three categories have been formed (Table 3.1). They are:

Agriculture : item g
Lower Non-Agriculture : item d-f, h, and i
Higher Non-Agriculture : item a and c.

Those who do not fall into these categories i.e., items j - o have been excluded for the purpose of analysis (466 cases).

3) **Standard of Living Index (SLI):** In the NFHS information on income of women or family was not collected. However, detailed data on housing conditions and ownership of certain assets were obtained. Therefore, it was felt appropriate to create a Standard of Living Index (SLI) on the lines suggested by Roy and Jayachandran (1996).

The details of items used for computing SLI and the weights (scores) assigned to each of the individual items are provided in the Appendix 1. The sum of all weights for each individual (woman) indicates her position in standard of living in terms of possession of household durable goods and household amenities. Based on these total scores, the women were classified into three categories of SLI as shown in Table 3.1.

4) **Sex-Composition of Living Children:** Based on the information available in the NFHS on total number of living sons and total number of living daughters (Q.203A, Q.203B, Q.205A and Q.205.B), the variable sex composition of living children has been generated as shown in Table 3.1.
5) **Exposure to Mass Media:** In the NFHS, information about the respondents’ exposure to mass media (Q.352A and Q.352B) was collected as given below:
Heard about Family Planing message on Radio (1 = yes, 2 = no)
Heard about Family Planing message on T.V. (1 = yes, 2 = no)

Combining these responses, a single variable viz., exposure to mass media, has been created. The categories with their codes of this variable are provided in Table 3.1.

**Analysis of Data**

Firstly, the gross differentials in age at marriage, children ever born, age at first birth, desire for additional children and use of contraception across the educational level of woman are examined through bivariate analysis. Later, in order to examine the net effects of educational attainment of spouses on these dependent variables, while controlling for other explanatory variables (see Table 3.1), different multivariate techniques have been adopted as given below:

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Technique Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children ever born</td>
<td>Multiple classification analysis</td>
</tr>
<tr>
<td>Age at effective marriage</td>
<td>Multiple classification analysis</td>
</tr>
<tr>
<td>Age at first birth</td>
<td>Multiple classification analysis</td>
</tr>
<tr>
<td>Desire: for additional children</td>
<td>Logistic regression</td>
</tr>
<tr>
<td>Contraceptive use</td>
<td>Logistic regression</td>
</tr>
<tr>
<td>Birth intervals</td>
<td>Life table and proportional hazards</td>
</tr>
<tr>
<td>Progression to next parity</td>
<td>Period parity progression ratios</td>
</tr>
</tbody>
</table>

The details of the statistical techniques used for the analysis of data are provided in the following lines.

**Multiple Classification Analysis (MCA)**

Multiple classification analysis is a special case of multiple regression with categorised explanatory variables. The predictor variables can be nominal, and a non-linear relationship between any predictor and dependent variables is permissible. MCA model has the grand mean as its constant term and main effects, or a set of additive coefficients for each predictor (for details see Andrews, et al., 1973). Using the MCA, the relative
influences of education of women and their husbands on age at first birth and children ever born have been assessed. In both cases (age at first birth and children ever born), at first stage only socio-economic variables viz., education level of women and their husbands, caste, religion, residence, husband's occupation, standard of living index are inducted in the model. Later, in addition to socio-economic variables, age at effective marriage is introduced as explanatory variable.

The net effect of education of the woman on the age at effective marriage has been evaluated with the help of MCA controlling for other factors (religion, caste and childhood place of residence). It has been carried out for five birth cohorts (1943-52, 1953-57, 1958-62 and, 1963-67, 1968-72) as well as marriage cohorts (1967-71, 1972-76, 1977-81, 1982-86 and 1982-91) of approximately equal size. The details regarding the choosing of respective cohorts are given in Chapter VI.

**Logistic Regression Analysis**

Owing to the dichotomous nature of the dependent variables contraceptive use and desired fertility (whether another child was desired), logistic regression has been performed. The logistic regression technique can be used not only to identify the risk factor but also to predict the probability of success. This technique expresses a qualitative dependent variable as a function of several independent variables, both qualitative and quantitative (Kendall, 1975; Fox, 1984).

If \( p \) is the probability of an event (for example, using of contraception or not desiring another child), then \( p \) is modelled as:

\[
\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \sum_{j=0}^{k} \beta_j x_j
\]

Where \( \beta \) is vector of the unknown coefficients and \( x \) is vector of the covariates that affect the probability. The model expresses the log odds of the event as a linear function of the independent variables.
Computation of Period Parity Progression Ratios

Parity progression ratio (PPR) is the proportion of women who move from a given parity to the next during their lifetime. The PPRs are useful to study the family building process and can be linked to the TFR. Generally, these are computed by cohorts of women or cohorts of births. However, in such analyses truncation caused by the survey makes it difficult to study recent changes. The period parity progression ratio (PPPR) methodology overcomes this problem by computing progression for synthetic cohorts (Feeney and Yu, 1987; Ni Bhrolchain, 1987). Suppose a group of women have their kth birth in calendar year 't'. Let the proportion that have the next, that is (k+1)th, birth in the same year be denoted by \( q_E(t) \). Of the remaining women, proportion who have the (k+1)th birth in year t+1 be denoted by \( q_0(t+1) \). Of the remaining women, proportion who have the (k+1)th birth in year t+2 be denoted by \( q_1(t+2) \), and so on. Thus \( q_E(t), q_0(t+1), q_1(t+2), \ldots \) give a series of duration specific PPRs for the cohort who had the kth birth in year 't'. If duration specific PPRs can be computed for various birth cohorts, a matrix \( \{q_i(u)\} \) where 'i' stands for duration \( (i = E,0,1,2,\ldots) \) and 'u' for calendar year, can be obtained. From these, synthetic PPR, called Period Parity Progression Ratio (PPPR) from kth to (k+1)th parity for various calendar years can be computed. The period parity progression ratio from kth to (k+1)th parity for calendar year 't', is computed as:

\[
1 - \{(1-q_E(t)) (1-q_0(t)) (1-q_1(t)) (1-q_2(t))\ldots\}.
\]

Since the chances of having a succeeding birth after a long time, say 10 years, are remote, values of \( q_i(t) \) became very small and nearly 0 for \( i>10 \). As a result, the number of terms in the product is not large (For details, see Feeney and Yu, 1987).

In the present study, the PPPRs have been computed for progressions 0-1, 1-2, 2-3, 3-4,\ldots, and 7-8 for single year periods starting from 1972 to 1991. However, in the case of educational attainment categories, progression ratios are computed only for 0-1, 1-2, 2-3, and 3-4. Further, such PPPRs are computed for five year periods i.e., 1972-76, 1977-81, 1982-86, and 1987-91, since a single year would be too short to provide a smooth series. The fertility histories from the NFHS provide the required data for the computation. The records have been used without employing sample weights. An examination of the NFHS data files revealed that the calendar year of birth was reported for almost all the births.

*This is because the SPSS software did not have provision for it. Besides, a major portion of computations was done manually. Though use of weights could have some influence on the results, the latter analysis of birth intervals in which weights were used yielded similar results.
In general, PPPRs ought to be obtained from all women including those never married. However, in India, where fertility outside the marriage is negligible, the progression from first parity onwards obtained for ever married women would not be different from all women. Therefore, the PPPRs have been obtained from the NFHS data files for ever married women. However, the progression to first parity is really progression from woman's birth to her first birth. This requires data on all women, including never married. However, in the NFHS fertility histories are obtained only from ever married women. Therefore, the proportions of women never married were computed by single years of age from individual records in household data files of the NFHS. The numbers of women, given in the ever married women's file, were then adjusted to account for never married women. This adjustment was made by single years of age.

Often, the progression from woman's birth to first birth is decomposed into progression from woman's birth to marriage and from marriage to first birth. This requires accurate information on the date of marriage. But the NFHS data files provide only the age at marriage, and not the calendar year of marriage. As a result, the PPPRs could not computed for woman's birth to marriage and then to first birth. Therefore, only the PPPR from woman's birth to first birth has been computed.

Birth Interval Analysis

(a) Life table analysis

The interval between ith and (i+1)th birth is called (i+1)th closed birth interval. However, if (i+1)th birth has not taken place by the date of the survey, the interval between survey date and ith birth is called the open birth interval. It is known that an analysis of closed birth intervals, though useful to examine the tempo of childbearing, does not reflect parity progression. A life table that can be constructed by pooling closed and open birth intervals (treating open birth interval as censored observation), is a better way of examining the family building process (Srinivasan, 1980; Rodriguez and Hobcraft, 1980). Therefore, in the present study this approach has been adopted (for details methodology of life table see Lee, 1993; Rodriguez and Hobcraft, 1980).

*All the PPPR analysis is based on the data for women surviving at survey, since this was a retrospective enquiry. Thus, the risk of woman's death is not taken into account, as is the case in all such analyses. The PPR (0-1) is simply from parity 0 to parity 1 for all women, hence the adjustment for never married women. For later progressions, this is not needed since fertility is presumed to be within marriage.
Proportional Hazards Model

In order to see the relative effects of educational level of woman and her husband on birth intervals, the proportional hazards model (Cox, 1972) has been employed. This technique makes it possible to calculate the risk of having a subsequent birth in a particular group, relative to that of a reference group, whilst controlling for other pertinent variables. Like the standard life table, it is assumed that there is hazard (or risk) at each duration 't', of the occurrence of the end-point event (a birth). The hazard function is the product of underlying duration-dependent risk $\lambda_0(t)$ and covariates ($z$) expressed as $\exp(\beta z)$. It is assumed that the duration specific rates or risks for a given individual’s characteristic are proportional. This is defined as:

$$\lambda(Z; t) = \lambda_0(t) \exp(\beta' Z)$$

$$= \lambda_0(t) \exp\left\{\sum_{j=0}^{\infty} \beta_j Z_j\right\}$$

Where $\lambda(Z; t)$ is the hazard of failure for an individual with covariate $z$ at time $t$. $\lambda_0(t)$ is the unspecified baseline hazard when $z = 0$, called reference group. $\beta$ is a column vector of unknown parameters to be estimated in the model.

The term $\exp(\beta z)$ is the relative hazard function or relative risk associated with having the characteristic $z$. Therefore, the hazard function enables one to estimate the relative risks of other groups in relation to the baseline group (reference group). When there is no covariate present in the model, then $\exp(\beta z)$ is unity. Values greater than unity indicate that the relative risk of having birth is greater for that group compared with reference group (for more details about this methodology, see Khan and Raeside, 1998).

The sample weights have been used in the computation of mean closed birth intervals. However, the life tables and proportional hazard analysis is based on the unweighted sample records. In order to avoid complications caused by marriage dissolution, the analysis has been restricted to currently married women who have been married only once (continuously married).
Though the size of the sample used here is large, the number of births for women who are educated up to high school and above becomes too small even at birth order 4. Therefore, education of wife is categorised into three categories - illiterate, primary school and middle school and above. However, in the case of husband's education, four categories are used; illiterate, middle school, and high school and above.

In order to see whether and how the differentials have changed over time, the intervals are analysed by the time period of the date of the beginning of the interval. Three periods are used; 1972-76, 1977-81, 1982-86. For a meaningful analysis, intervals beginning after 1986 are not used because many would be too short (last birth being too close to survey date).

Quality of Data

In any survey or Census, ages are poorly reported, particularly heaping on ages 0 and 5, and typical pattern of heaping on ages 8, 10, and 12. However, the NFHS age data are of considerably better quality than any other sources. In the NFHS for UP, age was missing for only 24 persons out of total 63,019 persons listed in the household schedule. By and large, the data are mostly complete for month and year of birth of child and respondent, age at first union, and women's education. The information on distribution of births by calendar year is complete and accurate. Overall, 99 percent of living children listed in birth histories had complete birth dates recorded. The percentage of missing information for birth taken place (in months) in last 15 years was only 1.14. Out of 11,438 ever married women, the percentage of missing information regarding their education was 0.02. For age at first union, it was only 0.03 percent.

Limitations

It is recognised that data collected in large surveys do have certain limitations. In particular, fertility histories obtained from populations with a low level of education suffer from inaccurate dating of events or omission of births especially of deceased children. This is more likely to be the case in a population in which registration of births is not rigorously done. Some researchers have found that the errors in dating of events could introduce a substantial bias in the estimation of change in fertility (Brass, 1971; Potter, 1977).
For instance, if women have a tendency to under-report the ages of very young children (and thus push births towards the date of survey), the birth-history data would provide an overestimate of the fertility in the most recent period and an under-estimate in the earlier period. Similarly, if displacement of births is away from the date of survey, recent fertility could be under-estimated. The shorter the period chosen, the greater would be the bias. Further, omission of births (of deceased children) could distort estimate of fertility and measures like birth intervals and parity progression ratios.

To the extent that the NFHS data on births suffer from displacement, the trends estimated from birth histories would also be affected. However, if the nature of displacement does not vary much by education classes, and the extent of the displacement is small, trends in differentials will not be much in error.

The NFHS, a major survey, conducted in 1990s had the benefit of the experience of the World Fertility Survey and Demographic Health Survey, since the instruments and procedures were based on these. The International Institute for Population Sciences (IIPS) was designated as a nodal agency for co-ordination and technical guidance to the NFHS, with technical assistance from the East–West Centre/ Macro International. The data collection for the NFHS was undertaken by various Consulting Organisations (COs) in collaboration with the Population Research Centre (PRCs) in each state. The representatives of each COs and PRCs were trained in a series of workshops organised by IIPS and the supervisors and investigators were given extensive field training. Thus, the quality of the data is expected to be quite good. However, the possibility of errors of the kind mentioned above can not be ruled out. To what extent such errors will distort the results will be discussed when the main findings are presented.