WORLD FERTILITY SURVEY

TECHNICAL BULLETINS

MAY 1980

No. 4/TECH. 1407

Basic Fertility Measures from Retrospective Birth Histories

VIJAY VERMA

INTERNATIONAL STATISTICAL INSTITUTE
Permanent Office · Director: E. Lunenberg
Prinses Beatrixlaan 428
Voorburg, The Hague,
Netherlands

WORLD FERTILITY SURVEY
Project Director:
Sir Maurice Kendall, Sc. D., F.B.A.
35-37 Grosvenor Gardens
London SW1W OBS, U.K.

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

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

This paper is one of a series of Technical Bulletins recommended by the WFS Technical Advisory Committee to supplement the document *Strategies for the Analysis of WFS Data* and which deal with specific methodological problems of analysis beyond the Country Report No. 1. Their circulation is restricted to people involved in the analysis of WFS data, to the WFS depositary libraries and to certain other libraries. Further information and a list of these libraries may be obtained by writing to the Information Office, International Statistical Institute, 428 Prinses Beatrixlaan, Voorburg, The Hague, Netherlands.

BASIC FERTILITY MEASURES FROM RETROSPECTIVE BIRTH HISTORIES

WFS/TECH.1407

By: VIJAY VERMA
WFS Central Staff
International Statistical Institute
35-37 Grosvenor Gardens
LONDON SW1W OBS, U.K.

CONTENTS

ACKNOWLEDGEMENTS

Several members of the WFS Professional Staff provided helpful suggestions and comments during the preparation of this document, among them Rod Little, John Cleland and Alan Sunter. Thanks are also due to Norman Ryder and Allen Hill for reviewing the document.

ABSTRACT

This Bulletin describes fertility rates of various types and other basic fertility measures which can be constructed from retrospective birth and marriage histories of the kind collected in WFS surveys. Age-specific, marriage duration-specific and parity-specific rates, on both a cohort as well as a period basis, are defined for several definitions of exposure to the risk of child-bearing. Birth intervals are also considered briefly. Assuming all necessary data to be available down to the level of the month, full computational details and numerical examples for the construction of the various direct measures of fertility are provided. Indirect estimation procedures employing extraneous data are not included in this document.

LIST OF MAIN SYMBOLS

a	retrospective age of woman
$b(a,p)$	aggregated number of births (at age a during period p, for example)
B	date of birth of woman (century-month)
$B(i)$	date of birth of i^{th} child (century-month)
c	birth cohort (index identifying calendar-year of birth or current age of woman)
d	retrospective duration since first marriage
$D(j)$	date of dissolution of j^{th} marriage (century-month)
$e(a,p)$	aggregated years of exposure (at age a during period p, for example)
e_0, e_1	months of exposure at specified ages (durations) to an individual woman
I	date of interview (century-month)
m	marriage cohort (index identifying calendar-year of marriage or current marriage duration)
$M(j)$	date of beginning of j^{th} marriage (century-month)
$n(c)$	number of women in cohort c
p	period (an index identifying a calendar-year or completed years before interview)
$P(i)$	date of i^{th} "fertile pregnancy" (century-month)
$r(a,p)$	fertility rate (for example age-period specific, $b(a,p)/e(a,p)$.)
$r_i(c,a)$	fertility rate by birth order (cohort-age specific, for example)
$\tilde{r}_i$	parity i specific rate
$s(c,p)$	mean cumulative fertility (for example, of cohort c by period p; $=\sum_p r(c,p)$)
$\tilde{s}(c,p)$	corresponding measure for a synthetic cohort.
$s_i(c,a)$	cumulative proportions having birth of order i (for example, in cohort c by age a; $=\sum_a r_i(c,a)$).
$T(i)$	length in months of i^{th} closed birth interval, $P(i)-P(i-1)$.
TL	length in months of the last closed birth interval
$U(i)$	length in months of the open birth interval for women of parity i

SUBSCRIPTS

i	birth order
m	calendar-month
y	calendar-year

1. INTRODUCTION

1.1 SCOPE OF THE BULLETIN

The purpose of this Technical Bulletin is to describe several measures of fertility which can be constructed from retrospective birth and marriage history data of the type collected in the WFS Individual Questionnaire[1], and to indicate in rather specific terms just how the calculations will be performed.

The scope of the document does not include the indirect estimation of measures, for which data from a vital statistics system or longitudinal survey are substantially more appropriate. Specifically, we are not concerned with estimating quantities such as rate of natural increase which are affected by mortality, nor with describing indirect estimation procedures which involve estimates of mortality levels. The direct measures of fertility described here are constructed basically by computing time intervals between pairs of events or the number of events in a specified time interval of "exposure", in retrospective histories of individual women, and then aggregating those for specified sub-populations in the sample.

In a sample survey, particularly one in a developing country, it is rarely possible to obtain complete and accurate data on dates of vital events. The lack of complete information is the most obvious problem in calculating measures of fertility. Ideally, dates in the birth history are specified as calendar-year and month of each birth. At the worst we may have cases where no information at all is available on the date of a birth; and even when reported, there can be cases where the given dates are obviously implausible in relation to each other (for example, reported births intervals smaller than the biologically possible minimum), or in relation to other events such as the woman's own birth date. Such inconsistencies also arise from coding and punching errors. This requires editing and correction of birth history data, procedures for which are complicated due to the lack of complete information. In the present document it will be assumed that all relevant dates have been edited for interval consistency, imputed where necessary, and *coded down to the level of the month**.

The problem of incomplete information and of data obviously inconsistent are relatively easy to detect and even to "correct", though the effect of imputation on the interpretation of the data are not easy to investigate. The birth history data can also suffer from other shortcomings, such as omission of births and systematic displacement in reporting of dates. These effects are more difficult to detect and can bias the levels, trends and differentials in fertility derived from retrospective data.

---

* *In most WFS surveys to date, the incidence of completely undated births has been extremely low; rather, the problem has been that of the failure to obtain, for appreciable proportions of reported births, dates down to the level of the month. Vijay Verma and Rod Little at the WFS have developed a comprehensive procedure for editing birth and marriage histories in the presence of incomplete data. The procedure is also used to assign months where not available, the objective being to provide aggregate measures and rates which are approximately unbiased -- though clearly, any specific imputed month can differ significantly from its "correct" (but unknown) value. Details of the edit and month-imputation procedure are given in the WFS Guidelines on Data Processing[2], and in the Users' Manual for the WFS Date Edit, Imputation and Recode (DEIR) computer program developed for applying the procedure to WFS data.*

This necessitates a thorough evaluation of the quality of the data before any firm conclusions can be reached regarding the prevailing patterns. While this Bulletin is not directly concerned with procedures for evaluating and adjusting birth history data, the calculation of the detailed fertility measures described here has a central role in such an evaluation. In the description to follow, the data will be taken at their face value; it is a matter of researchers' judgement whether for a given data set it is justifiable to compute all the measures described below, and even more so, whether the data are of sufficient quality to warrant a more sophisticated analysis.

Child-bearing has two components which are difficult to disentangle: quantity and timing. The final completed family size, for example, can be achieved by a wide variety of timing patterns, ranging from having all children closely spaced at early ages to having them spaced throughout the child-bearing ages. The variety of measures described below may allow a certain degree of separation between the two components. Certain measures tend to be more sensitive to the first component while others to the second component of fertility; for example births of higher orders are indicative of the quantity dimension, while inter-birth intervals are more sensitive indicators of the timing of fertility. The following description however, is not directly concerned with demographic *interpretation* of the variety of fertility measures; the objective is rather to *specify* the relevant measures as completely as possible.

Finally, it should be mentioned that the measures considered here are in the main of a descriptive type. Some of the special analytic techniques employing these measures such as life-table methods and birth interval analysis will be described in other WFS Documentation.

Given these restrictions in the scope of the Bulletin, it is intended to offer a full, though not exhaustive, list of measures and to provide computational procedures and examples in sufficient detail to ensure a common understanding by different users. As will be seen later, many of the measures can be computed on the basis of simple cross-tabulations of the data; these will be specified below. A computer program (FERTRATE) has been developed at the WFS for computation of most of the measures described in the Bulletin.

1.2 DATA AND NOTATION

The WFS Individual Questionnaire is administered to a sample of women in the child-bearing ages, and obtains data on two sequences of events for each respondent:

1) A maternity or birth history, eliciting the date of occurrence of each live birth (or pregnancy), and data on sex, survivorship status and age at death (if applicable) of each child; and

2) a union or marriage history eliciting effective dates of beginning and termination and the outcome of each period of sexual union.

In some of the WFS surveys, the maternity history has been recorded in the form of a single sequence encompassing all pregnancies irrespective of the outcome (the so called 'integrated pregnancy history' approach). However, in view of the difficulty in obtaining accurate dates in many developing countries, the main sequence in many surveys has been confined only to *live births*. Reference is made only to live births in the measures described below.

In the most elaborate form, the union history identifies dates of all sexual unions, distinguishing each union by its type (distinguishing, for example, legal marriages from common-law unions and 'visiting relationships'), and recording substantial periods of temporary separation within unions. In the less elaborate and more usual form, the sequence identifies periods spent within marriage, irrespective of union type and ignoring periods of temporary separation within marriage. In the extreme case, the data obtained or utilised in the 'marriage history' may be confined to a single event, namely the date of entry into *first* marriage.

Given the assumption that all relevant dates are available (or have been imputed) to the level of the month, a particularly convenient form of coding the date of occurrence of an event is to record the number of months elapsed since an arbitrarily fixed point in time. When the reference point is defined as the beginning of the current century (say in the Western Calendar), we obtain the *century-month-code* of the event. By definition, the century-month-code for January 1900 is '1', and, for example, for March 1950 it is (12x50 + 3) = 603.

Figure 1. Retrospective Birth and Marriage History of an Individual Woman

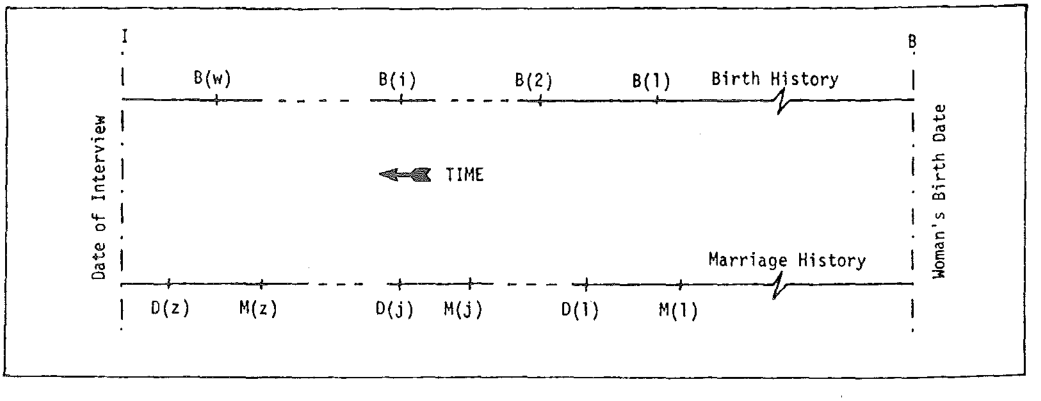


Figure 1 illustrates the birth and marriage history of a woman plotted on a straight line representing time. It also introduces the notation we will use. B is the date of birth of the woman, and I is the date of interview; a symbol such as B is used to refer to an event itself, as well as to the date (century-month code) of occurrence of the event. Dates are coded to the nearest month, and, for example, the numerical difference $(I-B)$ gives the rounded current age in months of the women. $B(1)$ to $B(w)$ are the dates of the woman's live births, with w as the number of children ever born (parity). Each birth is identified by its birth order, i, which is defined as the "numerical order (ie., first, second, third etc.) of a live born child in relation to all the previous live born children of the mother; where more than one child is born at the same confinement, each will be given a separate birth order".[3]

For certain measures reference may also be made to a projected event: the expected date of termination of a current pregnancy (CP), defined for women pregnant at the time of the interview.

In the marriage history, defined only for ever-married women, $M(1)$ is the date of entry into the first union, and for the jth marriage, $M(j)$ is the date of beginning and $D(j)$ the date of dissolution. For a woman in a union at the time of the interview, $D(z)$ is not defined, where z is the number of marriages.

With the data described above, certain fertility measures at the level of the individual woman may be constructed, for example: the number of children born within a specified age, marriage duration or calendar-period; the length of interval between births of specified order etc. An increasing amount of fertility research takes the individual woman as the unit of analysis and applies statistical techniques such as multiple regression to individual level responses. However, various fertility measures will be presented here in the form of *aggregate measures* defined for specified sub-populations of women. Since all standard aggregate measures of fertility are built-up from individual level data, the correspondence between the two levels of analysis is an immediately obvious one.

The basic fertility measures may be divided into two types:

1) Fertility rates, defined as the ratio of live births to women's intervals of exposure to child-bearing: the numerator consists of the number of live births during (say) a specified period to an aggregate of 'exposed' women, and the denominator consists of the total interval of exposure during the same period for these women. A variety of rates can be constructed corresponding to the different definitions of exposure, different periods considered, and the specific categories of women and live births included.

2) Birth intervals, which in the general case refer to the time elapsed between two events in the birth and marriage histories, at least one of which is a live birth.

1.3 FERTILITY RATES

Fertility rates are defined basically by "slicing" the birth and marriage histories of individual women by time intervals measured from certain specified points, and taking -- for aggregates of women -- the ratio of births to the length of exposure in each time interval. The time intervals may refer to historical locations, or to locations in individual women's life-cycles. Slicing in terms of fixed time periods (such as specified calendar-years, or intervals defined in relation to the date of interview) provides "period-specific" rates; intervals measured from women's birth date provide "age-specific" rates; and those from the date of (first) marriage give "duration-specific" rates. More than one of these controls may be applied simultaneously, for example to obtain age-period specific rates. Also, rates from one period or age to the next may be added together to provide measures of cumulative fertility.

Further, at the aggregate level women may be grouped according to their birth dates (or current age) or according to their dates of marriage (or current marriage duration) to obtain rates specific to birth or marriage cohorts.

For certain applications the numerator may be restricted to births of a particular sex, survivorship status or birth order; such detailed classification is a useful tool in investigating the quality of the birth history data. Fertility rates decomposed by birth order can be particularly useful in elucidating the pattern and trends in fertility.

Exposure (the denominator) may be measured in a number of ways. The crudest measure is the unconditional time elapsed, or 'unrestricted exposure'; in this case the numerator should include all births - including those to women not married by the time of the survey. At the next level, exposure may be defined as the total time elapsed following first marriage; in this case the numerator will exclude pre-marital births. Next, for marital fertility rates, exposure will be measured by the time spent within unions, with the numerator restricted to marital births. When the data are of sufficient quality, one may exclude from exposure periods of temporary separation within unions, as well as sterile intervals; or one may distinguish intervals of exposure by type of union. The main principle in defining intervals of exposure is that *if there is a restriction on the base interval, then a corresponding restriction should be placed on births included in the numerator.* Where this cannot be achieved exactly due to limitations of the available data, an attempt should be made to keep the incongruence as small as possible.

1.4 BIRTH INTERVALS

Live birth intervals should be distinguished from pregnancy intervals, the latter being defined in terms of any category of pregnancy termination, including pregnancies not resulting in live births. While it may sometimes be more appropriate to use pregnancy intervals as opposed to birth intervals, the available survey information on wasted pregnancies is frequently of poor quality. The measures recommended here refer only to intervals between live births (strictly speaking, between separate pregnancies resulting in live births).

Periods when the woman was not menstruating regularly or was not engaging in sexual intercourse are sometimes excluded from the total interval length to obtain "net" interval. In most WFS surveys, however, accurate data on periods of non-exposure to be excluded are not available, and the preferred practice has been to compute intervals simply as the total duration between births (or other events) defining the interval.

While in general terms a birth interval may be defined as the time elapsed between a birth and some other event (another birth of whatever order, an event from the marriage history, or some arbitrary point in time such as the date of interview), it is useful to distinguish various types of intervals defined more specifically. The interval between first marriage and first birth is called the *first birth interval.* The interval between one live birth and the next is called an inter-birth *interval;* these are designated according to the order of the birth which terminates the interval. The *last closed interval* is that between the most recent two live births; sometimes this definition is extended to include the first marriage and/or the expected termination of a current pregnancy as valid events for defining the interval. Finally, the *open interval* is the time elapsed since the last live birth.

The basic fertility measures considered in this document are those describing the distribution of interval lengths in terms of statistics such as the mean, median, variance etc. In view of the fact that a cross-sectional survey captures only an incomplete, hence somewhat biased, selection of the total life-experience of women, we will also consider briefly a simple life-table procedure for estimating this distribution. More elaborate analysis of the birth-interval data is a separate area of study in its own right.

2. AGE-SPECIFIC FERTILITY RATES FOR UNRESTRICTED EXPOSURE

2.1 DESCRIPTION OF THE MEASURES

By 'unrestricted exposure' is meant that the length of exposure to child-bearing is taken simply as the total time elapsed, irrespective of details of the marriage history. The most commonly used type of these rates is the conventional Age-Specific Fertility Rate (ASFR) defined as the "fertility rate with the number of live births during a given year born to women of a given age (or age group) as the numerator, and the number of person-years lived by that age (or age group) of women during the year as the denominator"[3]. In this section we will provide a more general formulation of age-specific rates for unrestricted exposure. Generally speaking, the measures described in this section are appropriate to a universe of all women, irrespective of current marital status. In Section 2.3 we will comment on samples confined to ever-married women.

Births occurring to a specified aggregate of women may be classified in terms of one or more of the following demographic controls:

1) When the birth occurred, ie, *period* of occurrence;
2) *age* of mother at birth of the child; and
3) current age or *cohort* of the mother.

Age is measured in completed years (single or grouped) at the time of the survey; the point of reference in general varies from one woman to another in the sample. By contrast, 'cohort' is customarily used to refer to a group of women born during the same calendar-year(s), and 'period' to refer to specific calendar-year(s) of occurrence. In the present context it is convenient to extend these terms to refer also to corresponding quantities defined in reference to the date of interview: 'cohorts' to groups of women in the same age range at the time of the interview; and 'periods' to specified durations before the interview. Reference to calendar years is more relevant to registration data, while that to the date of interview is more apposite to retrospective birth history data from surveys.

Specifically, three types of rates may be defined by pairing the three controls listed above. Births classified by period of occurrence and age of mother at child's birth will constitute the numerator in computing *age-period specific* rates (the conventional ASFRs); similarly, births may be classified in terms of mother's cohort and period of occurrence to compute *cohort-period specific* rates; or in terms of mother's cohort and her age at child's birth to compute *cohort-age specific* rates. The denominator in each case is the total number of person-years lived, classified in the same way as the numerator.

There is an overlapping redundancy in the three sets of rates since the basic classification variables satisfy the relation:

$$c = a + p \tag{1}$$

Where c ('cohort') refers to the time elapsed since the woman's birth, a ('age') to the time from her birth to the birth of the child, and p ('period') to the time elapsed since the birth of the child. The above relationship holds exactly only when these time intervals are measured exactly. In practice, of course, they are defined in terms of single years or in 5-year groups, and the above relationship is only an approximate one. For example if cohorts are defined as five-year groups by current age or calendar years of birth, and ages at child-bearing and periods of occurrence also defined in groups of five years, then births to women of a specified cohort at specified age-group will span a period of *ten* rather than five years. Similarly, births during a specified period at specified age-group are contributed by two separate (adjacent) cohorts of women.

DEFINITION OF COHORTS AND PERIODS

Noting that 'cohorts' and 'periods' may be defined either in terms of specific calendar-years or in reference of the date of interview, we will give full computational details for the following two schemes:

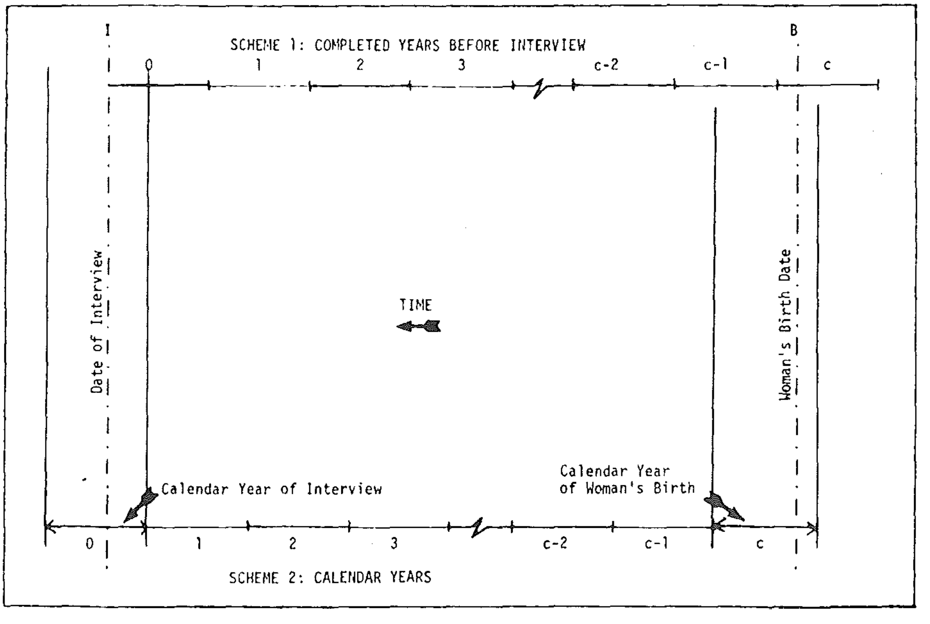
Scheme 1. Cohorts defined in terms of the woman's age at the time of interview, and periods defined as completed years before the interview.

Scheme 2. Cohorts defined as groups of women born during specified calendar-years, and periods defined as calendar-years.

Scheme 1 makes full use of the most recent data in retrospective birth histories and is recommended in the present context. However, Scheme 2 is also employed frequently, partly because of convention, but also for being more convenient for comparison with external data such as from vital registration or other surveys.

It is important to clarify the two schemes of classification since frequent reference will be made to these in the following sections. The schemes are illustrated in Figure 2 on the time-axis. As before, *B* is the date of birth of a woman and *I* the date of her interview. In Scheme 1 (frequently employed in WFS First Country Report[4]) periods are measured in single completed years before the interview; we number these sequentially backwards starting with '0' (meaning up to 12 months before the interview). A child born during period *p* (and still surviving) will be aged *p* completed years at the time of the interview. Similarly the period of a woman's own birth is in fact her current age, and, by definition identifies her cohort *(c)*. Note that since different women may be interviewed at different times, the actual time covered by the same period in Scheme 1 may not exactly coincide for different women.

Figure 2. Two Schemes for Defining Periods and Cohorts



In Scheme 2 periods refer to fixed calendar-years, which are the same for all-women in the sample. Cohorts refer to women born during the same calendar year(s). To make the computational details for the two schemes formally similar, we will number 'periods' (p) in Scheme 2 sequentially *backwards* starting with *'0'* for the year of interview; the index p is related to its corresponding calendar-year Y as follows:

$$p = I_y - Y,$$

Where I_y is the calendar-year of interview.

(In either scheme, a woman's cohort, a period and her age during that period are related according to equation (2) given later in the section).

Consider, for example, a woman born in March 1943 who is interviewed in July 1980. Under Scheme 1, yearly periods are defined as follows:

$p = 0$: July 1979 to June 1980 (inclusive)*

$p = 1$: July 1978 to June 1979

⋮

$p = 37$: July 1942 to June 1943

---

* *Since all dates are assumed coded to the level of the month, the exact date of an event within a given month is ambiguous. This ambiguity can be reduced by assuming the interview to be held at the beginning of the month and rejecting all events in the month of interview itself. This practice will be followed throughout. All other events will be assumed to occur on the average at the middle of the month.*

The woman's cohort (current age) is c = 37. During any period, say p = 15 (completed years before interview), she passes through two ages

a_1 = (37-15)-1 = 21, from July 1, 1964 to mid-March 1965; and

a_0 = (37-15)=22, from mid-March to June 30, 1965

Under Scheme 2, we define periods as:

p = 0 : 1980 (calendar year of interview)

p = 1 : 1980-1 = 1979

⋮

p = 37 : 1980-37 = 1943 (the year of woman's birth).

The woman's cohort is c = 37. During any period, say p = 15 (calendar-year 1980-15 = 1965), she passes through the same two ages a_1 = 21 (from January 1 to mid-March) and a_0= 22 (from mid-March to December 31).**

LEXIS DIAGRAM

A pictorial representation of aggregate data useful for clarification of the basic concepts involved in computing fertility rates is the *Laxis diagram.* The diagram displays the three inter-related dimensions - cohort, age and period - simultaneously, and can be constructed in various forms. In Figure 3 we present a form closely related to a convenient cross-tabulation of the data for computing the various fertility rates. Figure 4 illustrates the corresponding cross-tabulation.

** *The numerical value of c in the two schemes will differ by 1 for the same individual if her month of birth coincides with or is after the month of interview.*

FIGURE 3: LEXIS DIAGRAM Showing Cohorts, periods and Retrospective ages.

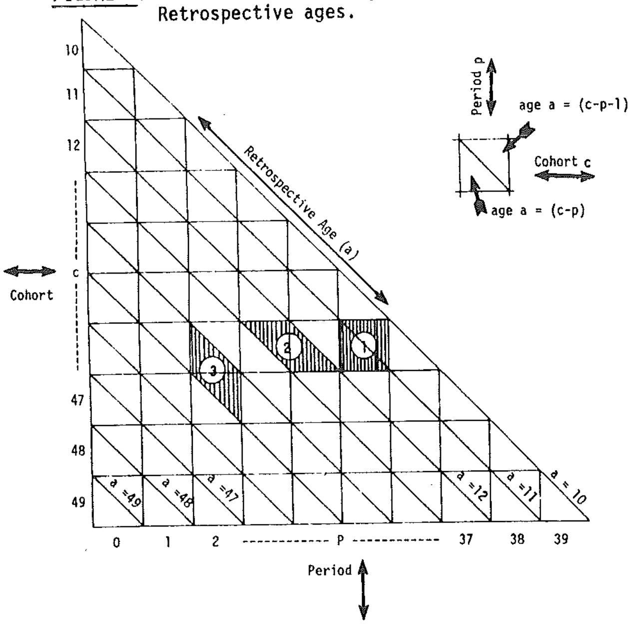
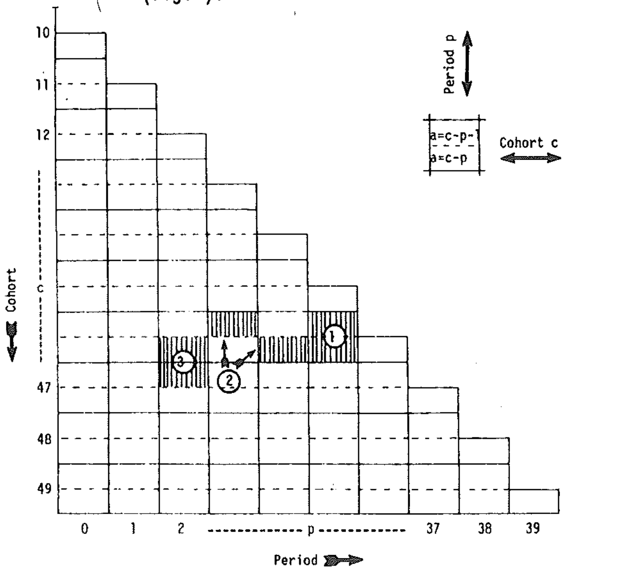


FIGURE 4: Cross-Tabulation Equivalent to Lexis Diagram - (fig.3).



Rows in Figure 3 represent fixed cohorts for women currently aged 10-49; columns represent fixed periods, 0-39 completed years before the interview. The life-experience of an individual woman is represented by a horizontal line (running from right to left), the vertical position of the line depending upon the exact birth date, with older women lower down in the diagram. The intersection of these 'life lines' with diagonals (top left to bottom right) indicates points at which women attain specified ages; the diagonal represent fixed ages.

The birth and exposure histories may be classified unambiguously by any two of the three dimensions: cohort, retrospective age and period of occurrence, giving three types of rates. Considered together, the three dimensions divide Figure 3 into triangles. Identifying a rectangle of type '1' in the figure by its coordinates (c,p), we note that it consists of two triangles: an upper triangle, identified say as $(c,p)^*$, in which the woman's age a is related to c and p as:

$$a = (c-p)-1$$

and a lower triangle, $(c,p)_*$, in which the woman is aged (2)

$$a = (c-p).$$

Whether an event falls in the upper or the lower triangle depends upon whether its month of occurrence is before or after the month of the woman's births.

The three types of rates can be identified in terms of these triangles as follows:

1. For a *cohort-period specific rate*, the numerator is the number of births $b(c,p)$ to women in cohort c during period p, ie during a rectangle of type '1' in Figure 3; the denominator is the number of person-years $e(c,p)$ lived during this period. Hence the cohort-period specific rate (for cohort c during period p) is:

$$r(c,p) = \frac{b(c,p)^{*} + b(c,p)_{*}}{e(c,p)^{*} + e(c,p)_{*}} = \frac{b(c,p)}{e(c,p)} = \frac{b(c,p)}{n(c)}. \qquad (3)$$

Since each woman in the cohort lives for exactly one year during any period of the same duration, $e(c,p)$ simply equals the number of women, say $n(c)$, in the cohort. It is customary to quote rates as births per 1,000 women-years of exposure, ie after multiplying (3) by 1,000.

2. For a *cohort-age specific rate*, the numerator is births occurring to women of cohort c, at age a, ie during a parallelogram of type '2' in Figure 3. The denominator is the number of person-years lived at this age, which for a one-year duration again equals the number of women, $n(c)$, in the cohort. In terms of the notation introduced above, the parallelogram corresponding to cohort c and age a is the sum of two triangles $(c, c-a-1)^{*}$ and $(c, c-a)_{*}$, since 'p' equals $(c-a-1)$ in the upper triangle and equals $(c-a)$ in the lower. Hence, cohort-age specific rate (for cohort c at age a) is:

$$r(c,a) = \frac{b(c,c-a-1)^{*} + b(c,c-a)_{*}}{e(c,c-a-1)^{*} + e(c,c-a)_{*}} = \frac{b(c,a)}{e(c,a)} = \frac{b(c,a)}{n(c)}. \qquad (4)$$

For notational simplification we have written the numerator of (4) as $b(c,a)$ and the denominator as $e(c,a)$, where the coordinates (c,a) refer to a parallelogram of type '2' in Figure 3.

3. For an *age-period specific rate* (ie conventional ASFR) the numerator is the total number of births at mother's age a during period p, ie during a parallelogram of type '3' in Figure 3. The denominator is the number of person-years lived during this period (which in this case does not reduce to a simple number such as $n(c)$). Hence, age-period specific rate (for age a during period p) is:

$$r(a,p) = \frac{b(a+p+1,p)^{*} + b(a+p,p)_{*}}{e(a+p+1,p)^{*} + e(a+p,p)_{*}} = \frac{b(a,p)}{e(a,p)} \tag{5}$$

since 'c' equals $(p+a+1)$ in the upper triangle, and equals $(p+a)$ in the lower. Again for simplicity we have written the numerator of (5) as $b(a,p)$ and the denominator as $e(a,p)$, where the coordinates (a,p) identify a parallelogram of type '3'.

As noted earlier, there is an overlapping redundancy in the three types of rates since the same basic information is being classified in different ways. Numerically, the difference become more significant when rates are aggregated over a number of years, for example over five-year groups by age or period. Substantively, the different forms are of interest when the rates are summed up from one age or period to another to obtain measures of cumulative fertility.

2.2 COMPUTATIONAL DETAILS

It remains to define variables c and p, as well as a (which identifies whether an event belongs to an upper or a lower triangle in Figure 3), in terms of the given data for each individual woman -- her date of birth, the dates of birth of her children, and the date of interview. From these variables we can identify a woman's contribution of births and lengths of exposure to cells of the cross-tabulation illustrated in Figure 4. Aggregated over a specified population of women, this provides us with quantities such as $[b(c,p)^*,\ b(c,p)_*]$ from which rates (3)-(5) can be computed.

Necessary computational details along with numerical examples for individual level data are given in Appendix I. Appendix II illustrates aggregate level measures such as (3)-(5).

2.3 ON THE NATURE OF THE INTERVIEWED SAMPLE

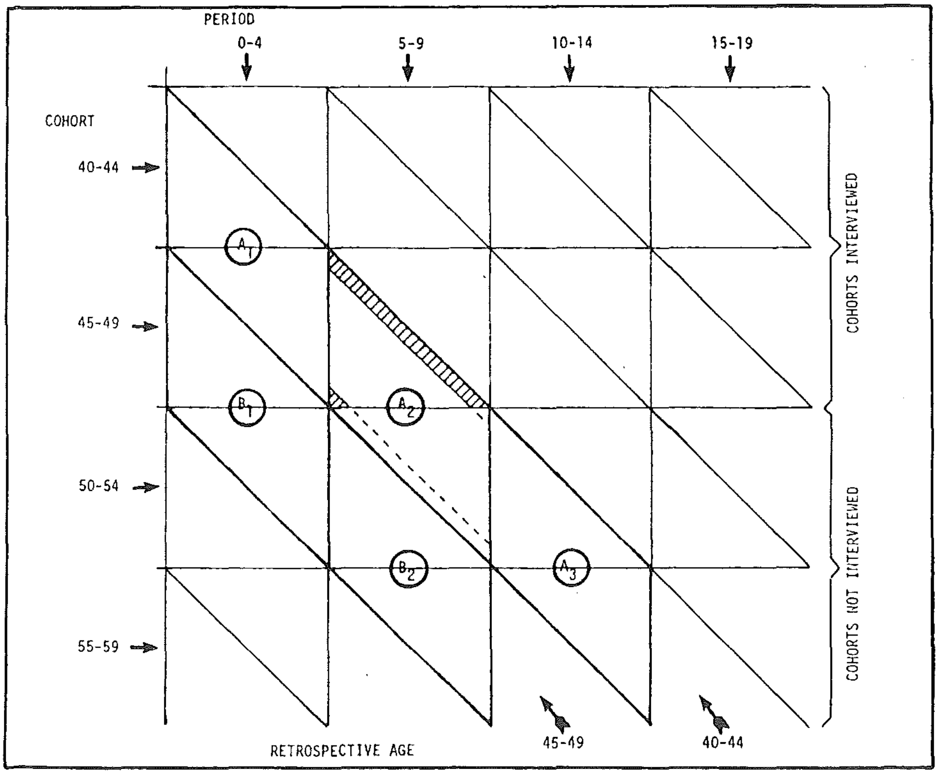
SELECTION BIAS

The sample for the individual interview is confined to women in the child-bearing ages (usually defined as 15-49), surviving at the time of interview (frequently, it is further restricted to ever-married women). As such the sample provides an incomplete representation of the total study population of women, particularly in relation to past fertility.

1. The effect of confining the sample to the child-bearing ages is most obvious when conventional age-specific fertility rates (equation (5)) are considered. Figure 5 shows a segment of the Lexis diagram (fig. 3) with data aggregated over 5-year groups. Consider fertility at ages 40-44. For the period 0-4 years before the survey, two cohorts, 40-44 and 45-49, contribute to this fertility (parallelogram A_1), and all necessary data are available from the survey. However, for the period 5-9 years before the survey only incomplete information is available for computing fertility at ages 40-44, since the 50-54 cohort is excluded from the sample (lower triangle of A_2); further, available data are biased towards exposure at younger ages -- for example, there is approximately $4\frac{1}{2}$ person-years of exposure per woman at age 40

(the longer shaded area in A_2), but only ½ years at age 44 (the smaller shaded area in A_2). If a rate based on incomplete data is required, it will be necessary to weight these data to compensate for differences in the length of exposure at different ages*. (The same is true of fertility at ages 45-49 for the period 0-4 years before the survey ie parallelogram B_1). No information is available for computing fertility at ages 40-44 for the period 10-14 years before the survey (parallelogram A_3). The data become increasingly incomplete as we proceed further back from the interview.

Figure 5. Illustration of Incomplete Data Due to Upper Age Limit for Eligibility for the Interview



* *One may, for example, compute numbers of births by single-years of age 40 to 44 and take a weighted sum, with weights inversely proportional to 4½ (for age 40), 3½ (for age 41) etc; similarly for the total length of exposure at these ages. For a similar procedure see Piampti and Knodel[6].*

Considerations similar to the above will apply if the *lower* age limit for interviewing exceeds the minimum age at child-bearing; in fact the effect in this case will be more serious as the most recent periods and younger ages will be affected.

2. An association between the level of women's fertility and their mortality can result in a bias in fertility trends and differentials when estimated from retrospective histories of women surviving at the time of the interview. However, the following illustration will show that even under strong association between fertility and mortality the resulting bias is likely to be small, particularly compared to other sampling and non-sampling errors inherent in retrospective histories based on sample surveys involving personal interviewing.

Suppose that 20 years ago, women then aged 25 consisted of two equal subgroups of the study population, the first half reproducing at *twice* the rate of the second half. Assume further that the high fertility group had a life expectation at birth of only 25 years (corresponding to life expectation at age 25 of 28.6 years: "west" model life-table[7], level 3), compared to 50 years for the low fertility group (corresponding to expectation at age 25 of 40.1 years). After 20 years the relative size of the two groups would have changed from 50:50 to around 44:56. It is the latter composition which will be reflected in the sample, while the former is the true composition 20 years ago. An over-representation of low fertility women in the surviving sample will result in an under-estimation of the past fertility of the study population, but only by around 4% even under the rather extreme conditions assumed in this example. Its effect on the estimated trend in fertility will be negligible.

A similar conclusion will be reached if we specifically considered maternal mortality (mortality resulting directly from the experience of child-birth) in its plausible association with the level of fertility. The mother's death following child-bearing means that the birth is not enumerated in the survey. An extreme difference of 20 per thousand in the maternal mortality rate between two sub-populations will introduce a differential bias of under 2% in the observed fertility levels for the two groups.

SAMPLING ERROR

It is not in place here to discuss procedures for estimating sampling errors for estimates based on complex multistage sample designs. Below we give a very approximate indication of the magnitude of the quantities involved.

For a typical WFS survey with a sample of, say, 5,000 women, there will be on the average 100-150 women at any single-year of current age. A single-year period fertility rate may be considered equivalent to the proportion of women having a birth during one year. Typically this proportion is around 0.2, and its standard error from the well known binomial formula $(pq/n)^{\frac{1}{2}}$ for $n \sim 100$ is around 0.04. In other words, standard error relative to the estimate for a single-year age-period rate is likely to be of the order of 20-25%, or even higher depending upon the increase in sampling error due to clustering of the sample.

Clearly it is necessary to aggregate data over several years. For a sample of 5,000 women, relative standard error for a single-year period rate but aggregated over five-year age groups may typically be of the order of 10%; to limit this to within 5%, it will be usually necessary to aggregate data over a period of 3-4 years.

Certain measures (such as the General or the Total Fertility Rates - see below) involve aggregation over all ages; for these, relative standard error for a sample of 5,000 women may be expected to be of the order of say 4-5% for single-year periods, and of the order of 2-3% when aggregated over a period of 2-3 years. For multistage clustered samples, the actual values of the error may be substantially higher depending on the efficiency of the sample design.

WEIGHTED SAMPLES

In the presence of departures from equal probability samples of individuals, we assume that individual contributions to the aggregated numerators and denominators are multiplied by appropriate sample weights to compensate for differences in selection probabilities. Beyond that, sample weights in no way modify the computational forms given here.

SAMPLES RESTRICTED TO EVER-MARRIED WOMEN

Fertility measures for unrestricted exposure are based on all women, irrespective of their marital status; an all women universe is assumed in the present description. However, a common arrangement in WFS surveys is to confine the detailed individual interview to ever-married women. On the basis of this interview alone, neither the fertility of never married women nor their contribution to the total person-years of exposure can be included in the computation of the rates. (Note, however, that pre-marital exposure and fertility of women who subsequently marry *are* included, at least in principle).

The household interview, which precedes the individual interview in WFS surveys, records the entire household population by age, sex and marital status, from which proportions of women ever-married by current age can be estimated. These proportions can be used to inflate appropriately the size of each ever-married cohort c to represent the entire birth cohort as follows: if $f(c)$ is the proportion ever-married among women in cohort c, then the denominators $e(c,p)^*$ and $e(c,p)_*$ in (3)-(5) are inflated by the factor $1/f(c)$ for all p. This amounts to multiplying the rates (3) and (4) by $f(c)$, while the age-period specific rate (5) becomes

$$r(a,p) = \left[b(a+p+1,p)^* + b(a+p,p)_* \right] \Big/ \left[\frac{e(a+p+1,p)^*}{f(a+p+1)} + \frac{e(a+p,p)_*}{f(a+p)} \right],$$

since two different cohorts ($c = a+p+1$ and $c = a+p$) are involved in the computation.

Note that proportions ever-married, $f(c)$, refer to current cross-sectional data, irrespective of any nuptiality trend. Also, their source can be external to, even independent of, the retrospective birth history data. The proportions estimated from a relatively small scale household survey may require smoothing, particularly when rates are to be computed for different socio-economic categories. The smoothing may be achieved by using moving averages, or by fitting a standard nuptiality schedule to the data[5], where the available sample size permits such fitting with reasonable confidence.

With the retrospective birth history data confined to a sample of ever-married women, it is usually not possible to adjust the numerator in (3)-(5) for the fertility of married women never-married by the time of the survey, particularly for their retrospective fertility. In any case the very basis for excluding never-married women from the detailed individual interview is the assumption that they do not make a significant contribution to the fertility of their cohort.

2.4 RELATED MEASURES

BIRTHS BY ORDER, SEX AND SURVIVORSHIP STATUS

In the foregoing discussion the numerator for a fertility rate consisted of all births, irrespective of the child's sex, birth order or survivorship status. It is substantively interesting, as well as useful for investigating quality of the birth history data, to compute fertility measures specific to sex and/or birth order. The numerator will then be the same array $[b(c,p)^*, b(c,p)_*]$ described earlier, but confined to births of a specified category; the denominator will be same as before, ie, the number of person-years of exposure irrespective of the particular category of births being considered (cf. parity specific rates, Section 5.1).

In a similar way we may compute proportions of children deceased for the retrospective arrays - probably classified by age at death, sex and birth order. The numerator will then be the numbers of dead children, and the denominator will be the total number of births (in the specified category by sex, birth order etc.). Such measures provide direct estimates of infant and child mortality levels; they also throw light on the completeness of reporting of child deaths in retrospective histories.

CUMULATIVE COHORT FERTILITY

Cohort-specific rates (3) and (4) for a given cohort can be cumulated across retrospective ages or periods to obtain a time-series of cumulative cohort fertility. For example, cohort fertility cumulated by age

$$s(c,a) = \sum_{a'=a_0}^{a} b(c,a')/n(c) = \sum_{a'=a_0}^{a} r(c,a'), \tag{6}$$

gives the mean parity achieved by (the end of) age a by cohort c (a_0 is the minimum age at child bearing, and $n(c)$ is the number of women in the cohort). For fixed values of a the series of mean values $s(c,a)$ can be compared across different cohorts. Note that data for age $a = c$ are censored by the interview (the left-most lower triangles in Figure 3), and for a given cohort the summation can be carried out only up to age $a = c-1$.

Alternatively a cohort's fertility may be cumulated by period:

$$s(c,p) = \sum_{p'=p}^{c-a_0} b(c,p')/n(c) = \sum_{p'=p}^{c-a_0} r(c,p'). \tag{7}$$

The limit $(c-a_0)$ is arbitrary and merely identifies the period of beginning of fertility following age a_0; cumulation up to $p = 0$ gives simply the current mean parity of the cohort.

If the objective is to compare the age pattern of cumulative fertility across real cohorts, form (6) is more suitable compared to (7), as the former controls for age at child-bearing more precisely. On the other hand, (7) has the advantage in that it makes fuller use of the most recent data: the summation can be performed from $p = 0$ as no censoring is involved. (7) also provides perhaps the most convenient form for constructing cumulative fertility measures for 'synthetic' cohorts as described below.

CUMULATIVE MEASURES BY BIRTH ORDER

Disagregation of births according to birth order and cumulation along cohorts or periods provides measures which bring certain features of the fertility pattern into sharper focus. These measures can be computed from the basic cross-tabulation described earlier (Figure 4), repeated for each birth order separately.

One such measure is the Parity Progression Ratio (PPR), defined as the proportion of women of a given parity who proceed to have at least one additional live birth. The PPR may be computed on a cohort basis or on a period basis. For a *cohort* it is computed as the ratio of the number of women in the cohort who have had at least $(i+1)$ live births (by a certain age), to the number of women in the cohort who have had at least i live births (by that age)[3].

Using i to refer to quantities specific to a birth order, the basic statistics required are

$$r_i(c,a) = b_i(c,a)/n(c), \quad i=1,\ 2,\ \ldots \tag{8}$$

ie, the proportion of women in cohort c having a birth of order i at age a. Cumulation along the cohort gives

$$s_i(c,a) = \sum_{a'=a_0}^{a} r_i(c,a') = \frac{1}{n(c)} \cdot \sum_{a'=a_0}^{a} b_{i(c,a')}, \tag{9}$$

which is the proportion of women who have had a birth of order i by (and including) age a -- that is, the *proportion who have had at least i births*. The parity i progression ratio for cohort c by age a is then

$$PPR_i(c,a) = s_{i+1}(c,a)/s_i(c,a). \tag{10}$$

Since by definition $s_0(c,a) = 1$, an important special case of (10) is:

$$PPR_0(c,a) = s_1(c,a)/s_0(c,a) = s_1(c,a), \qquad (10')$$

being equal to the proportion of women in the cohort who become mothers by (the end of) age a.

Though there is no formal difficulty in computing parity progression ratios for incomplete upper age limits, such indices are analytically not easy to interpret. Hence PPRs for cohorts are frequently computed for *completed fertility*, and can be done only for the oldest women in the sample. In populations with longstanding fertility control, a sharp drop in the measure after a certain parity may be expected; in the presence of a more recent decline from high fertility, low values at intermediate parities may be expected.

The proportion of women in the cohort who have had *exactly* i births at age a is the difference between the proportion with at least i births and the proportion with at least $(i+1)$ births, ie $[s(c,a)-s_{i+1}(c,a)]$. As a measure of dispersion in achieved parity by age for the cohort, we may compute the variance of live birth parity $s(c,a)$ as follows:

$$\begin{aligned} v(c,a) &= \sum_{i=0}^{W} \left\{ [s_i(c,a)-s_{i+1}(c,a)] \,.\, [i-s(c,a)] \right\}^2 \\ &= \sum_{i=0}^{W} i.s_i(c,a) - s(c,a).[s(c,a)+1], \end{aligned} \qquad (11)$$

where W is the maximum value of parity for any woman in the sample, ie $s_i(c,a) = 0$ for $i>W$.

The above measures are defined on a cohort basis; to construct birth order specific cumulative measures specific to a *period*, it is desirable to define rates more precisely on the basis of *parity-specific* exposure. This leads to the parity-specific rates described in Section 5 below.

CUMULATIVE PERIOD FERTILITY

Fertility rates may be cumulated across ages or cohorts for fixed periods to construct measures for 'synthetic' cohorts. The concept of the synthetic cohort is based on consideration of the experience of successive real cohorts in their respective life-cycle stages within a particular period, as if it were the consecutive experience of a single cohort. The objective is to provide measures sensitive to period trends in fertility.

Consider first the cumulative of age-period rates across ages:

$$\tilde{s}(a,p) = \sum_{a'=a_0}^{a} [b(a',p)/e(a',p)] = \sum_{a'=a_0}^{a} r(a',p). \tag{12}$$

(We use the symbol $\tilde{s}$ as distinct from s to stress that cumulation is along a *synthetic* cohort). As illustrated in Section 2.3 above, the available data are truncated due to the upper age limit (say c_0) for eligibility; complete data are available only up to age

$$a = (c_0 - p) - 1,$$

and only partially for $a = (c_0 - p)$, while none at all for higher ages. In WFS surveys c_0 is normally taken as 49, occasionally 50, which in practical terms is several years (say 5-10) higher than the upper age limit of the reproductive span. Hence for a number of years preceding the survey, cumulation (12) can be performed upto the end of the child-bearing ages. Such a cumulation gives the Total Fertility Rate (TFR) specific to period p. As an index of fertility, the TFR is independent of the age and sex structure of the population, and may be considered equivalent to the mean parity of a group of women who have passed through the reproductive period experiencing the given (period) age-specific fertility rates. Frequently the TFR is computed from births (and exposure) aggregated by 5-year age groups, rather than from data by single years, particularly when single-year rates tend to be unstable.

It is important to note that the data become progressively more incomplete as we proceed further back from the interview, and in estimating the TFRs, some arbitrary imputation of fertility at higher ages becomes necessary. Hence it is not desirable to go back more than 10 years or so prior to the interview.

While the age-specific rates $r(a,p)$ - the conventional ASFRs - are in themselves of considerable interest, for several reasons equation (12) is not the most convenient form for cumulation: (i) due to censoring at age $a = (c_0 - p)$, it does not make full use of the data for oldest ages; (ii) the denominator $e(a,p)$ is somewhat cumbersome to compute; and (iii) it is analytically desirable to define synthetic cohort measures in a form analogous to those for real cohorts. Hence it is more convenient to construct total fertility from *cohort-period* rates cumulated across cohorts for a fixed period:

$$\tilde{s}(c,p) = \sum_{c'=a_0+p}^{c} b(c',p)/n(c') = \sum_{c'=a_0+p}^{c} r(c',p). \qquad (13)$$

The cumulation can be carried out up to the oldest cohort c_0. In exactly the same way as (12), (13) provides TFRs for several years preceding the survey, the data becoming progressively less complete as we proceed further back. As noted above, (13) is preferable as it makes fuller use of the available data, and the problem of incomplete data is encountered in a less cumbersome way. A major advantage of (13) is that it is analogous to (7) for real cohorts. Comparable indices for real and synthetic cohorts provide measures of change through time. Consider for example the most recent period, $p = 0$, ie the year preceding the interview. The two indices, $\tilde{s}\,(c_0,0)$ from (13) and $s(c_0,0)$ from (7), cover the entire temporal range recorded in the survey: the former is the total fertility according to the currently prevailing age specific rates, and the latter is the mean completed fertility of the oldest women in the sample. In a similar way one can compare $\tilde{s}\,(c,0)$, $s(c,0)$ for a whole range of c values in an attempt to isolate changes in the timing of fertility. In terms of Figure 3, we are comparing cumulations along rows (ie $s(c,0)$ or mean parity of real cohorts) with downward cumulation along the left most column (for period $p = 0$) upto and including row c.

Similar comparisons can also be made for earlier periods. In other words, we compare cumulation along a row (say c) of Figure 3 with that along a column (say p), each cumulation proceeding up to the cell (c,p) where the row and the column concerned intersect.

We may note a basic distinction between a measure such as (7) based on a real cohort of women, and a synthetic measure such as (13) constructed by putting together period specific experience of a number of different cohorts. For a more detailed analysis one may decompose the total cumulative fertility $s(c,p)$ by, for example, birth order or into separate indices of marital fertility and nuptiality; such decomposition may provide greater analytic insight, but in no way alters the overall level $s(c,p)$ - which is based on the actual experience of a real cohort of women. By contrast, even the *level* of a synthetic measure such as $\tilde{s}(c,p)$ corresponds only to a given degree of specificity. For example, (13) is defined specific to the prevailing age-pattern of fertility, and takes no *separate* account of birth-orders of the children being born, or of the prevailing levels of marital feritlity and age at marriage. It is possible to define the synthetic cohort measure with a greater degree of specificity - as we will indicate subsequently in relation to parity specific fertility rates - to obtain analytically more precise indices, indices which are more reflective of the conditions pertinent to the period concerned and less dependent on past history. On the other hand, it should be noted that an analytically more precise index of this type can also have serious disavantages in certain circumstances: it can be misleading of the underlying trend due to excessive sensitivity to short run fluctuations from year to year, as well as to longer term reporting errors.

LESS REFINED MEASURES

One may also wish to construct certain less refined measures for the purpose of comparison with other sources of data. It should be noted at the outset that except for such comparison, the following measures are of little relevance when more refined measures can be constructed from detailed retrospective data.

The ratio of the total number of live-births during a given period to the total number of person-years lived by women in the child-bearing ages during that period is called the General Fertility Rate (GFR):

$$GFR(p) = \sum_{a=a_0}^{a_m} b(a,p) \Big/ \sum_{a=a_0}^{a_m} e(a,p),$$

where a_0 is the lower limit and a_m the upper limit of the reproductive span. Like the TFR, the GFR can be constructed only for a few years prior to the survey, ie for years for which data are available for essentially all reproductive ages. However, unlike the TFR, the GFR is not independent of the female age distribution in the sample.

Another commonly used measure, dependent on both sex and age composition of the population, is the Crude Birth Rate (CBR), for which the numerator is the same as that for the GFR, but the denominator is the estimated total population at the mid-point of the period concerned. With the type of data on population available in the WFS Household Schedule, the CBR can be estimated only for the period immediately prior to the interview; retrospective estimates will require reverse survival of the population on the basis of extraneous data.

Finally, we may also mention standardised birth rates which, for the purpose of constructing measures comparable across populations, are adjusted to take account of differences in the structure of two populations by age, sex marital status or any other such characteristic.

3. AGE-SPECIFIC MARITAL FERTILITY RATES

The fertility rates described in the previous section were based on all births irrespective of marital status, with exposure to child-bearing defined as the total time elapsed; the universe was all women, ever-married as well as never-married. In this section marital fertility rates will be defined with exposure and births restricted to periods after first marriage, or to periods spent with marriage; the relevant universe in constructing the rates is confined to ever-married women. Use of marital fertility rates permits separating the effect on overall fertility of changes in nuptiality (age at marriage, propensity to marry and marriage stability) from that of changes in the level of fertility within marriage.

Age-specific marital fertility rates for young ages are by definition restricted to women who marry early. The selectivity of the measures decreases with age and eventually excludes only those who never marry. In defining and interpreting the rates it is generally necessary to take this selection bias into account explicitly by controlling for age at marriage.

Age-specific rates may be converted to approximate marital rates by restricting the base population to currently married or ever-married women (but with no other restriction on births), on the expectation that most births occur within marriage. Such measures are approximate in the sense that in the interest of simplicity, an exact correspondence between the numerator (births) and the denominator (periods of exposure) is not sought.

A refinement would be to compute the "legitimate fertility rates" in which the numerator is restricted in addition to legitimate births. However there can be important socio-cultural and national differences in the definition of "marital status" and "legitimate" which have a serious effect on the comparability of marital fertility rates across different populations.

In the following, marital fertility rates will be defined with greater numerical precision. One may consider marital exposure at different levels of refinement:

1) Exposure defined in terms of the total time elapsed since entry into the first union. This definition makes the minimum possible use of the marriage history data, and disregards marriage dissolution and remarriage subsequent to first marriage.

2) Exposure defined in terms of the time spent within marriage, ie, excluding periods of non-exposure between the end of one marriage and the beginning of the next marriage (if any).

3) Where the availability and quality of the data permits, one may also exclude periods of non-exposure due to prolonged but temporary separations within unions as well as known periods of infecundity.

4) Sexual unions may be distinguished by type (for example, formal marriages from common-law unions from more casual 'visiting relationships'), and rates computed separately for each union type.

The main principle in defining a marital fertility rate is that if there is a restriction on the base interval of exposure, then a corresponding restriction should be placed on the births included in the numerator. Hence it is necessary to identify births in relation to marital status of the mother. This may be done with reference to her status at the time of conception leading to the birth concerned, or to that status at the time of delivery. We will follow the second of these alternatives.

By comparing the dates in the birth history with those in the woman's marriage history, we can identify pre-marital births (occurring before entry into the first union), 'extra-marital' births (those occurring in a subsequent interval when the mother was not in the married state), and marital births. Where relevant, extra-marital births may include those occurring at the time of prolonged separation within a union, while marital births may be classified further by union type.

3.1 'EVER-MARRIED' EXPOSURE

This refers to the exposure defined in terms of the total time elapsed since the woman's entry into her first union. The numerator of the rate excludes pre-marital births ie births for which the date $B(i)$ is prior to $M(1)$, the date of the mother's first marriage:

$$B(i) < M(1) \tag{14}$$

The procedure for defining the appropriate length of exposure to individual women for various periods, ages and cohorts is slightly more involved and is described in Appendix I in some detail. Essentially the proposed procedure is first to compute 'unrestricted' exposure (as for the rates described in the previous section) and then to modify it to exclude all exposure prior to first marriage.

Apart from the above-mentioned modifications of excluding pre-marital births and exposure from individual women's contribution to the aggregate rates, the scheme for classification of the data by birth cohorts, periods and mother's age at child's birth, and the associated cross-tabulations (Figure 4) etc. will be exactly as before and need not be repeated here.

3.2 EXPOSURE WITHIN MARRIAGE

This refers to the time spent within de facto unions. The numerator of the rate will include only marital births, ie births which occurred at a time the mother was in a union (alternatively one may take the date of conception rather than the date of occurrence in determining whether a birth is 'marital'). The births included satisfy the condition

$$M(j) \leqslant B(i) < D(j), \qquad (14')$$

for any marriage j of the women beginnig at date $M(j)$ and dissolving at date $D(j)$*.

---

* *For the current marriage, we may regard for convenience* $D(j) = I$, *the date of interview.*

To compute the woman's periods of exposure, the procedure proposed in Appendix I considers one marriage at a time. The elapsed times following the two events *M(j)*, the beginning of the marriage, and *D(j)*, the termination of marriage, are computed in the same way as for *M(1)* in section 3.1; their difference gives the time spent within marriage *j*; these are added together for all marriages to obtain the total length of marital exposure for the woman.

Though the computation of proper marital fertility rates presents no conceptual problem, practical difficulties can arise from the fact that dates in birth and marriage histories in WFS surveys are collected, edited and imputed (where required) independently of each other, resulting in uncertainty whether in fact a particular birth is or is not within marriage.

It is perhaps worthwhile to compute both ever-married (section 3.1) as well as proper marital fertility rates in most circumstances. However, the latter are more taxing on the quality of date reporting. When the incidence of marriage dissolution is low the two sets of rates can be expected to be very similar, and it may be sufficient to compute the simpler ever-married rates. Similarly, though at the other extreme, where marriage is unstable and the actual marital status following first marriage is a poor indicator of exposure, it will be more meaningful to compute only ever-married rates.

3.3 COMMENTS

1. Age-specific fertility of a birth cohort of women is determined by (1) fertility within marriage, (2) the incidence of marriage dissolution and remarriage, and (3) the proportion of the cohort who are ever-married at each age. Proper marital fertility rates measure (1); ever-married fertility rates confound (1) and (2); and rates for unrestricted exposure confound all the three factors.

   A basic objective of introducing age-specific marital fertility rates is to separate out overall fertility into marital fertility and nuptiality components. By considering these components separately and explicitly introducing age at marriage, more refined alternatives to the conventional ASFRs can be developed[8].

2. Measures of cumulative fertility may be constructed by summing up rates up to specified ages along cohorts (ie within given cohorts: rows in Figure 3), or along periods (ie within given periods: columns in Figure 3) in the form of equation (13) above. Cumulation of ever-married rates from age say 20 to age a gives the mean number of children born by age a to women first married at age 20; similar cumulation of proper marital rates gives that mean for women continuously in the married state since first marriage. Cumulation along a given period to the end of the child-bearing span gives the Total Marital Fertility Rate (TMFR).

   Due to the selection effect operating at young ages noted at the beginning of Section 3, the interpretation of these cumulative measures is not straightforward. Cumulative measures can, however, be instructive of the pattern of marital fertility, if these exclude very young ages; a reasonable approach would be to start cumulation from around the median age at first marriage. Control for age at marriage, at least in broad groups, will also be generally desirable. For comparison across populations, marital rates are sometimes standardised on some 'standard' distribution of proportions ever-married or currently married by age.

3. It has been observed that due to inaccuracies, as well as incompleteness, in the reporting of dates of vital events in developing country surveys, the proportion of first births classified as 'premarital' according to equation (14) is frequently much too high compared to what might be reasonably expected from the socio-cultural context. In other words (14), which compares dates of two events - events which are in the relatively distant past but generally close to each other -at the level of the *month* of occurrence, is too strict a condition and can result in misclassification of some marital births as premarital. One way to make the condition less strict would be to replace the century-month codes $B(i)$ and $M(1)$ by the (12 month) periods of occurrence of the respective events (as defined for example by equation I.2 in Appendix I). A birth would be defined as premarital only if its *period* of occurrence is prior to that of first marriage.

It may be appropriate to make a similar modification to equation (14') a birth would be classified as being outside a marriage only if its period of occurrence is *prior* to the period of beginning of the marriage or *after* the period during which the marriage terminates.

4. Where quality of the marriage history data do not permit computation of detailed marital fertility rates as decribed above, a useful approximation to these rates may be obtained by computing rates for unrestricted exposure but confined to women married at the time of the interview (for proper marital rates), or to ever-married women (for ever-married rates). This approximation is meaningful only for periods immediately prior to the interview, and is comparable to conventional marital fertility rates from other sources such as vital registration.

5. In situations where the incidence of marriage disolution is relatively high, but at the same time fertility following first marriage is known largely to be within marriage, the following simplification may be introduced in the computation of proper marital fertility rates: the numerator will consist of all births following first marriage (as in the case of ever-married rates), or of all births if pre-marital fertility is also negligible; the denominator will consist of the time actually spent within marriage (as for proper marital rates). This simplification partially overcomes the difficulty, noted earlier, in unambiguously classifying births as marital or extra-marital in the presence of independent errors and incompleteness in birth and marriage histories.

6. A simple, but crude estimate of current marital fertility may be based on the proportion of currently married women reporting a current pregnancy, classified by age-group (or marriage duration group). As current pregnancies of short duration tend to be particularly under-reported, it is preferable to restrict the calculation to pregnancies

of longer duration only. For example, the proportion of currently married women reporting a current pregnancy of durations 4 to 8 months (five out of the possible nine months in all), multiplied by 12/5 give an approximate value of the current marital fertility rate --assuming no seasonality, pregnancy wastage, or under-reporting at these pregnancy durations.

Too much reliance should not be placed on this crude measure as the completeness of reporting of current pregnancies may vary by age, background of the woman, and from country to country.

4. DURATION-SPECIFIC RATES

4.1 MARRIAGE DURATION AS A CONTROL FOR ANALYSIS

Apart from the woman's current age, a basic demographic control used in presentation and analysis of fertility is the duration since her first marriage. This duration provides, in most circumstances, a more precise indication of the length of exposure to child-bearing than does age. It is not infrequent to find that even among sub-populations with substantially different fertility, there is a considerable uniformity in the rate of child-bearing during the first years of marriage, with fertility differentials emerging only at later durations. Women marrying about the same time also tend to share certain values and experiences at similar points in their family building process, a consideration which can be particularly important in a developing country where many of the relevant facilities such as family plannning services, maternal and child health care etc. are of recent origin.

On the other hand, a marriage cohort lacks a strict biological basis due to differences in age at marriage among individual women. For example, women marrying very young may have substantially lower fertility in the first years of marriage owing to adolescent sub-fecundity; at the other end, women marrying very late will have lower fertility at any marriage duration simply because of their age. Secondly, while age-specific marital fertility rates at younger ages are restricted to early marrying women, an opposite bias is present for marriage cohorts: since women currently above the child-bearing ages are excluded from the individual interview sample, those at the longest marriage durations are selectively the early marrying ones. For example women at marriage duration 30 year must all have married before age 20, given that the sample is confined to women currently aged under 50. Due to these reasons, it is usually desirable to compute duration-specific fertility measures after *controlling for age at first marriage,* at least in broad groups.

4.2 COMPUTATIONAL FORMS

The procedure for computing duration specific rates is very similar to that for age-specific rates. The basic classification of the data is in terms of the three related variables:

- m, the woman's marriage cohort (replacing birth cohort c used earlier);
- d, mother's duration since first marriage at child's birth (replacing age a); and
- p, the period of occurrence (defined exactly as before).

As before, cohorts and periods may be defined either in terms of completed years before the interview (Scheme 1; see section 2.1), or as fixed calendar-years (Scheme 2). In either case, a marriage cohort consists of women married during the same period, and (m,d,p) are related in a form similar to equation (2):

$$d = (m-p) - 1 \quad or \quad d = (m-p)$$

As for age-specific marital fertility rates, exposure for duration-specific rates may be defined either as ever-married exposure (ie time elapsed since first marriage) or as marital exposure (ie time spent within unions).

Appendix I indicates the procedure for the classification of an individual woman's births and periods of exposure into the (m,p,d) array which can be cumulated into a cross-tabulation of the form illustrated in Figure 4 to compute (marriage) cohort-period specific, cohort-duration specific and duration-period specific rates (analogous to equations (3)-(5)). Appendix II provides numerical examples of the aggregate level measures.

4.3 RELATED MEASURES

CUMULATIVE COHORT AND PERIOD MEASURES

Cohort specific rates may be cumulated, for a given marriage cohort, across durations or periods to obtain measures of cumulative cohort fertility; similarly, cumulation of period rates across durations provides measures of cumulative period fertility. For example, analogous to equation (12) above, we define

$$\tilde{s}(d,p) = \sum_{d'=0}^{d} r(d',p) \tag{15}$$

where $\tilde{s}(d,p)$ may be considered equivalent to the mean parity after d years of marriage of a group of women experiencing the duration specific rates prevailing at period p.

To construct comparable measures for real and synthetic cohorts, we may cumulate cohort-period specific rates in forms similar to equations (7) and (13):

For a real cohort m (ie for women first married m years ago), mean parity at (the end of) period p is

$$s(m,p) = \sum_{p'=p}^{m} b(m,p')/n(m) = \sum_{p'=p}^{m} r(m,p'), \qquad (16)$$

where $n(m)$ is the number of women in marriage cohort m, and $b(m,p)$ is the number of births during period p to these women.

For an equivalent synthetic cohort corresponding to period p, we have

$$\tilde{s}(m,p) = \sum_{m'=p}^{m} [b(m',p)/n(m')] = \sum_{m'=p}^{m} r(m',p). \qquad (17)$$

The synthetic cohort measure can be refined by explicitly introducing age at first marriage distribution corresponding to the period concerned. Consider a (real) cohort m classified into age-at-marriage subgroups; using subscript g to refer to quantities relating to a particular subgroup, we have by definition:

$$n(m) = \sum_{g} n_g(m) \; ; \; b(m,p) = \sum_{g} b_g(m,p);$$

and

$$s(m,p) = \frac{1}{n(m)} \sum_{p'=p}^{m} \sum_{g} b_g(m,p') = \frac{1}{n(m)} \sum_{p'=p}^{m} \sum_{g} n_g(m).r_g(m,p'),$$

where $r_g(m,p) = b_g(m,p)/n_g(m)$ is the rate specific to the particular age-at-marriage subgroup. We may rewrite $s(m,p)$ as

$$s(m,p) = \sum_g \left\{ \frac{n_g(m)}{n(m)} \cdot \sum_{p'=p}^{m} r(m,p') \right\} = \sum_g \left\{ \frac{n_g(m)}{n(m)} \cdot s_g(m,p) \right\}. \quad (16')$$

In other words, the mean parity of a (real) marriage cohort is expressed as the weighted sum of the mean parities of its age-at-marriage subgroups, the weights being the proportional distribution of the cohort according to age at marriage. The equivalent form for a synthetic cohort is[8]

$$\tilde{s}(m,p) = \sum_g w_g(p) . \tilde{s}_g(m,p) \quad (17')$$

where $\tilde{s}_g$ is defined in the same way as (17) ie

$$\tilde{s}_g(m,p) = \sum_{m'=p}^{m} [b_g(m',p)/n_g(m)] = \sum_{m'=p}^{m} r_g(m',p),$$

and weights $w_g(p)$ are the proportional distribution according to age-at-marriage of first marriages which occur *during period p*. Note the measure defined by (17') is not necessarily numerically identical to that defined by (17), while that is the case with (16) and (16').

CUMULATIVE MEASURES BY BIRTH ORDER

Duration-specific rates classified by birth-order can provide powerful measures for elucidating trends in the timing and level of fertility. Rates by birth order can be cumulated along marriage cohorts, giving proportions of women in the cohort who have had a birth of a given order by specified marriage durations; cumulative rates can be used to define measures such as marriage-cohort specific parity progression ratios (see equation (10)).

As noted in Section 2.4, to construct parity-specific period cumulative measures, it is desirable to define rates based on parity-specific exposure as done in the following section.

5 PARITY-SPECIFIC RATES

5.1 DEFINITION OF EXPOSURE

A parity-specific rate is computed "with the number of live births of order *i* to women in a given age group or duration of marriage group during a year as the numerator, and the women-years lived during the year by women of parity *(i-1)* in that age or duration group as the denominator"[3].

It will be useful to clarify the similarities and differences between a parity-specific rate and other types of rates previously discussed:

1. For parity-specific rates, the retrospective birth histories can be classified in terms of the mother's age cohort, her age at birth of child and period of occurrence, as in the case of age-specific rates (section 2); or in terms of marriage cohort, duration at birth, and period, as in the case of duration-specific rates (section 4). However, it is more common here to aggregate data over several years.

2. For parity-specific rates, births in the numerator are classified by birth order, and the denominator is confined to person-years lived *at the previous parity*. These differ from age/duration-specific rates by birth order described earlier in that the denominator for the latter is the person-years of exposure irrespective of the woman's parity at the time. For this reason parity-specific rates are sometimes referred to as "true birth order rates"[3].

3. Age/duration-specific rates described earlier classified by birth order can be cumulated along *cohorts* to obtain cumulative proportions

of women in the cohort achieving certain parity by specified age/duration. For parity-specific rates, however, the denominator for a given cohort changes from one age/duration to another, and the rates can not be meaningfully cumulated along a cohort.

On the other hand, the former classified by birth order can not be meaningfully cumulated along *periods* to provide order specific measures. One of the main objectives of computing parity-specific rates described in this section is to obtain more precise measures of period fertility. For this reason it is more useful to compute parity-specific rates on cohort-period basis (rather than age/duration basis).

4. Parity-specific exposure is usually taken as 'unrestricted' exposure in so far as the mother's marital status is concerned; as a simple extension, it may be restricted to time elapsed since first marriage.

   However, restriction of the parity-specific exposure to time spent within unions is computationally complicated, and is likely to be specially taxing on the quality of birth and marriage history data.

Appendix I provides the necessary computational details. Births in the numerator are classified in exactly the same way as for age/duration-specific rates by birth order described earlier. A woman's length of exposure at a given age/duration during a given period is conditional on her parity: for parity i specific rate, she becomes exposed only after having achieved parity *(i-1)* and ceases to be exposed after having achieved parity i*.

5.2 PERIOD OR "SYNTHETIC COHORT" MEASURES

It was noted earlier that age/duration-specific rates by birth order (such as $r_i(c,a)$ in equation (8)), cumulated along a *cohort*, provide the proportions (such as $s_i(c,a)$; equation (9)) in the cohort who achieve certain parities by specified age/duration. In a similar way, parity-specific rates defined in this section can be used to obtain measures of (parity-specific) cumulative *period* fertility.

Let $\tilde{r}_i(d)$ be the parity i specific rate at marriage duration d for a given period**. Given $\tilde{r}_i(d)$, our objective is to compute the cumulative proportions $\tilde{s}_i(d)$ who have had a birth of order i by (the end of) marriage duration d, among a "synthetic cohort" of women experiencing the given duration-parity specific rates $\tilde{r}_i(d)$ for the period. The measure $\tilde{s}_i(d)$ is analogous to $\tilde{s}(d,p)$ in equation (15), except for being parity-specific.

---

* *We may note in this context the distinction between a parity i specific rate for a given period and the measure of probability of having a birth of order i during that period. The denominator for the latter is the number of women at parity (i-1) at the beginning of the period concerned; these women are taken to remain 'exposed' throughout the period irrespective of whether any of them have a birth of order i during the period; similarly, no other women become 'exposed' by having a birth of order (i-1) during the period concerned.*

Numerically, the two measures mentioned above should not differ much specially for single-year periods.

** *In the following, all quantities refer to a fixed period; hence subscript 'p' has been dropped for simplicity. Also, we use the symbol $\tilde{r}_i$ as distinct from r_i used earlier to emphasise the fact that rates here are based on parity specific exposure and refer to a fixed period ('synthetic cohort'). Note that the following measures may also be defined in terms of retrospective age rather than marriage duration.*

In fact the former provides a more precise alternative to equation (15) for the mean achieved parity by duration d for the synthetic cohort (see below).

To express $\tilde{s}_i(d)$ in terms of the known $\tilde{r}_i(d)$, we begin with the definition of the former. Since $\tilde{s}_{i(d)}$ is the proportion of the synthetic cohort who have had a birth of order i by (the end of) duration d, $\tilde{s}_i(d-1)$ is that proportion by duration $(d-1)$, the proportion who have a birth of order i during duration d is:

$$\tilde{s}_i(d) - \tilde{s}_i(d-1). \qquad (18)$$

The proportion who have exactly i births by (the end of) duration d is the difference of those with at least i births and those with at least $(i+1)$ births, ie

$$\tilde{s}_i(d) - \tilde{s}_{i+1}(d). \qquad (19)$$

Equation (18) is the numerator in the definition of $\tilde{r}_i(d)$. The denominator is the appropriate number of women exposed during d to the risk of having a birth of order i, and is made up of the following three components:

(1) The number who have had a birth of order $(i-1)$ but not of order i by the end of duration $(d-1)$, ie

$$\tilde{s}_{i-1}(d-1) - \tilde{s}_i(d-1);$$

(2) *Plus* half the number who become exposed by having a birth of order $(i-1)$ during the interval (these women are exposed at parity $(i-1)$ on the average for half the length of the interval):

$$\tfrac{1}{2}.[\tilde{s}_{i-1}(d) - \tilde{s}_{i-1}(d-1)] ;$$

(3) *Minus* half the number who leave the exposed state by having a birth of order i during the interval:

$$\tfrac{1}{2}.[\tilde{s}_{i}(d) - \tilde{s}_{i}(d-1)] .$$

By definition of the parity-specific rate $\tilde{r}_i(d)$, we have after some rearrangement:

$$\tilde{s}_i(d) = s_i(d-1) + \left(\frac{\tilde{r}_i(d)}{1+\frac{1}{2}\tilde{r}_i(d)}\right).\left(\tfrac{1}{2}\tilde{s}_{i-1}(d-1)-\tilde{s}_i(d-1)+\tfrac{1}{2}\tilde{s}_{i-1}(d)\right). \qquad (20)$$

For first-order births $(i=1)$, equation (20) reduces to the following (by definition $\tilde{s}_0(d)=1$ for all d, since all women have had at least zero births):

$$\tilde{s}_1(d) = \tilde{s}_1(d-1) + \left(\frac{\tilde{r}_1(d)}{1+\frac{1}{2}\tilde{r}_1(d)}\right).\left(1-\tilde{s}_1(d-1)\right) . \qquad (21)$$

For the first year following marriage $(d=0)$, one may reasonably take births of any order to be concentrated at the end of the interval, giving the simple relationship.

$$\tilde{s}_1(0) = \tilde{r}_1(0);\ \text{in general}\ \tilde{s}_i(0) = r_i(0). \qquad (22)$$

Starting with (22), first-order specific rates $\tilde{r}_1(d)$ can be used in (21) to compute for the synthetic cohort cumulative proportions $\tilde{s}_1(d)$ having first birth by (the end of) duration d; then (20) can be applied successively to obtain cumulative proportions for each higher parity in turn.

These cumulative proportions form the basis for computing other measures for the synthetic cohort. For example, duration-specific parity i progression ratios are:

$$PPR_i(d) = \tilde{s}_{i+1}(d)/\tilde{s}_i(d).$$

The mean parity by duration d for the synthetic cohort is:

$$\sum_{i=1}^{W} i\,[\tilde{s}_i(d) - \tilde{s}_{i+1}(d)]$$
$$= \sum_{i=1}^{W} i.\tilde{s}_i(d) - \sum_{i=2}^{W+1} (i-1)\tilde{s}_i(d) = \sum_{i=1}^{W} \tilde{s}_i(d), \qquad (23)$$

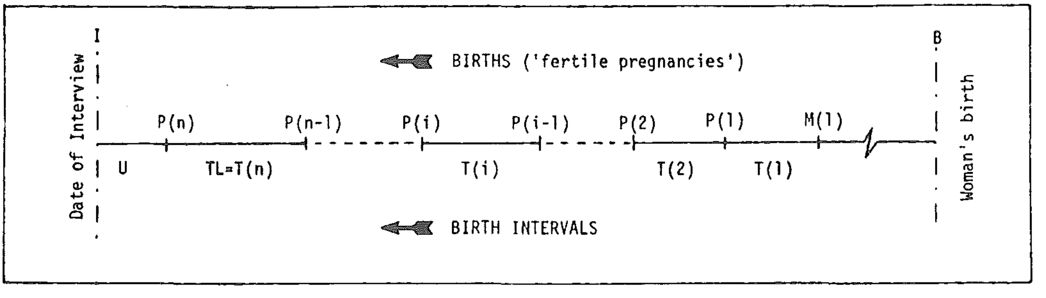
where W is the maximum value of parity for any woman in the sample (ie $\tilde{s}_i(d) = 0$ for $i > W$). Equation (23) is a more precise form compared to (15), being based on the prevailing *parity-specific* rates.

6 BIRTH INTERVALS

6.1 TYPES OF BIRTH INTERVALS

Retrospective birth history data permit computation and analysis of birth intervals which can be divided into two broad types: *closed intervals* terminated by a live birth; and open intervals censored by an arbitrary point in time such as the interview. Figure 6 provides a definition of the various types of intervals. It shows the retrospective history of a woman in a form similar to Figure 1. *B* is the woman's birth date, *M(1)* the date of her first marriage, and *I* the date of interview. The sequence *P(1)* to *P(n)* indicate the dates of her live births; these differ from *B(i)* in Figure 1 only in that here any set of multiple births is treated as a **single event**. In other words, *P(i)* are 'fertile pregnancies', or confinements leading to live births; the total *n* is less than the number of children born (*w* in Figure 1) to the extent multiple births (twins etc.) have occurred. In the following we will use the term "births" to refer actually to fertile pregnancies.

Figure 6. Closed and Open Birth Intervals



$T(2)$ to $T(n)$ are inter-birth (closed) intervals. These are designated according to the order of the birth which terminates the interval.

$$T(i) = P(i) - P(i-1), \quad 2 \leq i \leq n, \tag{24}$$

is the gross length of the interval, measured as the time elapsed (usually in months) from the $(i-1)^{th}$ birth to the i^{th} birth, irrespective of any periods of non-exposure within the interval. Note that an equation such as (24) gives the length of the interval to the nearest month (and not in completed months).

The interval from first marriage to first birth is termed as the first birth interval, $T(1)$,

$$T(1) = P(1) - M(1), \tag{25}$$

and is defined only for ever-married women who have had at least one live-birth. This interval -- which begins from an event other than a live-birth -- is qualitatively different from inter-birth intervals in several respects. First, in some contexts the date of first marriage -- even when elicited as the effective date of entry into a union -- may not be a good indicator of the onset of sexual activity and exposure to child-bearing. At any rate, it is not likely to be as good an indicator of initial exposure as a later birth is of resumption of exposure. Secondly, unlike an inter-birth interval, the first birth interval does not include a period of post-partum sterility. Thirdly, there is no biologically determined minimum length of the interval; in fact, in the presence of pre-marital births (as well as common inaccuracies in the reporting of dates), first birth intervals will be negative. For these reasons, it is best to analyse the first interval separately from inter-birth intervals; actually even for the latter a control for birth order is highly desirable.

An interval of special interest is the last closed interval, TL (often referred to simply as *the* closed interval). For a woman with n births ('fertile pregnancies'), it is the interval between last two births,

$$TL = T(n) = P(n) - P(n-1). \qquad (26)$$

As an inter-birth interval, TL is defined only for women with at least two births $(n \geq 2)$. However this definition is sometimes extended to include also the case of ever-married women with only one live birth, the date of first marriage, $M(1)$, defining the beginning of the interval:

$$TL = P(1) - M(1), \text{ for women of parity } n=1. \qquad (27)$$

Another possible extension is to include the expected date of termination (CP) of a current pregnancy (if any) as the prospective 'last birth'. Hence for a currently pregnant woman with at least one live-birth $(n > 0)$

$$TL = CP - P(n);$$

for a currently pregnant ever-married woman with no live births

$$TL = CP - M(1).$$

With these extensions the interval is defined for all women who satisfy at least two of the following three conditions: (1) are ever-married, (2) are currently pregnant, and (3) have had at least one live birth.

While for certain purposes it is useful to extend the definition of the last closed interval as indicated above, it is best to confine the analysis to inter-birth intervals defined by equation (26), specially when the objective is to study fertility differentials. Current pregnancy is frequently under-reported, the extent of which may differ from one sub-population to another. And as already mentioned, intervals beginning with marriage are qualitatively different from inter-birth intervals and should be analysed separately in any case.

We also define the *open birth interval* as the time elapsed since last birth:

$$U(n) = I - P(n). \qquad (28)$$

An extension of (28) for ever-married women with no live birth is to define the open interval as measured from first marriage

$$U(0) = I - M(1).$$

If a current pregnancy has been used to define the last closed interval (see above), the open interval is then not defined for currently pregnant women.

It is also possible to define an open interval retrospectively, for example as the interval between an arbitrarily chosen point (*'t'* in Figure 6) and the birth immediately preceding it (birth $P(i-1)$ in Figure 6). The objective is to compare the distribution of the open interval which would have been observed at some past moment with the distribution observed at the time of the interview. In many contexts, however, the available data are not of sufficient quality to permit a meaningful comparison of this type.

6.2 AGGREGATE MEASURES BASED ON INTERVALS

Data on birth intervals can be employed in a variety of ways in fertility analysis. At the level of the individual or the aggregate, these may be used to construct predictor or explanatory variables, classificatory or control variables, and dependent variables. Detailed analysis of birth interval data is a separate area of study in its own right, and beyond the scope of this Bulletin. The following comments are confined to uses of the data to construct certain descriptive measures at the aggregate level; in particular measures (such as the mean, median, other percentiles, standard deviation, and possibly higher moments) relating to the distribution by interval length for various types of intervals.

In constructing aggregate measures such as the mean or median length of intervals, it is necessary to recognise the presence of a selection or truncation bias in data from a cross-sectional survey. The timing of the interview is arbitrary with respect to a woman's reproductive history, and the observations are restricted to limited periods of time in the total (prospective) reproductive span. Suppose, for example, that the mean interval from first birth to second birth is calculated for successive marriage cohorts. The estimates must be restricted to women who have had at least two births, which, for recent cohorts, is restricted selectively to women with short birth intervals. The bias is a function of the length of exposure (as represented by, say, the woman's age or marriage duration), and of parity, with larger bias among groups with shorter durations of exposure and higher parity.

It is therefore necessary for birth interval analysis to be order-specific. Measures such as the mean should be confined to interval of the same order. Secondly, it is desirable to reduce the truncation effect by controlling age at entry into the parity in question as well as the length of the observation time. At a less refined level of analysis, for example, measures of interval length distribution may be based simply on the observed distribution, with controls to limit the selection bias as far as possible. For example, in estimating the mean length of the last closed interval, one may exclude intervals longer than say 5 years, and also restrict the calculation to women for whom the interval began at least five years ago. In general, however, it is not feasible to remove the selection bias altogether as it operates throughout the reproductive history; also, the available sample size limits the degree to which controls can be introduced.

A more appropriate approach is to construct life-tables *combining data on order-specific closed and open birth intervals*. Life-table techniques are beyond the scope of this Bulletin and are discussed in other WFS documents[10]. However, below we briefly outline a simple procedure[9] based on straightforward cross-tabulation of the birth history data. Apart from its relevance in the present context, the procedure is of considerable interest in the study of post-partum phenomenon such as lactation, post-partum abstinence and amenorrhoea, data on which are frequently collected in WFS surveys.

Consider a cohort of women who have achieved or surpassed parity i at the time of the interview. Following a birth of order i, any women in the group experiences one of the two events: *(1)* a birth of order $(i+1)$, or (2) the interview at parity i. Taking the date of occurrence of birth i as the point of reference, let period p be measured (say, in *months)* from this date. At the beginning of period p, let $f(p)$ be the proportion in the group who have not experienced another birth nor the interview. During period p itself, let $g(p)$ be the proportion who have a birth of order $(i+1)$ and $h(p)$ be the proportion who experience the interview (but no birth) during this period; that is, at the beginning of the next period we have:

$$f(p+1) = f(p) - [\, g(p) + h(p) \,]. \tag{29}$$

Noting that by definition $f(0)=1$, equation (29) gives $f(p)$ for all p in terms of $[g(p), h(p)]$. The latter quantities are obtained from a simple cross-tabulation as follows:

$h(p)$ is the proportion of women who were interviewed p months after their i^{th} birth and were still at parity i. In other words, the classification of women of current parity i according to the length of the open interval

$$p = I - P(i),$$

gives $h(p)$, where $P(i)$ is the date of the i^{th} birth.

Similarly, $g(p)$ is given by the frequency distribution of women of current parity greater than i according to the length of the $(i+1)^{th}$ closed interval

$$p = P(i+1) - P(i).$$

During period p, $[f(p) - \frac{1}{2} h(p)]$ is the approximate proportion exposed to the risk of having a birth; hence the probability, say $q(p)$, of an interval being "closed" by a birth during period p is:

$$q(p) = g(p) \,/\, [f(p) - \tfrac{1}{2} h(p)] . \qquad (30)^*$$

The probability of not having a birth during period p is $[1-q(p)]$, and that of not having a birth till the beginning of period p is the product

$$\prod_{p'=0}^{p-1} [1-q(p')] = l(p), \text{ say, with } l(0) = 1 \text{ by definition.}$$

$l(p)$ is the life-table distribution of interval length at specified parity (i).

---

* *An alternative form is:*

$$q(p) = 1 - \left[1- \frac{g(p)+h(p)}{f(p)}\right]^{\frac{g(p)}{g(p)+h(p)}}$$

7. CONCLUSION

A large number of fertility measures which can be constructed from retrospective birth history data have been specified in considerable detail with the objective of promoting a common understanding of the basic concepts and definitions involved. At the same time it is necessary to bear in mind the limited scope of this Bulletin. We have assumed data coded down to the level of the month, and in describing the various measures, have taken data at their face value. However, in real-life surveys the problem of missing values and imputation are far from trivial. For dates coded in various forms such as calendar-years, years ago, ages etc., the exact interpretation of the data available is not always unambiguous. Obviously, with data of suspect quality, with for example a substantial proportion of months imputed, it will not be easy to draw firm conclusions, particularly concerning fertility trends.

In analysis of data sets of varying quality, two distinct approaches are possible. One approach would be to proceed step by step starting with the crudest measures least taxing on the quality of the data, evaluate data quality at each step and determine whether more refined measures are justified. Alternatively one may construct a variety of measures and use those simultaneously to evaluate the data as well as to draw substantive conclusions. A priori, we recommend the second approach.

Whatever the approach, the measures described here are nevertheless of direct relevance. Analysis of birth history data, particularly as relating to trends, must proceed simultaneously with evaluation of the quality of the data; it depends upon that evaluation and at the same time provides means for the evaluation. Different sources of bias can produce similar distortions in the observed pattern of fertility, so that separation of the various sources of bias is not easy. The variety of measures described here allows a certain degree of separation since different measures tend to be more sensitive to different sources of bias.

Finally, it may be useful to list a few sources describing procedures for *indirect estimation* of fertility measures[7,11,12,13]. Such measures are not considered in this document as they refer to mortality levels which are not obtainable from a cross-sectional survey designed to obtain retrospective birth history data.

REFERENCES

1. International Statistical Institute, *Core Questionnaires*, WFS Basic Documentation, 1975.

2. World Fertility Survey, *Guidelines for Data Processing*, WFS/TECH. 1010, 1979.

3. Grebenik, E. and A. Hill, *International Demographic Terminology*, IUSSP Papers No. 4, 1974.

4. International Statistical Institute, *Guidelines for the Country Report No. 1*, WFS Basic Documentation, 1977.

5. Coale, A. Age Pattern of Marriage, *Population Studies* 25(2) 193-214, July 1971.

6. Piampiti, S. and J. Knodel, *Revised Estimates of Age-Specific Fertility Rates from the Survey of Population in Thailand*, Chulalongkorn University and National Statistical Office, Thailand, 1978.

7. United Nations. Manual IV: *Methods of Estimating Basic Demographic Measures from Incompleted Data*, ST/SOA/Series A/42, 1967.

8. Ryder, N.B., *Measuring Trends in Marital Fertility*, IUSSP Committee on Comparative Analysis of Fertility, Seminar on Analysis of Maternity Histories, London 1980.

9. Ryder, N. *Tabulation Plan*. Unpublished document prepared for the World Fertility Survey, 1974.

10. Smith, D.P., *Life-Table Analysis*, World Fertility Survey Technical Bulletins (forthcoming).

11. Brass, W., *Methods for Estimating Fertility and Mortality from Limited and Defective Data*, Laboratories of Population Statistics, The University of North Carolina at Chapel Hill, October 1975.

12. Grabill, W.M., and Lee-Jay Cho, *Methodology for the Measurement of Current Fertility from Population Data on Young Children*, Demography 2(1)50-73, 1965.

13. National Academy of Sciences, *Demographic Estimation: A Manual on Indirect Techniques* (forthcoming).

APPENDICES

APPENDIX I

COMPUTATIONAL DETAILS AND EXAMPLES: INDIVIDUAL LEVEL DATA

I.1 INTRODUCTION

The objective of this appendix is to illustrate in detail how an individual woman's contribution to births and to periods of exposure accumulated in the array illustrated in Figures 3 and 4 (page 17) can be computed on the basis of data from her birth and marriage history.

We assume that for each woman dates of the following events are available down to the level of the month:

B,	The woman's birth date;
$B(i)$,	dates of births of her children;
$M(j)$, $D(j)$,	dates of beginning and termination (if applicable) of her marriages;
I,	the date of interview.

For any event say E in the woman's life, its date (also denoted by 'E') is assumed coded in the century-month from:

$$E = 12.E_y + E_m, \qquad (I.1)$$

or conversely, $E_y = \text{Integer}(E/12)$ and $E_m = E - 12.E_y$,

where E_y stands for (the last two digits of) the calendar year, and E_m for the calendar month of occurrence of the event.

The array concerned involves classification of the data in terms of cohorts *(c)* and periods *(p)*; in addition, each cell *(c,p)* is further divided into two parts depending upon whether the woman's age, a, (or marriage duration) equals *(c-p)* corresponding to a "lower triangle" in Figure 3, or equals *(c-p-1)* corresponding to an "upper triangle". We may say that in general a equals *(c-p-k)*; the two cases mentioned above corresponding respectively to $k=0$ and $k=1$, so that *the array involves classification in terms of (c,p,k) with k=0 or 1*. We will assume throughout that ages (and durations) are measured in completed years, and that cohorts and periods refer to single year intervals. The two schemes of classification (see section 2.1) will be described in detail: cohorts and periods defined in terms of completed years before the interview (Scheme 1); or defined as calendar-years (Scheme 2).

To provide an illustrative example, reference will be made throughout to the following hypothetical history of an individual woman.

Event	Month, Year	Century-Month	Event	Month, year	Century-Month
Woman's birth, B	November 1939	479			
			1st marriage, M(1)	July 1957	691
1st birth, B(1)	May 1958	701			
2nd birth, B(2)	August 1961	740			
			End of marriage, D(1)	August 1961	740
3rd birth, B(3)	April 1963	760			
			2nd marriage, M(2)	May 1964	773
4th birth, B(4)	January 1967	805			
5th birth (twin)	" "	"			
6th birth, B(6)	December 1971	864			
			End of marriage, D(2)	October 1973	886
			3rd marriage, M(3) (current)	September 197	933
7th birth, B(7)	September 1977	933			
			Interview, I	June 1980	966

I.2 AGE-SPECIFIC FERTILITY: UNRESTRICTED EXPOSURE

1.2.1 SCHEME 1

Here periods *(p)* are defined as completed years before the interview and numbered sequentially backwards starting with *0*. Cohorts *(c)* refer to women's age in completed years at the time of interview. The basic relationships are:

$$
\begin{aligned}
&p\text{, period of occurrence of any event } E &&= \text{Integer}\left(\frac{I-1-E}{12}\right); \\
&c\text{, woman's cohort or current age} &&= \text{Integer}\left(\frac{I-1-B}{12}\right); \text{ and} \\
&a\text{, her age when } E \text{ occurred} &&= \text{Integer}\left(\frac{E-B}{12}\right).
\end{aligned}
\qquad (I.2)^*
$$

It can be shown that c, p and a defined above satisfy the relation:

$$k = (c-p-a) \qquad (I.3)$$

= 0 (corresponding to a "lower triangle" in Fig. 3),

or 1 (corresponding to an "upper triangle")**.

---

* *Since all dates are assumed coded to the level of the month, the exact date of an event within a given month is ambiguous. For events occurring during the same calendar month, (I.2) assumes that (1) the day of occurrence of event E is after the day of birth of the woman, and (2) the day of interview is before the day of the woman's birth or the day of any other event. In fact we will assume throughout that the interview is held at the beginning of the month, while any other event on the average occurs at the middle of the month. The convention, though arbitrary, reduces the ambiguity in our calculations. At the same time it amounts to rejecting events occurring during the month of interview itself.*

** *The value of $k = (c-p-a)$ depends upon the relationship between E_m, B_m, I_m (month of the event, of woman's birth and of interview). It can be seen from (I.2) or Fig. A.1 that*

i) For $B_m < I_m$: $k=0$ if $I_m > E_m \geq B_m$; $k=1$ if $E_m \geq I_m$ or $E_m < B_m$.

ii) For $B_m \geq I_m$: $k=0$ if $E_m < I_m$ or $E_m \geq B_m$; $k=1$ if $B_m > E_m \geq I_m$.

In our example, $I_m = 6$ (June) and $B_m = 11$ (November); hence case (ii) applies; only two births - Nos. 2 and 7 - satisfy the relation for $k = 1$.

CLASSIFICATION OF BIRTHS INTO THE (c,p,k) ARRAY

The substitution of $B(i)$ for E in equations (I.2) and (I.3) identifies the cell in Fig. 3 or 4 to which that birth belongs.

For the hypothetical history given in the previous section, with I = 966 and B = 479, the woman's cohort (current age) is:

$$c \;=\; \text{Integer}\left(\frac{966-1-479}{12}\right) \;=\; \text{Integer}\left(\frac{486}{12}\right) \;=\; 40.$$

The period of occurrence of each birth and the mother's age at the time of birth are given as follows:

Birth Order	1	2	3	4	5	6	7
$B(i)$, date of birth	701	740	760	805	805	864	933
$p = \text{Int}\left(\frac{965-B(i)}{12}\right)$	22	18	17	13	13	8	2
$a = \text{Int}\left(\frac{B(i)-479}{12}\right)$	18	21	23	27	27	32	37
$k = (40-p-a)$	0	1	0	0	0	0	1

All births in the example belong to the row c=40; for any birth the column is given by p; the second and the seventh births belong to upper triangles, and the rest to lower triangles in Fig. 3.

THE LENGTH OF EXPOSURE

The length of unrestricted exposure during any one year period is, by definition, 12 months per woman. For a woman in cohort c, the total exposure during p can be divided into two parts: e_0 months at age $a_0 = (c-p)$;

and e_1 months at the previous age $a_1 = (c-p-1)$. (In Fig. 3, e_0 is the exposure during a lower and e_1 during an upper triangle).

As illustrated by Fig. A.1, the components (e_0, e_1) are given as follows:

For $B_m < I_m$: $e_0 = (I_m - 1) - (B_m - \tfrac{1}{2}) = I_m - B_m - \tfrac{1}{2}$;

for $B_m \geqslant I_m$: $e_0 = (I_m - 1) + (12 - B_m + \tfrac{1}{2}) = (I_m - B_m - \tfrac{1}{2}) + 12$. (I.4)

And $e_1 = (12 - e_0)$.

The quantities (e_0, e_1) for a given woman depend only on the month of her birth and the interview month, and are constant from one period to another.

In our example $B_m = 11$ and $I_m = 6$, so that

$e_0 = (6-1) + (12-11+\tfrac{1}{2}) = 6\tfrac{1}{2}$ *months*

(from mid November to beginning of June) and

$e_1 = (12-6\tfrac{1}{2}) = 5\tfrac{1}{2}$ *months*

(from beginning of June to mid November).

During p = 15 say, the woman spent $5\tfrac{1}{2}$ months at age (40-15-1) = 24; and spent $6\tfrac{1}{2}$ months at age 25.

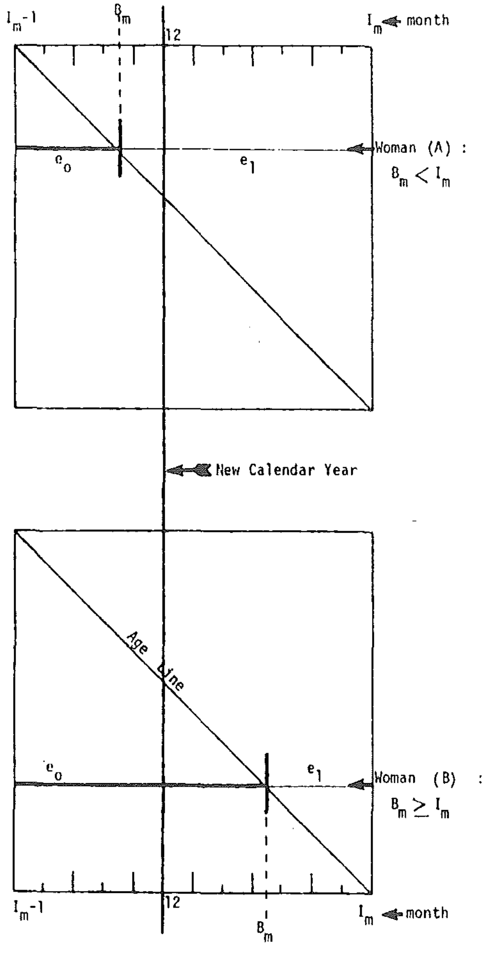


Figure A.1. Components of Exposure During a One Year Period at Two Ages.

1.2.2 SCHEME 2

Here periods refer to calendar years. For any event E occurring in calendar year E_y, we define its period p

$$p = I_y - E_y. \qquad (I.5)$$

A woman's birth cohort (c) refers to the calendar year of her birth, i.e.

$$c = I_y - B_y;$$

and her age when E occurred is given by (I.2) as before*.

In our example (with $I_y = 80$, $B_y = 39$), the woman's birth cohort is

$$c = 80 - 39 = 41, ,$$

and p, a and k for her births are as follows:

Birth Order	1	2	3	4	5	6	7
B_y, Year of birth	58	61	63	67	67	71	77
$p = 80 - B_y$	22	19	17	13	13	9	3
a (as before)	18	21	23	27	27	32	37
$k = (41-p-a)$	1	1	1	1	1	0	1

* *I_y, the year of interview, is a fixed quantity if all women in the sample are interviewed during the same calendar year. If, however, the interviewing spreads over more than one calendar year, we will take I_y in (I.5) as the year in which the last of the interviews took place. In general however we will use I_y to indicate the woman's own interview year, unless stated otherwise.*

Except for the 6th birth, all births belong to upper triangles in Fig. 3 $(k = 1)$.*

As before we divide the 12 months of exposure during any one year period into two parts e_0 at age $(c-p)$ and e_1 at age $(c-p-1)$. For any year *prior* to the year of interview we have

$$e_0 = 12 - B_m + \tfrac{1}{2}, \quad e_1 = 12 - e_0 = B_m - \tfrac{1}{2}. \qquad \text{(I.6)}$$

During the year of the woman's interview, she is not in general exposed for full 12 months. As illustrated by Fig. A.2, (e_0, e_1) for the interview year are as follows:

$$\text{If } I_m > B_m : \; e_0 = (I_m - B_m) \quad \text{and} \quad e_1 = (B_m - \tfrac{1}{2});$$

$$\text{if } I_m \leqslant B_m : \; e_0 = 0 \quad \text{and} \quad e_1 = (I_m - \tfrac{1}{2}).$$

Or written more concisely

$$e_0 = \max(I_m - B_m, 0), \; e_1 = \min(I_m, B_m) - \tfrac{1}{2}. \qquad \text{(I.7)**}$$

In our example, for a year prior to 1980, equation (I.6) gives

---

* *The value of k 1 (c-p-a) depends upon the month of the event* (E_m) *and the month of the woman's birth* (B_m):

$k = 0$ *if* $E_m \geqslant B_m$ *(as for birth 6 in our example)*

$k = 1$ *if* $E_m < B_m$ *(the remaining births in our example)*

** *The function "max" means the larger of the two values in paratheses, and "min" the smaller of the two.*
Note that as an exception, in (I.7), the interview is taken to be on the average at the middle of the month, as is appropriate here.

$$e_0 = 12 - 11 + \tfrac{1}{2} = 1\tfrac{1}{2} \text{ months},$$
$$e_1 = 11 - \tfrac{1}{2} = 10\tfrac{1}{2} \text{ months}.$$

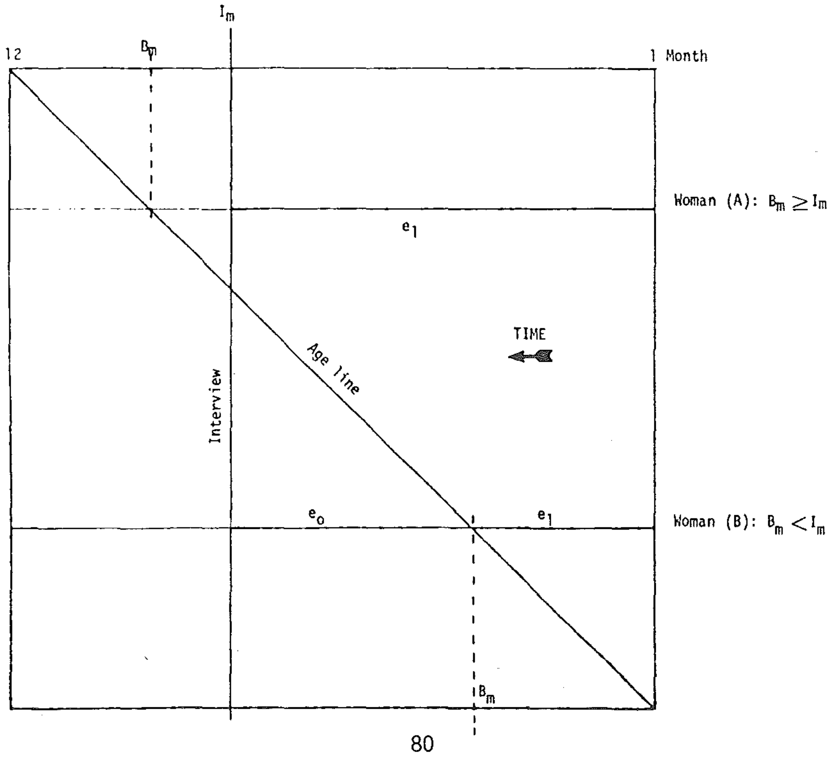
For example during the year 1971 (p = 80-71 = 9) the woman is exposed for 10½ months at age $(c-p-1)$ = (41-9-1) = 31, and for 1½ (from her birthday in mid-November to the end of the year) at age 32.

During the year of the interview (1980, i.e. period p = 0), equation (I.7) gives

$$e_0 = 0, \quad e_1 = 6 - \tfrac{1}{2} = 5\tfrac{1}{2} \text{ months}$$

That is, she is exposed for 5½ months (from the beginning of the year to mid-June) at age $(c-p-1)$ = (41-1) = 40, which is her current age; obviously she has had no exposure at the next age, $(c-p)$ = 41.

Figure A.2. Components of Exposure During Calendar Year of Interview



I.3 AGE-SPECIFIC MARITAL FERTILITY

Here we consider two definitions of exposure: (i) exposure defined as the total time elapsed since first marriage ('ever-married exposure'); and (ii) exposure confined to time spent within marriage or sexual union ('marital exposure').

I.3.1 EVER-MARRIED EXPOSURE

Births are classified into the (c,p,k) array exactly as described above. The only change is that premarital births ($B(i) < M(1)$) are excluded*. There are no such births in our example.

To compute the length of exposure, we will modify quantities e_0 and e_1 (see equations (I.4)-(I.7)) for the restriction that time elapsed following a certain event, E (in this case the date of first marriage) only is counted. Distinction will be made again between the two schemes for defining periods and cohorts.

SCHEME 1

For a given woman, let $e_0(p), e_1(p)$ be the quantities corresponding to (e_0, e_1) in equation (I.4), the former set being restricted to time elapsed since E. By definition, the period during which E occured is

$$p_E = \text{Integer}\left(\frac{I-1-E}{12}\right). \tag{I.8}$$

---

* *See, however, comment (3) in Section 3.3.*

Obviously, during any period more recent than p_E $(p < p_E)$

$$e_0(p) = e_0 \quad \text{and} \quad e_1(p) = e_1, \tag{I.9}$$

while for any period prior to p_E $(p > p_E)$ there is no exposure by definition, i.e.

$$e_0(p) = e_1(p) = 0. \tag{I.10}$$

Now consider period p_E during which the event E occurs. By definition, p_E covers the following 12 months (in century-month code)*:

$$(I-12p_E) - 12 \quad \text{to} \quad (I-12p_E) - 1, \text{ inclusive.}$$

Following event E, the number of months elapsed *within* p_E is

$$X = (I-12p_E-1) - E + \tfrac{1}{2} = (I-12p_E-E) - \tfrac{1}{2}. \tag{I.11}$$

With (e_0, e_1) defined by equation (I.4), two cases can be distinguished (see Fig. A.3).

(i) $X \geqslant e_0$, for which $e_0(p) = e_0$ and $e_1(p) = X-e_0$;

(ii) $X < e_0$, for which $e_0(p) = X$ and $e_1(p) = 0$.

---

* *As elsewhere, we take the interview to be at the beginning of the month, and all other events to be on the average at the middle of the month.*

Or written more concisely

$$e_0(p) = \min(e_0, X), \quad e_1(p) = \max(X - e_0, 0) \quad | \text{for } p = p_E. \qquad (I.12)$$

The three expressions (I.9), (I.10) and (I.12) may be written as a single expression valid for any p

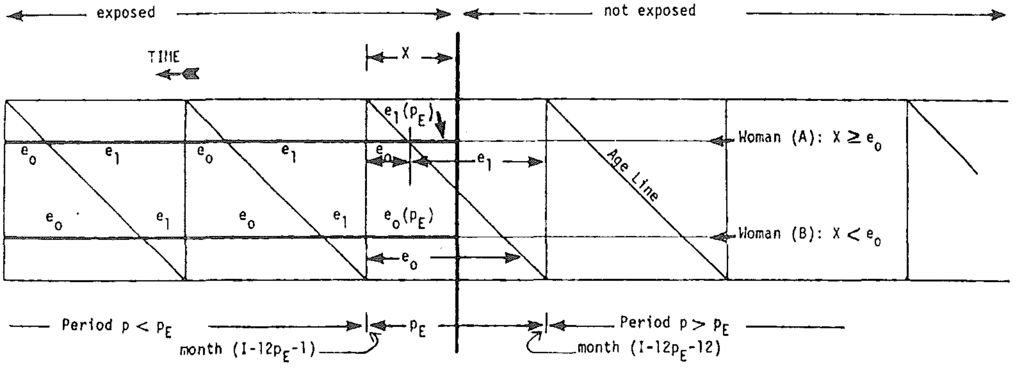
$$\boxed{e_0 = \min\left\{e_0, \max(X, 0)\right\}; \quad e_1 = \max\left\{\min(X, 12) - e_0, 0\right\}} \qquad (I.13)$$

where (e_0, e_1) are defined by (I.4), and X is redefined as

$$X = (I - 12p) - E - \tfrac{1}{2}.$$

Taking $M(1)$ as the event E in (I.13), we obtain a woman's contribution of ever-married exposure in the (c,p,k) array.

Figure A.3. Lexis Diagram Illustrating Components of Exposure Following Event E.



In our example, the period of the woman's first marriage is

$$p_E = \text{Integer}\left(\frac{966-1-691}{12}\right) = 22,$$

the one-year period spanning from the beginning of June (month of interview), to the end of May the following year (1980-22 = 1958).

For any more recent period $(p < 22)$, she is exposed for $e_0(p) = e_0 =$ $6\frac{1}{2}$ months at age ($40-p$) and for $e_1(p) = e_1 = 5\frac{1}{2}$ months at age ($40-p-1$). For any earlier period $(p > 22)$, she has no ever-married exposure.

For p = 22 (the period of her first marriage) the total number of months elapsed following first marriage is from equation (I.11)

$$X = (966-12\text{x}22-691) - \tfrac{1}{2} = 10\tfrac{1}{2} \textit{ months (mid July to May 31).}$$

From equation (I.12), X can be divided into two parts: $e_0(p) = 6\frac{1}{2}$ months (from her birthday in mid-November to 31st May) at age $(c-p)$ = (40-22) = 18; and $e_1(p)$ = 4 months (from her marriage day in mid-July to her birthday in mid-November) at age $(c-p-1)$ = 17.

SCHEME 2

For any calendar year *prior to* I_y, it can be verified that equation (I.13) applies also to Scheme 2, with (e_0, e_1) given by (I.6), and X by

$$X = 12(I_y-p) - E + \tfrac{1}{2}.$$

For the year of the interview $(p = 0)$, equation (I.13) is replaced by

$$e_0(p) = \min(e_0, X), \quad e_1(p) = \min\left\{\max(X - e_0, 0), e_1\right\}, \quad \text{(I.14)}$$

with (e_0, e_1) given by (I.7) and X defined as

$$X = I - E$$

I.3.2 MARITAL EXPOSURE

The only change concerning births is that those occurring outside marriage are excluded. A birth i occurring within marriage j satisfies the condition*

$$M(j) \leqslant B(i) < D(j).$$

In our example, birth number 3 is excluded as it occurs outside marriage.

Exposure within marriage j can be obtained by: (i) computing $e_0(p)$, $e_1(p)$ for the time elapsed since the beginning of the marriage, from an equation such as (I.13) with $E = M(j)$; (ii) computing corresponding quantities for the time elapsed since the termination of marriage, with $E = D(j)$; and (iii) subtracting the latter from the former. For the total length of exposure within marriage, quantities (iii) for individual marriages can be added together.

---

* *See comment (3) in Section 3.3.*

As an illustration consider the woman's exposure during period $p = 2$ (defined according to say Scheme 1, covering the full 12 months June 1977 to May 1978). The quantities $(e_0(p), e_1(p))$ for this period for any of the events $M(1)$, $D(1)$, $M(2)$ or $D(2)$ are the same (being respectively, $6\frac{1}{2}$ months and $5\frac{1}{2}$ months), resulting in no net contribution of the woman's marital exposure. Exposure begins following the third marriage, $M(3)$, during the period concerned. The number of months spent in marriage during $p = 2$ is given by (I.11):

$$X = (966 - 12\text{x}2 - 933) - \tfrac{1}{2} = 8\tfrac{1}{2} \textit{ months (mid October to May 31)},$$

which is divided into two parts (equation (I.12)):

$$e_0(p) = e_0 = 6\tfrac{1}{2} \text{ months of marital exposure at age } (c-p) = 38,$$

and

$$e_1(p) = X - e_0 = 2 \text{ months of marital exposure at age 37.}$$

Suppose now that (altering our example slightly) the third marriage had dissolved in April 1978 (century month, $D(3) = 940$). The number of months elapsed following $D(3)$ during period $p = 2$ is from (I.11):

$$X = (966 - 12\text{x}2 - 940) - \tfrac{1}{2} = 1\tfrac{1}{2} \textit{ months (mid April to May 31)}.$$

Hence corresponding to $D(3)$ we have from (I.12)

$$e_0(p) = X = 1\tfrac{1}{2}, \quad e_1(p) = 0.$$

Subtracting the above from corresponding quantities for $M(3)$ computed earlier gives marital exposure during $p = 2$:

$$e_0(p) \quad = \quad 6\tfrac{1}{2} - 1\tfrac{1}{2} \quad = \quad 5 \text{ months (mid November to mid April) at age 38;}$$

$$e_1(p) \quad = \quad 2 - 0 \quad = \quad 2 \text{ months (mid September to mid November) at age 37.}$$

I.4 DURATION SPECIFIC FERTILITY

The procedure for computing duration specific rates is practically identical to that already described for age-specific rates: birth cohorts are replaced by marriage cohorts *(m)* and retrospective age by retrospective duration since first marriage *(d)*. In the various computational forms given above, a woman's birth date, B, is replaced by the date of her first marriage, $M(1)$. Hence, for the woman in our example (assuming Scheme 1);

Marriage cohort, $m = \text{Integer}(\frac{I-1-M(1)}{12}) = \text{Integer}(\frac{965-691}{12}) = 22;$

duration since first marriage at birth of the first child,

$$d = \text{Integer}(\frac{B(1)-M(1)}{12}) = \text{Integer}(\frac{701-691}{12}) = 0,$$

and the period of occurrence of the birth

$$p = \text{Integer}(\frac{I-1-B(1)}{12}) = \text{Integer}(\frac{965-701}{12}) = 22.$$

Quantities (e_0, e_1) - corresponding to equation (I.4) - refer to the time elapsed since first marriage, i.e. "ever-married exposure". Equation such as (I.13) gives elapsed time conditioned on some other event having been occurred.

I.5 PARITY SPECIFIC EXPOSURE

For parity i specific rate, a woman is exposed during the interval between her $(i-1)^{th}$ and i^{th} births. Equation (I.13) provides the necessary computational form: we compute $(e_0(p), e_1(p))$ with $E = B(i-1)$ and subtract from it the corresponding quantities for $E = B(i)$.

Consider for example parity specific rate for birth order 3 - by say woman's birth cohort defined according to Scheme 1. In our example, the woman has her second birth during period p = 18 (see Section I.2.1) and becomes exposed to the risk of having a birth of order 3. She has her third birth during p = 17 and ceases to be exposed. Hence she is not exposed for any period $p > 18$ or $p < 17$.

During p = 18, the number of months spent at parity 2 is (equation I.11)

$$X = (966 - 12\text{x}18 - 740) - \tfrac{1}{2} = 9\tfrac{1}{2} \text{ months},$$

which is divided into two parts:

$$e_0(p) = 6\tfrac{1}{2} \text{ months (mid November to mid May)}$$

$$\text{at age } (c-p) = (40-18) = 22; \quad \text{and}$$

$$e_1(p) = X-e_0(p) = 3 \text{ months (mid August to mid November)}$$

$$\text{at age } 21.$$

During any period $p \leqslant 17$, months elapsed since attaining parity 2 are obviously

$$e_0(p) = e_0 = 6\tfrac{1}{2}\ ; \quad e_1(p) = e_1 = 5\tfrac{1}{2}.$$

Similarly, we define the number of months spent at parity 3. For periods $p \geqslant 18$, obviously $e_0(p) = e_1(p) = 0$. During $p = 17$, the number of months at parity 3 is

$$X = (966 - 12\text{x}17 - 760) - \tfrac{1}{2} = 1\tfrac{1}{2} \text{ (mid April to May 31)},$$

giving from equation (I.12)

$$e_0(p) = X = 1\tfrac{1}{2} \text{ months at age } (c-p) = 23; \quad e_1(p) = 0.$$

During any period $p < 17$, time elapsed since attaining parity 3 is

$$e_0(p) = 6\tfrac{1}{2}, \quad e_1(p) = 5\tfrac{1}{2} \text{ months}.$$

Subtracting months spent at parity 3 from those at parity 2, we obtain the duration of exposure to the risk of having third birth as follows:

$$p > 18 : \quad e_0(p) = e_1(p) = 0;$$

$$p = 18 : \quad e_0(p) = 6\tfrac{1}{2}, \quad e_1(p) = 3 \text{ months};$$

$$p = 17 : \quad e_0(p) = 6\tfrac{1}{2} - 1\tfrac{1}{2} = 4, \quad e_1(p) = 5\tfrac{1}{2} - 0 = 5\tfrac{1}{2} \text{ months};$$

$$p < 17 : \quad e_0(p) = e_1(p) = 0.$$

Hence the total interval between second and third births

$$B(3) - B(2) = 760 - 740 = 20 \text{ months}$$

is divided into four components:

Period $p = 18$: $e_1(18) = 3$ *months at age* $(c-p-1) = 21$
(mid August to mid November 1961);

$e_0(18) = 6\frac{1}{2}$ *months at age* $(c-p) = 22$
(mid November 1961 to end May 1962);

Period $p = 17$: $e_1(17) = 5\frac{1}{2}$ *months at age* $(c-p-1) = 22$
(June 1 to mid November 1962);

$e_0(17) = 4$ *months at age* $(c-p) = 23$
(mid November 1962 to mid April 1963).

In relation to marital status, parity-specific exposure considered may be (i) unrestricted i.e. without any reference to marital status, or (ii) it may be confined to time elapsed following first marriage or (iii) to time spent within marriage. For reasons noted in Section 5, we will not discuss the last mentioned case.

UNRESTRICTED EXPOSURE

For first birth rates, the starting point of exposure (at parity zero) may be based on some suitably determined minimum age at child bearing. For samples confined to ever-married women, parity specific rates can be computed only on the assumption that the fertility of never-married women is negligible, and that, consequently, these women contribute to exposure only at parity zero. It is also necessary to have data on proportion never-married by current age, for example from the household interview.

If $f(c)$ is the proportion ever-married at the time of the interview (estimated from the household schedule) and $n(c)$ is the number of ever-married women (in the individual interview) then the total exposure (denominator for the rate) at parity zero during any one year period must be augmented by:

$$n(c) . \quad \frac{1-f(c)}{f(c)} \; years$$

EVER-MARRIED EXPOSURE

An alternative way of defining parity-specific exposure is to confine it to the time elapsed following first marriage. This is preferable to unrestricted exposure if the fertility of never-married women is not negligible, and specially if the individual sample is confined to ever-married women. Secondly, with ever-married exposure parity-specific rates can be computed with either form of classification of the data - by age cohort or by marriage cohort. The procedure is identical for all-women and ever-married samples, since only ever-married women appear in the calculation.

If s is the number of premarital births to an (now) ever-married woman, then she contributes to exposure only at parities $i \geq s$. Her ever-married exposure at parity $i = s$ begins at $M(1)$ and terminates at the occurrence of her $(s + 1)th$ birth (if any). Exposure at any parity $i > s$ begins at $B(i)$ and terminates at $B(i + 1)$ or the interview.

I.6 BIRTH INTERVALS

Birth intervals are defined in terms of "fertile pregnancies" which differ from births to the extent multiple births have occurred. In our example

Birth Order	1	2	3	4,5(twins)	6	7
Pregnancy Order, i	1	2	3	4	5	6
Date (century month), $P(i)$	701	740	760	805	864	933

Given that $M(1) = 691$ and $I = 966$, we have

First birth interval: $P(1) - M(1) = 701-691 = 10$ *months.*

Last closed interval: $P(6) - P(5) = 933-864 = 69$ *months.*

Open birth interval: $I - P(6) = 966-933 = 33$ *months*

And, for example, the fifth inter-birth interval:

$$T(5) = P(5) - P(4) = 864-805 = 59 \text{ months}$$

The interval lengths computed above are on the average in rounded (not completed) months.

APPENDIX II

EXAMPLES OF AGGREGATE LEVEL MEASURES

Table 1 gives an example of aggregated live-births and women-years of unrestricted exposure, classified by birth cohorts, periods of occurence and mother's age at child-birth, all in 5-year groups. Cohorts are labelled '0' to '7' corresponding to current ages 10-14 to 45-49; retrospective age-groups are identified in the same way; and periods are labelled '0' to '7' corresponding to 0-4 to 35-39 completed years before the survey. In the table rows correspond to cohorts, columns to periods and diagonal to retrospective age groups. For example, women in cohort '5' (current age 35-39) during period '0' (0-4 years before interview) have had 309 births at ages 30-34 and 223 births at ages 35-39; the corresponding women-years of exposure are 1734 and 1445. For the same cohort, during period '1' (5-9 years before the interview) the number of births is 510 at ages 25-29, and 344 at ages 30-34*.

These data are used in Table 2 to construct cohort-age specific rates. For example, the rate for cohort '5' at age group '4' is

$$\frac{309 + 344}{1734 + 1445} \times 1000 \quad = \quad 205 \text{ births per 1000 women-years}$$

---

* *The data in the example are based on an individual survey confined to ever-married women; all figures in Table 1(B) have been inflated by the proportions ever-married by single years of age at the time of the interview, obtained from the household survey.*

In cummulating these rates along rows (cohorts) of Table 2(B), we multiply by (5/1000) to obtain the mean number of children born by specified age to women in the cohort. For example, for cohort '5' (women aged 35-39), the mean is 1.76 by exact age 25, 3.29 by age 30 and 4.32 by age 35; the last figure for the next older cohort (aged 40-44) is 4.83. Note that the entries in paratheses in the left-most cells are censored by the interview: rates in the corresponding cells of Table 2(A) are operative on the average for 2½ years (rather than 5 years).

The data in Table 1 are used in Table 3 to construct age-period specific rates (conventional ASFRs). For example, at period '0' (0-4 years before the interview) fertility at ages 35-39 is

$$\frac{223 + 245}{1445 + 1594} \times 1000 = 154 \textit{ births per 1000 women-years}$$

In cumulating these rates along columns (periods) of Table 3(B), we multiply by (5/1000) since the data are grouped by 5-years. These columns give the mean number of children born by a specified age to women experiencing the prevailing age-specific rates for a given period. Note that the rates in paratheses in Table 3 (the bottom row) are censored - they are biased towards younger ages within the age group. Further, the data become progressively more incomplete as we proceed further back from the interview.

The bottom figure (4.81) in the first column of Table 3(B) is the Total Fertility Rate for the period 0-4 years before the survey (the effect of censoring is negligible as there is little fertility at ages above 44). For the period 5-9 years before the interview, the prevailing rates

TABLE 1: LIVE-BIRTHS AND WOMEN-YEARS OF EXPOSURE CLASSIFIED BY COHORT, PERIOD AND RETROSPECTIVE AGE IN 5-YEAR GROUPS.

(A) Number of Births

Cohort \ Period	0	1	2	3	4	5	6	7
0	1							
1	6 / 130	1						
2	284 / 476	6 / 119	1					
3	589 / 452	281 / 503	8 / 105	1				
4	391 / 307	478 / 480	212 / 410	6 / 87	2			
5	309 / 223	510 / 344	491 / 460	188 / 362	3 / 68	1		
6	245 / 110	391 / 289	485 / 423	491 / 457	193 / 358	3 / 71	0	
7	80 / 20	245 / 132	375 / 257	456 / 338	426 / 328	124 / 242	3 / 46	4

(B) Women - Years of Exposure

Cohort \ Period	0	1	2	3	4	5	6	7
0	1024							
1	2432 / 2565	2565						
2	2522 / 2283	2522 / 2283	2283					
3	2460 / 2012	2460 / 2012	2460 / 2012	2012				
4	1728 / 1614	1728 / 1614	1728 / 1614	1728 / 1614	1614			
5	1734 / 1445	1734 / 1445	1734 / 1445	1734 / 1445	1734 / 1445	1445		
6	1594 / 1390	1594 / 1390	1594 / 1390	1594 / 1390	1594 / 1390	1594 / 1390	1390	
7	1293 / 1058	1293 / 1058	1293 / 1058	1293 / 1058	1293 / 1058	1293 / 1058	1293 / 1058	1058

Cohort

Period

Retrospective Age

TABLE 2: COHORT-AGE SPECIFIC FERTILITY

(A) Cohort-age specific fertility rates (Unrestricted exposure).

Cohort \ Retrospective Age	7	6	5	4	3	2	1	0
0	1							
1	55	3						
2	221	84	8					
3	235	244	86	7				
4	200	260	266	89	7			
5	159	205	305	268	81	4		
6	84	179	273	316	285	88	3	
7	20	90	214	303	334	284	72	3

(B) Cumulative Cohort fertility - mean number of children ever born by retrospective age.

Cohort \ Retrospective Age	7	6	5	4	3	2	1	0
0	(0)							
1	(.15)	.02						
2	(1.01)	.46	.04					
3	(2.28)	1.68	.46	.03				
4	(3.61)	3.11	1.81	.48	.03			
5	(4.72)	4.32	3.29	1.76	.43	.02		
6	(5.93)	5.72	4.83	3.46	1.88	.45	.02	
7	(6.55)	6.50	6.05	4.98	3.46	1.80	.38	.02

TABLE 3: AGE-PERIOD SPECIFIC FERTILITY

(A) Age-period specific fertility rates (conventional ASFRs).

Retrospective Age \ Period	0	1	2	3	4	5	6	7
0	2							
1	82	1						
2	225	84	2					
3	225	262	85	2				
4	184	296	269	82	1			
5	154	242	311	281	86	1		
6	71	199	297	340	292	73	1	
7	(19)	(125)	(243)	(319)	(310)	(229)	(43)	(4)

(B) Cumulative Period Fertility

Retrospective Age \ Period	0	1	2	3	4	5	6	7
0	0.01							
1	0.42	0.01						
2	1.55	0.43	0.01					
3	2.67	1.74	0.43	0.01				
4	3.59	3.22	1.78	0.42	0.01			
5	4.36	4.23	3.34	1.83	0.43	0.01		
6	4.72	5.42	4.82	3.53	1.90	0.37	0.01	
7	(4.81)	(6.04)	(6.04)	(5.12)	(3.45)	(1.52)	(0.22)	(0.02)

imply that a woman has an average of 5.42 births by age 40, and 6.04 births (except for the censoring effect noted earlier) by age 45.

Finally, Table 4 gives an example of marriage duration - period specific rates. Retrospective marriage durations are given in 5-year groups, while periods are defined in single years before the interview. The figures shown are the number of marital births per 1000 woman-years spent within marriage.

TABLE 4 DURATION-PERIOD SPECIFIC RATES

	Period (Years Before Interview)				
Marriage duration (completed years)	0	1	2	3	4
0-4	331	371	349	349	393
5-9	220	269	310	310	294
10-14	173	164	217	229	231
15-19	126	121	161	181	176
20-24	71	74	108	137	148
25-29	32	67	72	82	105
30+	0	10	24	30	32