WORLD FERTILITY SURVEY
TECHNICAL BULLETINS

APRIL 1980

NO. 6/TECH. 1365

Life Table Analysis

DAVID P. SMITH

INTERNATIONAL STATISTICAL INSTITUTE
Permanent Office · Director: E. Lunenberg
428 Prinses Beatrixlaan, P.O. Box 950
2270 AZ Voorburg
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.

WORLD FERTILITY SURVEY

TECHNICAL BULLETINS

Life Table Analysis

DAVID P. SMITH

APRIL 1980 NO. 6/TECH. 1365

CONTENTS

ACKNOWLEDGEMENTS

Several persons have assisted in the preparation of this document, among whom V. C. Chidambaram and German Rodriguez deserve special thanks. Both helped to smooth and correct the usual obscurity of my writing and Chid, over many months, helped me to expand and clarify the introductory and methodological materials.

1. INTRODUCTION

The origins of life tables rest with the money wagers on annuities and assurances of the late seventeenth century, which required for fairness a reasonably accurate assessment of the prospective lifetimes of the buyers. The formulas that have been developed since that time, and the contributions of sampling theory that began with Greenwood in 1926, continue to find their greatest use in the insurance and health fields, but they suit a variety of other problems equally well. In the WFS core questionnaires life tables can be used to measure infant and child mortality rates, marriage and dissolution-of-marriage rates, the interval between marriage and successive births or between births, and duration of breastfeeding. Where contraceptive histories or abortion are recorded, these may also merit life table treatment.

What life tables do for these analyses is to show event rates at fixed durations of exposure – e.g. the proportion of infants surviving from birth to age 1 month or 1 year, or the proportion of women marrying by age 15, 20, or 25. Other types of rates for these events exist and are as widely used as life table rates, such as the proportion of children who are still living for women in a particular age group, or the proportion ever married among women 15-19, 20-24 or 25-29. Though easier to calculate than life table rates, these give much less information and are often not as precise. Between two populations they may differ because the probabilities of the event occurring are different, or simply because the rates happen to be sensitive to the age distribution and this happens not to be the same for the populations being studied. Continuing with the example of marriage rates, if in one population women 15-19 are mostly close to age 15 and in the other population they are mostly close to age 20, and if women tend to marry in their late teens in both populations, then the proportions ever married can be expected to be quite different in the two cases.

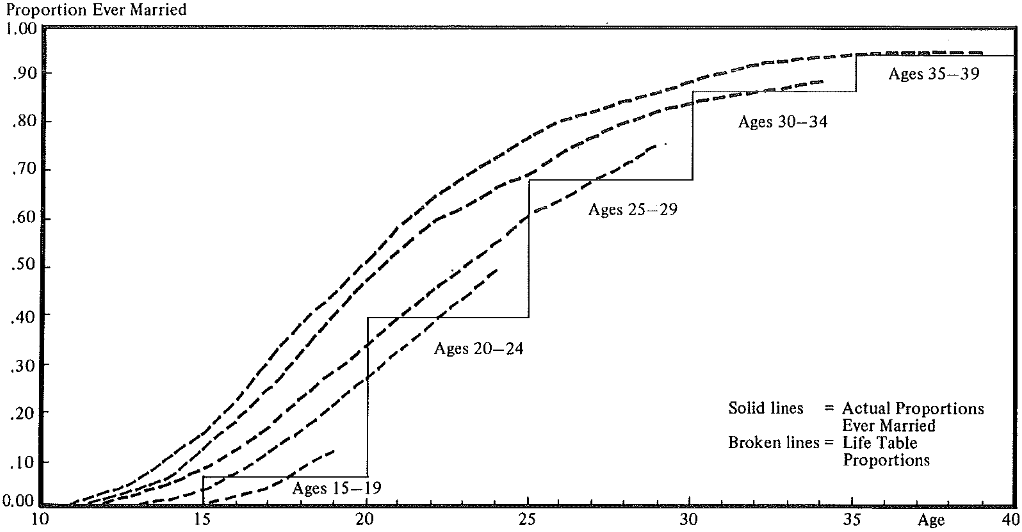
The way life tables get around this problem is to use the data in a different way. Instead of recording only whether women in a particular age group are married or still single, for life tables the ages at which the marriages took place are listed as well. This allows a year by year and age by age accounting, through which overall marriage patterns become clearer. Results will be precise to the extent that current ages and ages at marriage are correctly reported and can be compared with figures for other age groups, other periods of time, and other populations. The figures can be affected by age distributions, but this happens only when ages at marriage are shifting: it is not a general problem in the way the age distribution is for proportions ever married.

An illustration is provided in Figure 1, using the 1975 Sri Lanka Fertility Survey. In this case, age distributions within five-year age groups are reasonably balanced and it is the greater detail of the life table approach that most stands out. Using age at marriage information, the life table is able to show that marriage patterns have been changing, and that these changes have been relatively steady over quite a number of years. For younger age groups, the proportions married at each age are consistently lower. (It is possible that the proportion ever marrying is also changing, but this will not be known until all of the women reach about age 40). Other examples that make use of life tables are scattered throughout this bulletin.

It may be worth emphasizing that life tables are not suitable for all types of events. An age-specific fertility rate, for example, provides a much more compact summary of fertility behaviour than would life tables (which would show durations between births in the same way that they show ages at marriage, but with separate tables for each parity rather than one more useful overall table). The same is likely to be true of the mean number of children ever born or the mean number of living children, and figures like the proportion of women who currently use contraception or have had abortions. In all of these cases life table rates are possible, but besides being retrospective rather than current, the amount of detail they give will be likely to be more than we require and not necessarily easy to condense.

In addition, life tables will not work with variables that are abstract, or for which durations have little effect on the proportions in different categories. Ideals such as those relating to

FIGURE 1. Life Table and Actual Proportions Ever Married, by Age. 1974 Sri Lanka Fertility Survey



desired and optimal family sizes have both of these features. It is hard to determine when they change, or even whether changes are meaningful.

As a general rule, life tables will be appropriate when durations of exposure are central to whether the events being studied have occurred, when the durations are measurable, and when the events themselves are simple and unambiguous. Also, life table analysis is made easier if single and not repeated events are at issue. The types of events that are discussed in this bulletin were selected to meet these criteria.

An introduction to life table methodology is presented at the beginning of Section 2 for readers not already familiar with the mechanics of their construction. The treatment is brief, and assumes the reader has some acquaintance with common demographic notation. (The most important distinction we make is that a left-hand subscript refers to a duration and a right-hand subscript to an exact time or an exact age; thus, N_x and ℓ_x are counts or probabilities at age x or time x, while ${}_1D_x$ and ${}_1p_x$ are counts or probabilities over the interval from x to x+1). The remainder of the Section examines the types of data available in the WFS surveys and discusses variations in the basic life table methodology that suit each of the different data sets.

In Section 3 we illustrate the application of these methods to WFS data on infant and child mortality, age at marriage, marital dissolution and remarriage, birth and pregnancy intervals, and breastfeeding. The examples all draw on the 1975 Sri Lanka Fertility Survey. Where we point to problems that arise in each type of analysis owing to data limitations, the reader should bear in mind that the Sri Lanka survey is of good quality, and the concerns we voice can be expected to apply in equal or greater measure elsewhere.

2. METHODOLOGY

The life tables to be calculated from WFS data combine the experiences of one or more cohorts of women, followed from an initial event to a later termination or to interview. The rates found will be decrement rates in that they measure the duration to a single event, such as the i'th birth, the stopping of breastfeeding, or termination of an interval of contraceptive use. For repeated events, such as sporadic episodes of contraceptive use or multiple marriages interrupted by periods of non-cohabitation, more complex increment-decrement formulas are required. These formulas will not be reviewed here, but may be found in Schoen and Nelson (1974), and Schoen (1975: see also the comment by Rogers and Ledent, 1975). We will also not be concerned with the special problems that arise in life table construction from clinic records. For these the reader may consult Tietze and Lewit (1973) and Jain and Sivin (1977).

2.1 Single Decrement Life Tables

The simplest of cohort life tables require only two pieces of information: (1) the duration between an individual's becoming 'at risk' and experiencing a terminating event, if this has occurred; and (2) the duration between his or her becoming at risk and the interview. Durations are always to be measured in unrounded integer periods i.e. intervals in the range 0.0 to 0.99 are counted as being of length 0, intervals in the range 1.0 to 1.99 are counted as being of length 1, and so forth. We will frequently refer to these as 'completed' periods. 'At risk' will always mean simply that the individual is in the population to which terminating events occur. For example, a woman normally becomes at risk of having a live birth (or, of starting a pregnancy that will produce a live birth) when she marries, and remains at risk until the live birth (or the pregnancy) occurs.

If the interval from entry to interview is the same for all individuals the proportion terminating prior to time x will be

$$1 - \hat{\ell}_x = \sum_{i=0}^{x-1} {}_1D_i / N \quad , \tag{1}$$

where N is the initial sample size and ${}_1D_i$ is the number who terminate in the interval (i, i+1). The summation from i=0 to i=x-1 counts all terminations occurring up to exact time x. The expression is independent of the time units chosen; in what follows, we will nearly always use completed years or completed months.

This expression holds only when the length of observation is the same for all individuals, or when all have experienced the terminating event. Otherwise we require an alternate formula that takes into account the changes in sample size that occur as persons reach interview and cease to be observed. In this case, the data are said to be censored. The formula is:

$$1 - \hat{\ell}_x = 1 - \prod_{i=0}^{x-1} (1 - {}_1D_i^*/N_i^*) \tag{2}$$

The expression constructs $\hat{\ell}_x$ as a function of the monthly probabilities of non-termination, in which N_i^* is the number of individuals observed throughout month i to i+1 and ${}_1D_i^*$ is the number among them who terminate during that month.

The count for each month omits individuals who are reaching interview, which we will designate n_i^*, and those who both terminate and reach interview (d_i^*, a subset of n_i^*). For them the month of interview represents on average about a half-month's observation time, which makes d_i^* something like a half count of monthly terminations, and d_i^*/n_i^* something like a half rate.

(This is most apparent for d_0^* and n_0^*, which represent the experience of individuals interviewed in the same calendar month and year that they became at risk, hence at risk only briefly. The problem, however, carries over to any later months in which some members of the sample arrive at interview, since their exposure in that month will be less than for the rest of the sample[1]).

Expressions (1) and (2) are related, expression (2) being derived from (1) by setting:

$$ {}_1D_i^* = {}_1D_i - d_i^* \qquad (3a)$$

$$ N_i^* = N - \sum_{j=0}^{i} n_j^* - \sum_{j=0}^{i-1} {}_1D_j^* \qquad (3b)$$

When n_i^* and d_i^* are zero for all i, the cumulative termination rate may be found either by summing monthly terminations to the original sample, as in (1), or by cumulating monthly probabilities of avoiding termination for those still at risk and subtracting from unity, as in (2).

An example may help to bring out the equivalence of the two formulas, and to show more clearly what the ℓ_x and $1 - \ell_x$ rates represent. We will suppose we are given the dates of entry, termination and interview listed in columns (1) – (3) of Table 1, and have carried through the subtractions to find the durations to termination and interview shown in columns (4) – (5). The example is hypothetical but might represent times from marriage to first pregnancy, durations of post-partum amenorrhea following a birth, intervals of contraceptive use, or some similar event.

Disregarding for the present cases j-w, for whom intervals to interview are less than 3 months (they will enter in later examples), we regroup individuals according to their durations to termination and interview as shown in Table 2. (In the table, note that over the first 3 intervals

$$ {}_1D_i = {}_1D_i^* \text{ and } N_i^* = N - \sum_{j=0}^{i-1} {}_1D_j) \ .$$

The calculation of survival and termination rates is displayed in Table 3, using both expressions (1) and (2). The results say simply that 2 out of 9 individuals terminate in the first month they are at risk, 5 out of 9 do so in the first two months, and 6 out of 9 terminate in the first three. The survival rate $\hat{\ell}_x$ is thus only 1/3 after three months at risk. If the tables were of durations from marriage to first pregnancy, this would be the proportion not yet pregnant after three months of marriage; for amenorrhea it would be the proportion who had not yet resumed menstruation at three months after childbirth; for contraception it would be the proportion who remained active contraceptive users three months after having begun. Survival is interpreted in each case as remaining in the initial category.

If in place of cases a-i we bring in the entire sample from Table 1, allowing numbers to decrease as the individuals reach interview, we will have the distributions and rates of Tables 4 and 5. All rate calculations are now by expression (2), since for (1) we would need a constant N.

The handling of n_i^* cases, persons not terminating prior to month i and reaching interview in (i, i+1), is brought out by Table 4. As was noted earlier neither they nor the subsample d_i^* enter into rate calculations for the interval. [The reader will not have difficulty interpreting the Table if he notes that the number of persons observed at the start of interval i to i+1 will

---

[1] An alternative to omitting incomplete intervals is to weight them according to time spent at risk, which for N_i^* is the full month and for n_i^* perhaps half that. Then:

$$ 1 - \hat{\ell}_x = 1 - \prod_{i=0}^{x-1} [1 - ({}_1D_i^* + d_i^*) / (N_i^* + \tfrac{1}{2} n_i^*)] $$

This solution, and some others, are explored in Appendix 2. (Incomplete intervals will not pose difficulties in life table Method III, which does not overlap individuals with different durations of exposure. See p. 13).

be $N_i^* + n_i^*$ and is composed of those persons observed for the whole of the previous interval and not terminating during it, $N_{i-1}^* - {}_1D_{i-1}^*$].

TABLE 1

Case	Date of Entry	Date of Termination	Date of Interview	Duration to Termination	Duration to Interview
a	1.1.79	10.2.79	24.5.79	1.30 months	4.77 months
b	18.1.79	15.5.79	18.5.79	3.90	4.00
c	19.1.79	–	9.5.79	–	3.67
d	22.1.79	29.2.79	12.5.79	1.23	3.67
e	9.2.79	17.5.79	24.5.79	3.27	3.50
f	30.1.79	12.2.79	13.5.79	0.40	3.43
g	5.2.79	30.2.79	17.5.79	0.83	3.40
h	5.2.79	20.4.79	10.5.79	2.50	3.17
i	6.2.79	18.5.79	10.5.79	1.40	3.13
j	28.2.79	–	23.5.79	–	2.83
k	9.3.79	–	25.5.79	–	2.53
l	10.3.79	8.5.79	12.5.79	1.93	2.07
m	29.3.79	–	28.5.79	–	1.97
n	21.3.79	–	14.5.79	–	1.77
o	18.3.79	23.3.79	10.5.79	0.17	1.73
p	12.4.79	–	19.5.79	–	1.23
q	4.4.79	9.5.79	10.5.79	1.17	1.20
r	25.4.79	17.5.79	29.5.79	0.73	1.13
s	4.4.79	–	6.5.79	–	1.07
t	29.4.79	–	27.5.79	–	0.93
u	18.5.79	–	28.5.79	–	0.33
v	19.5.79	–	28.5.79	–	0.20
w	12.5.79	–	16.5.79	–	0.13

TABLE 2

Interval Span	Total	Number of Cases Observed Throughout Interval: Number not Terminated as of Start of Interval	Number Terminating During Interval	Number Terminating in This or Prior Intervals
(i, i+1)	N	N_i^*	${}_1D_i$	$\sum_{j=0}^{i} {}_1D_j$
0-1	9 (a-i)	9 (a-i)	2 (f,g)	2 (f,g)
1-2	9 (a-i)	7 (a-e,h,i)	3 (a,d,i)	5 (a,d,f,g,i)
2-3	9 (a-i)	4 (b,c,e,h)	1 (h)	6 (a,d,f-i)
3-4	2 (a,b)			

TABLE 3

Interval Span	No. of Cases Observed Throughout Interval:[1]				RATES: Monthly Termination	Monthly Survival	Cumulative Termination [Expression (2)]	Cumulative Termination [Expression (1)]
(i,i+1)	N	N_i^*	${}_1D_i^*$	ΣD_j	${}_1D_i^*/N_i^*$	$1-{}_1D_i^*/N_i^*$	$1-\Pi(1-{}_1D_i^*/N_i^*)$	$\Sigma_1 D_j/N$
0-1	9	9	2	2	$\frac{2}{9}$	$\frac{7}{9}$	$1 - \frac{7}{9} = \frac{2}{9}$	$\frac{2}{9} = .22$
1-2	9	7	3	5	$\frac{3}{7}$	$\frac{4}{7}$	$1 - (\frac{7}{9})(\frac{4}{7}) = \frac{5}{9}$	$\frac{5}{9} = .56$
2-3	9	4	1	6	$\frac{1}{4}$	$\frac{3}{4}$	$1 - (\frac{7}{9})(\frac{4}{7})(\frac{3}{4}) = \frac{6}{9}$	$\frac{6}{9} = .67$
3-4	2							

[1] In this Table, ${}_1D_i = {}_1D_i^*$

2.2. Multiple Decrement and Cause-Deleted Tables

With a small increase in complexity expressions (1) and (2) can be rewritten to distinguish more than one type of termination, in *multiple decrement* tables. For this, total terminations (${}_1D_i^*$) are separated into sub-categories ${}_1D_{i,1}^*$, ${}_1D_{i,2}^*$. . ., such that

$$\sum_j {}_1D_{i,j}^* = {}_1D_i^*$$

The cumulative termination rate for each sub-category will be

$$\hat{\tau}_{x,j} = \sum_{i=0}^{x-1} \hat{\ell}_i \; {}_1D_{i,j}^*/N_i^* \qquad (4)$$

where each of the terms in the summation is the incremental proportion terminating from the j'th cause. The increments are additive, so that

$$\hat{\ell}_x = 1 - \sum_j \hat{\tau}_{x,j}$$

For a single type of termination the expression becomes another way of writing (1) or (2).

TABLE 4

Interval	Number of Cases Observed Throughout Interval: Number Not Terminated as of Start of Interval	Number Terminating in This Interval	Number of Cases Observed During Part of Interval only: Total	Number Terminating in This Interval
(i, i+1)	N_i^*	${}_1D_i^*$	n_i^*	d_i^*
0-1	19 (a-s)	4 (f,g,o,r)	4 (t-w)	0
1-2	10 (a-e,h-1)	4 (a,d,i,l)	5 (m,n,p,q,s)	1 (q)
2-3	4 (b,c,e,h)	1 (h)	2 (j,k)	0
3-4	1 (b)	1 (b)	2 (c,e)	1 (e)
4-5	0	0	0	0

TABLE 5

Interval Span (i,i+1)	No. of Cases Observed Throughout Interval: N_i^*	$_1D_i^*$	RATES Monthly Termination $_1D_i^*/N_i^*$	Monthly Survival $1-{_1D_i^*}/N_i^*$	Cumulative Termination [Expression (2)] $1-\Pi(1-{_1D_i^*}/N_i^*)\doteq 1-\hat{\ell}_{i+1}$
0-1	19	4	$\frac{4}{19}$	$\frac{15}{19}$	$1-(\frac{15}{19})$ = .21
1-2	10	4	$\frac{4}{10}$	$\frac{6}{10}$	$1-(\frac{15}{19})(\frac{6}{10})$ = .53
2-3	4	1	$\frac{1}{4}$	$\frac{3}{4}$	$1-(\frac{15}{19})(\frac{6}{10})(\frac{3}{4})$ = .64
3-4	1	1	1	0	$1-(\frac{15}{19})(\frac{6}{10})(\frac{3}{4})(0)$ = 1.0

Cause-deleted life tables provide a way of estimating the sub-category rate $\hat{r}_{x,j}$ as it would appear if one or more of the other sub-categories, *not correlated with* $\hat{r}_{x,j}$, did not operate. (Examples would be the duration of breastfeeding in the absence of infant mortality, which seems to be distinct from other factors influencing breastfeeding intervals; or accidental pregnancy as a cause of contraceptive termination among couples not intending further births, a problem distinct from other medical or nuisance factors associated with discontinuation). The rates are found by transferring the types of termination that are to be disregarded from the category $_1D_i^*$ into n_i^* in expression (2). Under the reclassification both time spent at risk and the incidence of relevant terminating events remain correctly specified. Interpretation of the rates, however, requires care, since the possibility that the transferred individuals are particularly susceptible or particularly immune to the risks under study is not taken into account. If they are, an element of ambiguity is introduced in what is intended to be a simple concept. This consideration tends to limit the utility and use of cause-deleted rates.

2.3 Sampling Errors of the Estimates

Sampling variances of life table survival and termination rates can be estimated using Greenwood's formula:

$$\text{Var}(\hat{\ell}_x) = \text{Var}(1-\hat{\ell}_x) = \hat{\ell}_x^2 \sum_{i=0}^{x-1} \text{Var}({_1\hat{p}_i})/{_1\hat{p}_i}^2 , \qquad (5a)$$

where ${_1\hat{p}_i} = 1 - {_1D_i^*}/N_i^*$.

Under simple random sampling this becomes

$$\text{Var}(\hat{\ell}_x) = \hat{\ell}_x^2 \sum_{i=0}^{x-1} (1-{_1\hat{p}_i})/({_1\hat{p}_i}\, N_i^*) \qquad (5b)$$

$$= \hat{\ell}_x^2 \sum_{i=0}^{x-1} {_1D_i^*} / [N_i^* (N_i^* - {_1D_i^*})] \ . \qquad (5c)$$

(The variance of ${_1\hat{p}_i}$ is ${_1\hat{p}_i}(1-{_1\hat{p}_i})/N_i^*$).

For stratified random sampling of paired cluster units, used in a number of WFS surveys, the variance of ${_1\hat{p}_i}$ is estimated from the clusters. Letting h and h′ represent paired cluster units, for Var $(\hat{\ell}_x)$ we will have:

$$\text{Var}(\hat{\ell}_x) = \hat{\ell}_x^2 \sum_{x=0}^{x-1} \frac{1}{N_i^{*2}} \sum_h [{_1D_{i,h}^*} - {_1D_{i,h'}^*} - (1-{_1\hat{p}_i})(N_{i,h}^* - N_{i,h'}^*)]^2 \qquad (5d)$$

in self-weighted samples; and

$$\mathrm{Var}(\hat{\ell}_x) = \hat{\ell}_x^2 \sum_{i=0}^{x-1} \frac{1}{N_i^{**2}} \sum_h W_h [{}_1D_{i,h}^* - {}_1D_{i,h'}^* - (1 - {}_1\hat{p}_i')(N_{i,h}^* - N_{i,h'}^*)]^2 \quad (5e)$$

in samples with stratum weights W_h. The terms N_i^{**} and ${}_1\hat{p}_i'$ in the weighted sample variance are:

$$N_i^{**} = \sum_h W_h \ (N_{i,h}^* + N_{i,h'}^*) \ ,$$

$${}_1\hat{p}_i' = \sum_h W_h \ ({}_1D_{i,h}^* + {}_1D_{i,h'}^*) \ / \ N_i^{**} \ .$$

For the derivation of these estimates the reader may consult Potter (1969, pp. 476-477), Kish (1965, pp. 193-197), and Kish, Groves and Krotki (1976, pp. 16-19). The latter is a WFS Technical Bulletin.

2.4 Estimation from Retrospective Data (Methods I and II)

In the examples given previously the durations at risk, or intervals between entry into risk and termination or interview, were assumed to be known in completed months or years. Whenever this is the case (that is, when precise estimates of completed months or years can be made) we refer to the tables that are calculated as Method I life tables.

WFS data on first marriage ages and on child mortality are mostly of this kind. Since in both cases individuals can be thought of as being at risk from birth, the duration to interview in completed years will be the same as age at last birthday (for deceased children, age the child would be if still living). Similarly, the duration to termination will be the age at marriage or death. For infants, ages in months are used in place of ages in years[1].

For first marriage, we define the categories:

N_i^*	=	All women currently age i+1 and above who were not married before age i
${}_1D_i^*$	=	All women currently age i+1 and above who married at age i
n_i^*	=	All women currently age i who were not married before age i
d_i^*	=	All women currently age i who married at age i

The categories N_i^*, n_i^* will include never-married women, usually enumerated in the WFS household schedules. If their numbers are not available, the tables that are found will be for ever-married women only and this should be stated in the table headings[2]. (Tables for sub-categories of women, such as by educational levels, will frequently be of this type since the relevant data may not exist for single women).

For child mortality, ages have been grouped in some of the WFS surveys and interval widths must be interpreted accordingly. There is also the problem that children are listed by date of birth rather than age at interview, age being found by subtraction of the birth date from the interview date. This introduces an ambiguity: a child whose i'th birthday falls in the interview month may be either age i or age i-1, depending on the days of birth and interview, which are not stated. To avoid biases from this source, all births less than i years + 1 month before interview need to be excluded from the sample on which survival from birth to age i (or incremental survival from age i - α to age i) is being calculated. The categories to be formed are:

---

[1] In some of the WFS surveys, children's ages at death are given in grouped categories (i.e. under 1 month, 1-2.9 months, 3-5.9 months, and so forth). Provided that ages at death are not mis-stated, the use of variable intervals does not bias the estimated life table rates. Where the number of categories is small, however, the advantage of life tables over cross tables diminishes. An effective use of the latter approach is found in Somoza (1980).

[2] Such tables can be highly misleading. For an example, see Trussell (1980).

N_i^* = All children born at least i+1 years + 1 month before interview who did not die before age i

$_1D_i^*$ = All children born at least i+1 years + 1 month before interview who died at age i

n_i^* = All children born between i years + 1 month and i+1 years + 1 month before interview who did not die before age i

d_i^* = All children born between i years + 1 month and i+1 years + 1 month before interview who died at age i.

Where the data given to us are calendar dates rather than ages or durations, and intervals must be found by subtraction, the life table rates that are derived will be referred to as Method II rates. WFS data on marital dissolutions and remarriages and on birth intervals are of this type. To construct tables, we subtract the date of entry for the particular risk from the dates of termination and interview. This gives us the categories N_i^*, $_1D_i^*$, n_i^* and $_1d_i^*$ as in Method I tables, but with a critical distinction: subtraction of dates gives an approximate interval length that is sometimes wrong. For example, while the interval 0 to 1, corresponding to entry and termination or entry and interview in the same month, is always less than 30-31 days; the interval 1 to 2 may represent from 1 day (last day of month i to first day of month i+1) to about 60 days (first day of month i to last day of month i+1); the interval 2 to 3 from about 31 days (last day of month i to first day of month i+2) to about 90 days (first day of month i to last day of month i+2); and so forth.

If days of entry are uniformly distributed during each month, and if the true survival and termination rates are linear functions of the duration of exposure [i.e. if $\ell_x = a + bx$ and $1 - \ell_x = (1 - a) - bx$], it happens that the lack of precision with which the time intervals have been measured will not bias the rate estimates $\hat{\ell}_x$, $1 - \hat{\ell}_x$ of expressions (1) and (2). However, owing to the shorter first interval (interval 0), each $\hat{\ell}_x$ will correspond to a mid-month or mid-interval rate $\ell_{x-½}$ rather than to ℓ_x.

In the WFS, days of entry will normally be randomly distributed but the true event rate will be nonlinear, with the result that the correspondence of $\hat{\ell}_x$ to $\ell_{x-½}$ will only be approximate. Even so, beyond the first few intervals biases will be trivial in nearly all cases and need not be a matter of concern (this is demonstrated in Appendix 1). Linear interpolation between the calculated rates $\hat{\ell}_x$ and $\hat{\ell}_{x+1}$ may be used to find a close approximation to the correct whole-interval rate ℓ_x.

We may repeat the example of the previous section using approximate instead of exact intervals to see the differences that result. For years and months only, omitting days, we have the dates and intervals shown in Table 6. Continuing as in Tables 4 and 5, in Table 7 the categories N_i^*, $_1D_i^*$, n_i^*, and d_i^* are formed, and in Table 8 monthly and cumulative termination rates are derived. Although the correspondence between Tables 1 and 6 is not particularly close, the cumulative termination rates representing $1-\ell_{½}$ and $1-\ell_{1½}$ found in Table 8 (.10 and .31, respectively) and $1-\ell_1$ and $1-\ell_2$ found in Table 5 (.21 and .53) are not out of line with each other. Rates for later intervals compare poorly, but will not concern us as sample sizes are particularly small[1].

2.5 Estimation from Current Status Data (Method III)

In calculating life tables it is not necessary to make use of as much information as Methods I and II have called for. An alternative, helpful when termination dates and duration of use tend to be reported inaccurately, is to construct tables based only on year and month of entry into

---

[1] We will not always want to create tables using the smallest intervals the data allow, in this case months, and longer intervals are easily constructed. For widths $y \geqslant 2$ months we divide the durations calculated in Table 1 or 6 by y, dropping fractions so that intervals will be in completed units as before. The remaining steps are unchanged except that the intervals are now of a different size; the estimated rates being for periods ℓ_y, ℓ_{2y}, ℓ_{3y}, . . .under Method I, and $\ell_{y-½}$, $\ell_{2y-½}$, $\ell_{3y-½}$ under Method II.

TABLE 6

Case	Date of Entry	Date of Termination	Date of Interview	Duration to Termination	Duration to Interview
a	1.79	2.79	5.79	1 month	4 months
b	1.79	5.79	5.79	4 months	4
c	1.79	–	5.79	–	4
d	1.79	2.79	5.79	1	4
f	1.79	2.79	5.79	1	4
e	2.79	5.79	5.79	3	3
g	2.79	2.79	5.79	0	3
h	2.79	4.79	5.79	2	3
i	2.79	5.79	5.79	3	3
j	2.79	–	5.79	–	3
k	3.79	–	5.79	–	2
l	3.79	5.79	5.79	2	2
m	3.79	–	5.79	–	2
n	3.79	–	5.79	–	2
o	3.79	3.79	5.79	0	2
p	4.79	–	5.79	–	1
q	4.79	5.79	5.79	1	1
r	4.79	5.79	5.79	1	1
s	4.79	–	5.79	–	1
t	4.79	–	5.79	–	1
u	5.79	–	5.79	–	0
v	5.79	–	5.79	–	0
w	5.79	–	5.79	–	0

TABLE 7

Interval Span	Number of Cases Observed Throughout Interval: Number Not Terminated as of Start of Interval	Number Terminating in This Interval	Number of Cases Observed During Part of Interval Only: Total	Number Terminating in This Interval
$(i, i+1)$	N_i^*	${}_1D_i^*$	n_i^*	d_i^*
0-1	20 (a-t)	2 (g,o)	3 (u-w)	0
1-2	13 (a-f,h-n)	3 (a,d,f)	5 (p-t)	2 (q,r)
2-3	6 (b,c,e,h-j)	1 (h)	4 (k-n)	1 (l)
3-4	2 (b,c)	0	3 (e,i,j)	2 (e,i)
4-5	0	0	2 (b,c)	1 (b)

TABLE 8

Interval Span	No. of Cases Observed Throughout Interval:		RATES Monthly Termination	Monthly Survival	Cumulative Termination [Expression (2)]
(i,i+1)	N_i^*	${}_1D_i^*$	${}_1D_i^*/N_i^*$	$1-{}_1D_i^*/N_i^*$	$1-\Pi(1-{}_1D_i^*/N_i^*) \doteq 1-\ell_{i+½}$
0-1	20	2	$\frac{2}{20}$	$\frac{18}{20}$	$1-(\frac{18}{20}) = .10$
1-2	13	3	$\frac{3}{13}$	$\frac{10}{13}$	$1-(\frac{18}{20})(\frac{10}{13}) = .31$
2-3	6	1	$\frac{1}{6}$	$\frac{5}{6}$	$1-(\frac{18}{20})(\frac{10}{13})(\frac{5}{6}) = .42$
3-4	2	0	0	1	$1-(\frac{18}{20})(\frac{10}{13})(\frac{5}{6})(1) = .42$

the population at risk and status as of interview. This permits estimation of ℓ_x for intervals $x \geqslant 1$ as:

$$\hat{\ell}_x = 1 - {}_xD_0{}^{(x)} / N_0{}^{(x)}, \qquad (6)$$

where $N_0{}^{(x)}$ represents all women becoming at risk x months prior to interview and ${}_xD_0{}^{(x)}$ is the subset of these women who had terminated by the time of the interview. As with method II tables, interval widths are measured by subtraction of dates[1].

The formula differs from our earlier ones in that each rate is derived only from individuals who became at risk in a particular month, usually a small sample[2], and each counts terminations up to interview. We show in Appendix 1 that the expression yields rates $\hat{\ell}_x \doteq \ell_x$.

Current status life tables can be illustrated using our earlier example. In Table 9 dates of entry and interview are displayed and intervals to interview calculated as in Table 6, but in place of the termination dates we have substituted status as of interview (i.e. whether or not each individual had terminated). Table 10 then finds cumulative termination rates, which are simply the proportions in the successive duration groupings who have already terminated. The cost in sample size when Method III is used is immediately apparent when the N's of this Table are compared with those of Table 8.

[1] Cases whose durations are 0, for whom intervals are very short, are not used.

[2] Small sample sizes can best be dealt with by grouping calendar dates of entry, termination and interview into y-month blocks *before* subtracting to find the intervals. This produces rates ℓ_y, ℓ_{2y},. . .

TABLE 9

Case	Date of Entry	Date of Interview	Duration to Interview	Whether Terminated
a	1.79	5.79	4 months	Yes
b	1.79	5.79	4 months	Yes
c	1.79	5.79	4	No
d	1.79	5.79	4	Yes
f	1.79	5.79	4	Yes
e	2.79	5.79	3	Yes
g	2.79	5.79	3	Yes
h	2.79	5.79	3	Yes
i	2.79	5.79	3	Yes
j	2.79	5.79	3	No
k	3.79	5.79	2	No
l	3.79	5.79	2	Yes
m	3.79	5.79	2	No
n	3.79	5.79	2	No
o	3.79	5.79	2	Yes
p	4.79	5.79	1	No
q	4.79	5.79	1	Yes
r	4.79	5.79	1	Yes
s	4.79	5.79	1	No
t	4.79	5.79	1	No
u	5.79	5.79	0	No
v	5.79	5.79	0	No
w	5.79	5.79	0	No

TABLE 10

Duration to Interview (Interval) x	Number of Cases Observed for This Duration $N_0(x)$	Number Terminated ${}_xD_0(x)$	Cumulative Termination Rate [Expression (6)] $1-{}_xD_0(x)/N_0(x) \doteq \ell_x$
0	not used	not used	—
1	5 (p-t)	2 (q,r)	$\frac{2}{5}$
2	5 (k-o)	2 (l,o)	$\frac{2}{5}$
3	5 (e,g-j)	4 (e,g-i)	$\frac{4}{5}$
4	5 (a-d,f)	4 (a,b,d,f)	$\frac{4}{5}$

3. APPLICATIONS

In this section we outline the types of analysis to which life tables can contribute in the WFS. A number of warnings and qualifications are included, which concern problems and limitations that arise from the quality of WFS data or the manner in which particular items have been recorded. Users should give these some attention when preparing their own analyses.

Several tables from the 1975 Sri Lanka Fertility Survey have been included for illustration. Sampling errors are only sometimes shown; owing to the large number of cluster pairs (303) used in the Sri Lanka survey these will always be very close to simple random sampling estimates.

3.1 Infant and Child Mortality

Deaths are not events that respondents are comfortable recalling, or that interviewers are comfortable probing. In consequence they are not usually well reported. Whenever possible estimates from the WFS should be cross-checked against independent sources, both to gauge their quality and as a possible means of correcting age-specific or other mortality estimates when the omissions are serious[1].

TABLE 11. Proportion of Children Dying at Ages Under 1 and Under 5 by Year of Birth and Birth Order. 1975 Sri Lanka Fertility Survey

Year of Birth	Birth Order				Total
	1	2	3	4+	
		Ages Under 1			
1945–1949	.114	.100	.136	.082	.109
1950–1954	.068	.061	.054	.053	.060
1955–1959	.063	.058	.060	.074	.065
1960–1964	.074	.054	.046	.064	.061
1965–1969	.051	.062	.056	.061	.059
		Ages Under 5			
1945–1949	.151	.158	.167	.150	.155
1950–1954	.115	.101	.113	.112	.110
1955–1959	.091	.096	.086	.112	.099
1960–1964	.088	.069	.070	.094	.085
1965–1969	.074	.078	.076	.089	.083

Children's deaths in the Sri Lanka survey appear to have been well reported (Table 11). Omissions or misclassifications appear likely for births in the late 1940s and for low parity births before 1955, but subsequent rates are credible. [For comparison with the 1960-1964 and 1965-1969 rates, Keyfitz and Flieger (1971, pp. 378-381) estimate the 1963 proportions dying by ages 1 and 5 as .058 and .092, respectively; and the 1967 proportions as .042 and .066. The figures, which derive from Department of Census and Statistics tabulations, accord satisfactorily with the WFS results]. Standard errors for 1955+ are on the order of .005 – .007, and nothing will be said of fine differences between rates except that the general patterns by parity are reasonably consistent through time.

This Table was calculated by Method I, using the categories defined on page 11.

---

[1] Even when they are wrong the rates can be used as an index of reliability for reporting of earlier vs. more recent events. Making the assumption that mortality has been constant or improving in nearly all countries, we should expect it to show a reverse trend only in proportion as events in more distant periods have been passed over. A discussion of methods for estimating the quality of child mortality data is given in Somoza (1980).

3.2 Age Patterns of Marriage and Marriage Dissolution

For first marriage, life tables are calculated by Method I from the distribution of single women by age at interview and ever-married women by ages at marriage and interview, beginning from exact age 10 or 15 and continuing to about 30 or 35. The most useful piece of information these rates will provide is whether the age at marriage is falling, stable, or rising; a matter of particular relevance in developing countries for which early marital age-specific fertility rates tend to be high (c. 400). Marriage life tables, separated by cohorts as far as the data permit, are able both to detect and to quantify such changes, and may suggest their forward projection. In so doing, they permit refinement of hypothetical future fertility trajectories to accommodate likely effects of delayed childbearing as well as of changing family sizes. [A computer program, NUPTIAL, that utilises the Coale (1971) marriage function to generate completed marriage patterns has been devéloped by Rodriguez and Trussell (1979). For associated fertility patterns, the reader may consult Coale and Trussell (1974)].

These comments assume that the reporting of marriages is approximately correct. It sometimes is not, as when consensual unions are prevalent or when interviewers and respondents have manipulated dates to enclose premarital pregnancies or to cover lapses of memory. If this is suspected, the credibility of the given marriage data can be at least roughly ascertained from the levels of marital age-specific fertility they suggest, and also by the proportion of first births falling within the first year of marriage. The latter figure should not be above 60 per cent, Sheps' estimate (1965, p. 72) for an American Hutterite population. (Sri Lanka figures are shown in Table 13).

Rates of separation or divorce, widowhood, and remarriage are found by Method II, with entry from date of marriage for marital dissolution tables and from date of marriage dissolution for remarriage tables. These events have less demographic relevance than first marriages because they affect relatively small numbers of women, and older women disproportionately. They tend as well to be subject to omissions and incorrect dating.

For Sri Lanka, where rates of marital dissolution are low, we find a decrease since the 1940's in proportions widowed (Table 12: standard errors in this table are on the order of .01 for total rates less than .1, and on the order of .02 – .03 for higher rates). Rates of separation and divorce are seen to differ by age at marriage, women marrying at ages 25-29 being at a disadvantage relative to those who marry younger, but no certain trend appears in the rates over time at this level of aggregation. (Dissolution rates are sharply higher among marriages occuring in the 1970's, not shown in the Table. The rise appears to be largely confined to women under 25 at interview and is independent of their ages at marriage).

TABLE 12. First Marriage Dissolutions by Marriage Cohort and Age at Marriage. 1975 Sri Lanka Fertility Survey

Marriage Cohort	Age at Marriage	Proportion Terminated within							
		5 years		10 years		15 years		20 years	
		Wid-owed	Separated Divorced	Wid-owed	Separated Divorced	Wid-owed	Separated Divorced	Wid-owed	Separated Divorced
1940's[1]	15-19	.02	.03	.04	.04	.06	.06	.08	.06
	20-24	.02	.03	.05	.03	.07	.05	.07	.05
1950's[1]	15-19	.02	.02	.03	.03	.04	.05	.08	.06
	20-24	.01	.03	.04	.04	.07	.06	.13	.07
	25-29	.02	.07	.05	.10	.07	.10	.10	.12
1960's	15-19	.01	.03	.03	.05				
	20-24	.01	.03	.02	.05				
	25-29	.01	.05	.03	.08				

[1] Restricted to women under age 50 at interview.

In assessing remarriage, the main demographic concern will be wife's age at event, which in developing countries largely governs the potential for further childbearing. Rates for Sri Lanka are moderately high (cf., Smith, 1980).

3.3 Pregnancy and Birth Intervals

The interval from marriage to first pregnancy or birth is subject to the biases of marriage reporting, and both these and other invervals are subject to biases arising from birth misplacement or omissions. Particularly subject to error are early events to older respondents, especially where children have died. Because of this, age-specific and life table fertility rates are usually not traced back further than about 10 years.

For these and more recent events a good check can be found in surveys done prior to the WFS when the events were current. Within WFS child mortality rates, sex ratios at birth, and changes in the life table intervals of low-order births also can be used to estimate the quality of reporting of progressively earlier events.

In the analysis of first pregnancy and first birth intervals our main interest will be with timing, since decisions to postpone childbearing usually represent a break between older cultural values that emphasize immediate childbearing, and the exigencies of newer status patterns. For later births, changes in timing overlap with changes in completed family sizes and need to be considered together. Both types of changes appear in life tables as differences in the cumulative proportions of couples reaching each parity within a stated time, measured either from a given age, a lower parity, or marriage. A check for the validity of recent changes can be made by separating wives who have worked or who have used contraception, and examining the extent to which the remaining sample displays the older patterns.

Intervals between marriage and first live birth in Sri Lanka are shown in Table 13 according to marriage cohort and age at marriage. A clear increase can be seen in the proportion of wives having births within 1 and 2 years of marriage since the 1940's, probably a result of reporting errors in the older birth data. Across all periods wives married at 20-24 have had their first births sooner after marriage than those married at 15-19 or 25-29, though by 10 years of marriage fewer of the younger wives remain childless. Some adolescent sterility is implied in these findings, as seems reasonable. The results have not been controlled for marital dissolutions; this would be done through two factor net rates (for marriage dissolution or birth as terminating events) or cause-deleted rates (with birth as the terminating event and marriage dissolutions handled as arrivals at interview, that is, entered in n_i*).

TABLE 13. Proportion of Wives Having First Live Birth by Duration Since Marriage, Marriage Cohort and Age at Marriage. 1975 Sri Lanka Fertility Survey

Marriage Cohort	Age at Marriage	Proportion Having First Birth within: 1 year	2 years	5 years	10 years
1940's[1]	15-19	.34	.72	.94	.96
	20-24	.40	.75	.95	.96
1950's[1]	15-19	.35	.70	.93	.98
	20-24	.42	.78	.94	.97
	25-29	.32	.70	.87	.92
1960's	15-19	.38	.74	.94	.98
	20-24	.45	.77	.92	.96
	25-29	.41	.75	.89	.92

[1] Restricted to women under age 50 at interview.

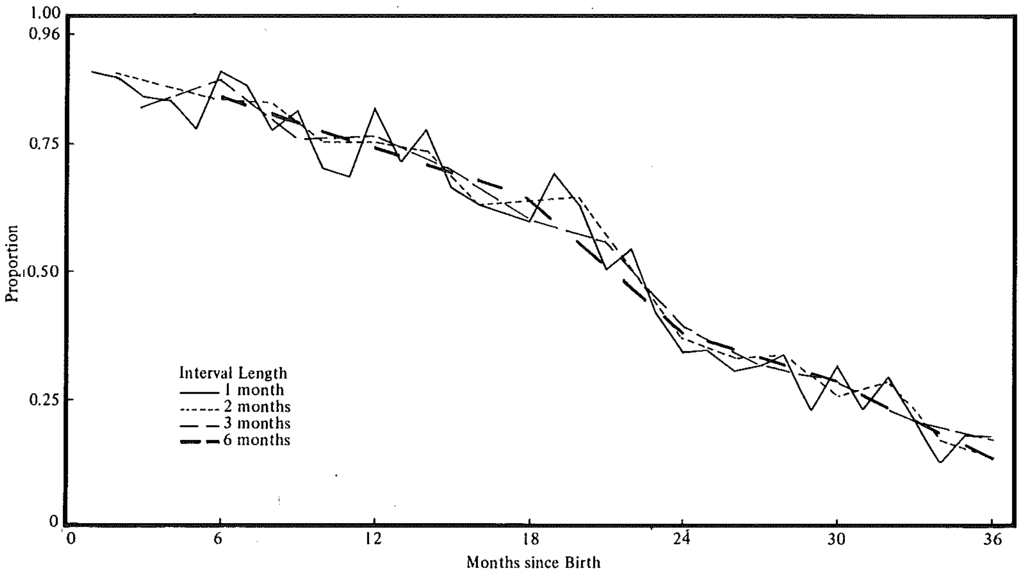
An important application of life table pregnancy rates for very high fertility areas is in the estimation of fecundability, for which see Sheps and Menken (1973, pp. 129-130, 398-399). A WFS Illustrative Analysis focusing on birth intervals has been written by Rodriguez and Hobcraft (1980: see also Hobcraft and Rodriguez, 1980).

All of these tables follow Method II.

3.4 Duration of Breastfeeding

Breastfeeding intervals are exceptionally prone to heaping biases, with 12, 18, 24 or 30 months commonly favored even when this is longer than the duration since the birth to which they refer. Method II life tables could only be made from these data after reassignment of durations obviously wrong, and would contain a further bias common to tables derived from 'most recent' event (as breastfeeding has been coded): this is that breastfeeding intervals are known only for some births – those of women who have had at most one subsequent birth. Because of this our sample, as we move further back in time from the interview, gradually shifts from high fertility to lower fertility women. The breastfeeding experience of the two groups is unlikely to be the same.

FIGURE 2. Proportion of Women Currently Breastfeeding by Duration since Birth, Wives 15-44 at Event. 1975 Sri Lanka Fertility Survey.



Method III life tables, which derive from all births at the different durations before interview, avoid both telescoping and latest-event biases. (Method III tables remain subject to serious error, as do Method II tables, when dates of birth are imputed). They yield the results shown in Figure 2 and Table 14 for Sri Lanka[1]. Note that in counting the number of women breastfeeding at interview (column 3), it is essential not to confuse births; a woman may or may not be currently breastfeeding her latest child, she cannot be breastfeeding an earlier child. Further detail on breastfeeding is provided in Lesthaeghe and Page (1980), a WFS Illustrative Analysis.

---

[1] Women pregnant at interview are excluded, since it is not made clear in the Sri Lanka recode tape whether they have continued to breastfeed their latest child. Owing to their absence the rates shown in the Table are slightly biased. The Table also does not distinguish whether breastfeeding was discontinued because of child mortality. To eliminate this cause, births at time x would be counted only if the child had survived to interview.

TABLE 14. Proportion of Women Currently Breastfeeding by Duration Since Birth, Wives Ages 15-44 At Event. 1975 Sri Lanka Fertility Survey

Duration Since Birth (Months)	Number* of Women	Number Breastfeeding	$\hat{\ell}_x$	Standard Error
1	97	87	.892	.030
2	113	99	.881	.030
3	105	89	.840	.036
4	103	86	.834	.037
5	116	91	.780	.040
6	113	101	.893	.030
7	104	90	.864	.034
8	89	69	.775	.042
9	105	86	.815	.037
10	107	75	.702	.042
11	70	48	.686	.055
12	87	71	.816	.043
13	81	58	.715	.049
14	74	58	.780	.045
15	79	53	.665	.054
16	80	51	.633	.052
17	88	54	.616	.052
18	99	60	.599	.042
19	81	56	.690	.056
20	77	48	.628	.054
21	85	43	.504	.057
22	76	41	.545	.056
23	68	28	.415	.061
24	68	24	.346	.062
25	87	30	.347	.053
26	59	18	.307	.060
27	96	30	.317	.049
28	95	32	.338	.045
29	75	17	.230	.049
30	101	32	.313	.044
31	88	20	.231	.047
32	110	32	.295	.044
33	92	20	.216	.044
34	65	8	.126	.042
35	83	15	.183	.041
36	92	16	.178	.039
37	98	8	.082	.027
38	78	10	.123	.036
39	102	12	.122	.031
40	103	15	.143	.033
41	113	13	.119	.031
42	98	12	.118	.032
43	82	10	.117	.039
44	84	7	.084	.033
45	74	8	.106	.036
46	99	2	.022	.015
47	107	3	.031	.016
48	92	5	.049	.022

**Currently pregnant women are omitted.*

As the Figure and the standard errors in the Table make clear, the small sample sizes of Method III create problems of their own. Smoothing or grouping, as in the Figure, will be required before the results are used. For further analysis the data could be separated into only a handful of sub-categories, and this greatly restricts their contribution. Note, however, that the smoothed tables are helpful in searching out bias in larger sample tables constructed by Method II, and may permit more extensive use of tables of that type.

3.5 Further Analysis

None of the examples we have given have been presented with more than cursory attention to intervening variables, several of which are central. Women's ages at event, durations of marriage, numbers of living children and residence are familiar intermediate variables, as also are labor force participation, occupational status and education (when groupings beyond the lowest are reasonably represented) and sometimes religion. Contraception and abortion, as means to an end, are worth investigating if sample sizes permit when birth intervals are being compared. The reader can add to this list.

For whatever analyses are planned, the reader should keep in mind that life table analysis requires certain minimum sample sizes. As Figure 2 illustrates, $\hat{\ell}_x$ rates can fluctuate substantially for numbers of cases near or below 100, and this figure will commonly be a good cut-off. This means starting with substantially larger numbers if cases are expected to be lost by arrival at interview.

When in doubt as to sample sizes, (5b) can be hand-calculated as a first approximation to the sample variance if (5d) is unknown.

REFERENCES

Breslow, N. and J. Crowley, 1974. "A large sample study of the life table and product limit estimates under random censorship". *Annals of Statistics 2:* 437-453.

Coale, Ansley J. 1971. "Age patterns of marriage". *Population Studies 25:* 193-214.

Coale, Ansley J. and James Trussell. 1974. "Model fertility schedules: Variations in the age structure of childbearing in human populations". *Population Index 40:* 185-258.

Hobcraft, John and German Rodriguez. 1980. *Methodological issues in the life table analysis of birth intervals.* World Fertility Survey, London (forthcoming).

Jain, Anrudh K. and Irving Sivin. 1977. "Life table analysis of IUD's: Problems and recommendations". *Studies in Family Planning 8:* 25-47.

Kish, Leslie. 1965. *Survey Sampling.* John Wiley & Sons, New York.

Kish, Leslie, R. M. Groves and L. P. Krotki. 1976. *Sampling Errors for Fertility Surveys.* World Fertility Survey, Occasional Paper No. 17, London.

Keyfitz, Nathan and Wilhelm Flieger. 1971. *Population: Facts and Methods of Demography.* W. H. Freeman & Co., San Francisco.

Lesthaeghe, Ronald and Hilary Page. 1980. *Breastfeeding patterns in Nepal.* An illustrative analysis. World Fertility Survey, London (forthcoming).

Potter, Robert G. 1969. "Use-effectiveness of intrauterine contraception as a problem in competing risks". Pp. 458-484 in Ronald Freedman and John Y. Takeshita: *Family Planning in Taiwan.* Princeton University Press, Princeton, New Jersey.

Rodriguez, German and John Hobcraft. 1980. *Life table analysis of birth intervals in Columbia: An illustrative analysis.* World Fertility Survey, London (forthcoming).

Rodriguez, German and James Trussell. 1979. *Maximum likelihood estimation of the parameters of Coale's model nuptiality schedule.* World Fertility Survey, London.

Rogers, Andrei and Jacques Ledent. 1976. "Increment-decrement life tables: A comment". *Demography 13:* 287-290.

Schoen, Robert. 1975. "Constructing increment-decrement life tables". *Demography 12:* 313-324.

Schoen, Robert and Verne E. Nelson. 1974. "Marriage, divorce and mortality: A life table analysis". *Demography 11:* 267-290.

Sheps, Mindel C. 1965. "An analysis of reproductive patterns in an American isolate". *Population Studies 19:* 65-80.

Sheps, Mindel C. and Jane A. Menken. 1973. *Mathematical Models of Conception and Birth.* University of Chicago Press, Chicago.

Somoza, Jorge L. 1980. *Illustrative analysis of infant and child mortality in Columbia.* World Fertility Survey, London.

Smith, David P. 1980. *Marriage dissolution and Remarriage in Sri Lanka and Thailand.* World Fertility Survey, London.

Trussell, James. 1980. *Age at first marriage in Sri Lanka and Thailand.* World Fertility Survey, London.

APPENDIX I

LIFE TABLE SURVIVAL RATES DERIVED FROM APPROXIMATE INTERVALS

In constructing life tables it is usual to have available the calendar dates (day, month, year) of entry into the at risk population, of termination or death, and of interview. Having these, essentially correct intervals of survival and unbiased monthly and yearly survival rates can be derived. When only the year and month of entry, termination and interview are known both sets of rates become less certain. Unbiased rates, it will turn out, can be made from this more restricted information only when the actual rate of survival or non-termination (ℓ_x) is a linear function of the duration of exposure; that is, when $\ell_x = a + bx$ for all x. In one other case, when survival takes the exponential form $\ell_x = ae^{bx}$, a simple bias correction can be introduced; and for a number of common distributions biases will exist but will be non-trivial only in early months, permitting adjustments of later intervals to be omitted.

All of these results assume that entries into the population at risk are spread uniformly through each month (as against being bunched toward selected weeks) and require that events in open intervals (the categories n_i* and d_i* defined earlier) not be used in the survival rate calculations.

We begin with the case $\ell_x = a + bx$[1]. If entry into the population at risk is uniformly distributed throughout month 0 [i.e. in the interval (0.1)], then the proportion continuing to the start of month x among all entrants, say $f(\ell_x)$, will be

$$f(\ell_x) = \begin{cases} \int_0^x \ell(x-t)\,dt = \int_0^x [a + b(x-t)]\,dt = ax + bx^2/2, & x \leqslant 1 \\ \int_{x-1}^x \ell(t)\,dt = \int_{x-1}^x (a+bt)\,dt = a + b(x - ½) = \ell_{x-½}, & x \leqslant 1. \end{cases}$$

These expressions find mean survival between two timepoints for individuals entering the at-risk population at time 0. To measure survival from an interval to a point, as we intend, is the reverse of this problem and formally indentical[2].

The result $f(\ell_x) = \ell_{x-½}$ for $x \geqslant 1$ informs us that when survival is a linear function of the duration of exposure, ℓ_x rates based on approximate intervals will be unbiased estimates of the true mid-interval rates $\ell_{x-½}$. [The reader may have guessed from our earlier discussion of Method II life tables that this would be at least approximately the case, since the first interval (interval 0 in our earlier notation) spans 0 to 30 days while all subsequent intervals span 0 to 60 days. The mean span is one-half month for the first interval and one month for all others. As terminations, unlike entries, can be arbitrarily distributed, only sometimes – i.e. when ℓ_x is linear – will rates derived from them happen to fall exactly at mid-intervals].

---

[1]This case implies a force of mortality $\mu_x = 1/(\omega - x)$, where $\omega = a/b$ and is the maximum duration that any member of the population survives.

Mathematically, $\mu_x = -\frac{d}{dx}\ln \ell_x$

[2]Readers with a demographic background may note the $f(\ell_x) = {}_1L_{x-1}$ and that the discussion which follows in the main text is concerned with the estimation of ℓ_x values from ${}_1L_x$. A useful general approximation formula, derived by a Taylor expansion of ${}_1L_x$, is:

$\ell_x = \frac{13}{24}({}_1L_{x-1} + {}_1L_x) - \frac{1}{24}({}_1L_{x-2} + {}_1L_{x+1})$.

This is the same as linear interpolation if the ℓ_x distribution is linear and slightly better if it is not.

The linear case can be generalized by setting $\ell_x = a + bx^c$ to generate concave or convex distributions. By about the sixth month, distributions for which $0 < c < 2$ display biases in timepoint location and survival probability that rarely reach 0.1 per cent. As $c >> 2$ however, which corresponds to an initial period of very high survival followed by one of catastrophic mortality, biases in the later survival rates can be on the order of several per cent. (In general, approximate interval data will not yield suitable estimates of survival where abrupt changes in patterns occur).

The distribution $\ell_x = 2b/(e^{cx} + e^{-cx}) = b \operatorname{sech}(cx)$ is concave from below at points near the origin and convex at farther distances, corresponding to distributions in which the hazard function or force of mortality rises sharply at first and subsequently stabilizes (Figure 2 is of this general shape). It yields rates whose accuracy is comparable to that for $\ell_x = a + bx^c$, given c in the same ranges.

To correct bias in these examples is tedious, requiring several iterations between timepoint estimates and regression coefficients. Our confidence that the ℓ_x distribution has been correctly specified will not usually be great enough to justify the effort required, even with respect to early intervals where the initial approximation $f(\ell_x) = \ell_{x-1/2}$ is least adequate.

The distribution $\ell_x = a\,e^{bx}$ (for which $\mu_x = -b$, a constant) is more tractable. For entrants uniformly distributed in the interval (0,1) we have:

$$f(\ell_x) = \begin{cases} \int_0^x a e^{b(x-t)}\,dt = \frac{a}{b}(e^{bx}-1) \ , & x \leqslant 1, \\ \\ \int_{x-1}^x a\,e^{bt}\,dt = \frac{e^b-1}{b}\,[\,a e^{b(x-1)}\,], & x \leqslant 1. \end{cases}$$

For $x \geqslant 1$ we find the point x^* such that $f(\ell_x) = \ell_{x^*}$ by setting

$$\ell_{x^*} = a\,e^{bx^*} = \frac{e^b-1}{b}[\,a e^{b(x-1)}] = f(\ell_x)\ ,$$

from which

$$\begin{aligned} x^* &= x - 1 + \ln[(e^b - 1)/b]\,/\,b \\ &= (x-1)^* + 1. \end{aligned}$$

All intervals in the exponential are one month in width except the first, which is

$\ln[(e^b-1)/b]\,/b = x_1{}^*$

The fact that the rates $f(\ell_x)$ fall one month apart with respect to each other, and are displaced with respect to the origin by the amount $x_1 - x_1{}^* = \frac{1}{2} - x_1{}^*$, implies that the regression coefficient $\hat{b}$ will be an unbiased estimator of b, while $\hat{a}$ will be biased owing to the displacement of the complete $f(\ell_x)$ series. The difference term $\frac{1}{2} - x_1{}^*$ can be used to find an unbiased estimator $\hat{a}$, which is:

$$\hat{a}' = \hat{a}\,e^{\hat{b}\,(\frac{1}{2}-x_1{}^*)} = \hat{a}\,\hat{b}\,/\,(e^{\frac{1}{2}\hat{b}} - e^{-\frac{1}{2}\hat{b}}).$$

As for other distributions, the correction is non-trivial only for intervals near the first.

These expressions, which relate to Method II life tables, have assumed that only part of the sample reach interview in a given month, and have avoided blending the short observation times in the interview month for these individuals with the full-month observation times for individuals reaching interview in later months. In Appendix 2 a reconciliation of observation times is considered which estimates full-month survival rates for persons reaching interview during the month. We now consider the different option that is available when the duration of observation is the same for all members of the sample, as is assumed in Method III life tables.

In this case the survival rate to the start of the interview month for the sample will be, as before:

$$f(\ell_x) = \begin{cases} \int_0^x \ell(x-t)\,dt, & x \leqslant 1 \\ \\ \int_{x-1}^x \ell(t)\,dt, & x \geqslant 1 \end{cases}$$

If, like entry into the at risk population, interviews are uniformly distributed during the month, then survival as of the interview date will equal

$$g(\ell_x) = \begin{cases} \int_0^x \frac{1}{\delta} \int_0^\delta \ell(t)\,dt\,d\delta, & x \leqslant 1 \quad \text{and month of entry = month of interview} \\ \\ \int_{-1}^0 \int_{x-1}^x \ell(t+\delta)\,dt\,d\delta, & x \geqslant 1 \quad \text{and month of entry = month of interview} \end{cases}$$

To find the point x^* such that $g(\ell_x) = \ell_{x^*}$, it is necessary as before to specify the form of the ℓ_x distribution. For the case $\ell_x = a + bx$, where month of entry precedes month of interview $g(\ell_x) = a + bx = \ell_x$, and therefore $x^* = x$. That is, with entry and interview in different months, with both uniformly distributed, and with survival a linear function the proportion surviving at interview $g(\ell_x)$ will be the same as the proportion surviving at x[1].

For the exponential distribution $\ell_x = ae^{bx}$, with entry and interview not in the same month we have

$g(\ell_x) = ae^{b(x-1)}[(e^b - 1)/b]^2 = ae^{bx^*}$

$x^* = x + \left\{(2/b)\ \ln\left[(e^b - 1)/b\right] - 1\right\}$.

In the region $0 < b \leqslant -.01$ the quantity in braces is negligible; and in the region $-.01 \leqslant b \leqslant -.1$, which brings in most cases of interest, it is near $b/12$.

As before, the quantities $\hat{a}, \hat{b}$ can be found by regression and $\hat{b}$, which will be unbiased, can be used to correct the bias in $\hat{a}$. The required expression is

$\hat{a}' = \hat{a}\, e^{\hat{b}\,(1 - x_1^*)} = \hat{a} e^{\hat{b}} / [(e^{\hat{b}} - 1)/\hat{b}]^2$.

Errors for $\ell_x = a + bx^c$ and $\ell_x = 2b / (e^{cx} + e^{-cx})$ will be on the order of those found previously.

---

[1] For month of entry = month of interview, $g(\ell_x) = ax + bx^2/4$. At $x = 1$, $g(\ell_x) = \ell_{1/4}$, but this result is not particularly helpful since we can rarely be confident of linearity in this region .

APPENDIX II

RATE ESTIMATES FROM EVENTS IN THE INTERVIEW MONTH

To approximate whole-month rates from survival and termination during the month of interview, we assign a half month of observation time either to all cases reaching interview (in our earlier notation n_i^*) or to those reaching interview who have not yet terminated ($n_i^* - d_i^*$).

Justification for the latter technique, the more common, is essentially pragmatic; the rate $d_i^*/[d_i^* + ½ (n_i^* - d_i^*)$ is usually a close approximation to the full month termination rate D_i^*/N_i^* and never exceeds 1.0. Merging these two sets of observations yields the combined monthly termination estimate

$$q_i = \frac{D_i^* + d_i^*}{N_i^* + d_i^* + ½ (n_i^* - d_i^*)} = \frac{D_i}{N_i - ½ (n_i^* - d_i^*)}$$

For this estimate to be unbiased with respect to D_i^*/N_i^* it is necessary that the survival rate for the complete month among n_i^* cases be the product of their survival rates during both its observed and unobserved segments. Letting d_i^{**} represent terminations falling after interview but within the calendar month or calendar interval, we have

$$1 - \frac{d_i^*}{d_i^* + ½ (n_i^* - d_i^*)} = \left(1 - \frac{d_i^*}{n_i^*}\right)\left(1 - \frac{d_i^{**}}{n_i^* - d_i^*}\right).$$

Solving for d_i^{**} we find

$$d_i^{**} = d_i^* \frac{(n_i^* - d_i^*)}{(n_i^* + d_i^*)}$$

For comparison, in the linear distribution $\ell_x = a + bx$, $d_i^{**} = d_i^*$; and in the exponential $\ell_x = ae^{bx}$, $d_i^{**} = d_i^* (n_i^* - d_i^*) / n_i^*$. Terminations in both are thus less sharply skewed toward the earlier part of each interval than is assumed in the standard method of half-month corrections.

Half-month corrected rates will be unbiased only when this condition is met. It cannot usually be assumed to hold *a priori*[1], and for that reason we have avoided using such rates.

---

[1] To continue our discussion in Appendix I, an unbiased half-month correction for linear ℓ_x distributions would be $q_i = (D_i^* + d_i^*) / (N_i^* + ½n_i^*)$, the form shown in footnote 1 on page 8. It assigns a half-month of observation time to all persons reaching interview. For exponential distributions $q_i \doteq [D_i^* + d_i^*] / [N_i^* + ½ (n_i^* + ⅓d_i^*)]$. A note on these distributions will be found in Breslow and Crowley (1974).