DEMOGRAPHY© Volume 21, Number 4 November 1984

CALCULATION OF LIFE TABLES FROM SURVEY DATA: A TECHNICAL NOTE

Noreen Goldman
Office of Population Research, Princeton University, Princeton New Jersey 08544
Anne R. Pebley
Office of Population Research, Princeton University, Princeton, New Jersey 08544
Graham Lord
Morgan Stanley and Company, Inc., New York, New York

Abstract—Life table calculations from survey data are frequently based on events for which exact dates are not available. When these dates are coded in monthly form (e.g., century months), estimates should take into account the fact that the first duration interval—the interval which captures events occurring in the first month of exposure—is half the length of all remaining intervals. Although failure to do so has a trivial effect on many demographic calculations, estimates which are based on events which occur with high frequency in the first few months of exposure can be substantially biased. Estimates of fecundability for four countries in the World Fertility Survey are used to illustrate this bias.

In recent years, the quantity of survey data available for analyzing demographic events has grown enormously. This dramatic increase is in large part due to the World Fertility Survey, which includes national fertility surveys from approximately 60 developed and developing countries. Accompanying this growth has been an increase in use of life table methods. Demographers have used life tables for several centuries to analyze mortality and for several decades to look at topics such as contraceptive efficacy and employment. Increasingly, life tables are now being used for the analysis of other phenomena such as childspacing, breastfeeding, marriage patterns, fecundability, and migration (see, for example, Smith, 1980). Part of this change may be due to (a) current research interests which can often be addressed more readily with survey data; (b) the fact that life tables can handle censoring problems which almost always arise in the use of survey data; and (c) the increasing use of computers.

In general, life table calculations require information on dates of events, in order to calculate lengths of exposure and age or duration at the time of the event. In many demographic surveys, these dates are recorded in monthly intervals. Most commonly, a century-month scheme is employed: e.g., the month of January 1900 is labelled as "1" and each subsequent month is labelled accordingly. Because information as to the exact dates of events has been lost by this type of coding procedure, most estimates of the frequency of vital events incorporate the assumption that events occur in the middle of the calendar month. This assumption is satisfactory for most purposes, especially since in many developing countries respondents would be unlikely to supply accurate information on exact dates. However, for the analysis of events which occur at short durations, a life table calculation based on this assumption can be substantially biased.

In order to illustrate various types of life table calculations, we consider, as an example, the estimation of fecundability—the monthly probability of conception in the absence of fertility control. When estimating fecundability from survey data, most techniques are based on the waiting time between marriage and the first birth. A common method involves a life table calculation of the probability of a conception occurring according to the number of months of exposure to conception. For the purposes of this paper, we consider the start of exposure to be the date of first marriage including both legal marriages and consensual unions. Thus, we assume that there is no premarital intercourse. Duration is calculated as the difference (in numbers of months) between the century month of the date of first fertile conception[1] (i.e., nine months before the date of first birth), and the date of the onset of marriage.

The first problem in life table construction results from those cases for whom the century month of marriage is the same as the century month of conception—duration equal to zero according to the assumption that events occur midmonth. Clearly we should not include these births in the numerator of a life table estimate if we attribute zero exposure to the denominator. The second problem results from the fact that, on average, all events occur at a duration whose length is an integral number of months. For example, women who become pregnant in the calendar month subsequent to the month of marriage have, on average, one month of exposure. According to a standard life table format with intervals 0–1 month, 1–2 months, 2–3 months, etc., the pregnancy could fall into either the interval (0,1) or the interval (1,2).

Consider the first month of exposure, i.e., the calendar month of marriage. Suppose that the reported month is January 1980 and let us assume that marriage (but not conception) occurs in the middle of the month—January 15, 1980. Then, during the month of January, women who do not conceive have, on average, only 15 days or a *half* of a month of exposure to the risk of conception. If we assume that conception occurs uniformly over the period *subsequent* to marriage, women who conceive during the same month as marriage have only a *quarter* of a month of exposure prior to conception, on average. This result may be surprising since we know that women who have a marriage and a conception during the same calendar month could have anywhere between zero exposure—if the two events occurred simultaneously—and one month of exposure—if marriage occurred on January 1, and conception on January 31.

Now consider women who have a conception in the month subsequent to marriage—February 1980. They could have little more than zero exposure if marriage occurred at the end of January and pregnancy at the beginning of February, or almost as much as two months. On average, they would have one month of exposure, half of which would be spent in January and half in February. Women who have not conceived by the end of February would be exposed for one-and-a-half months, half of a month during January and the entire month of February. The same type of calculation pertains to all subsequent months.

In summary, exposure to the risk of conception during the calendar month of marriage is, on average, half as long as exposure during all subsequent months. Hence, the correct construction of a life table by duration of marriage should be defined by the following labelling of intervals: 0–0.5 months, 0.5–1.5 months, 1.5–2.5 months, etc., with the corresponding probabilities of conception ${}_{.5}q_0$, ${}_1q_{.5}$, ${}_1q_{1.5}$, etc. With this format, no change is required in the basic calculation of life table values. Women are included in the denominators of q_x values up to and including the duration in which they conceive. By the very definition of the intervals, women will contribute half the amount of exposure to the first interval as compared with all subsequent intervals. Note that if one were to calculate average women-months of exposure in each interval (e.g., for the calculation of M_x or e_x values) women with pregnancies in the month of marriage would contribute roughly one-quarter of a month of exposure (half of the exposure in the interval [0,0.5]) and women with pregnancies in the month subsequent to marriage would have their pregnancies roughly halfway through the interval (.5, 1.5), as desired. The only problem with this calculation is that the final results are for nonstandard intervals. For example, we do not readily have the probability of a conception occurring within one, two or three months of marriage.

There are a number of ways to obtain life table values, all of which involve interpolation. Smith (1980) chooses to linearly interpolate between l values (e.g., l_1 is calculated as the average of $l_{.5}$ and $l_{1.5}$). A number of more sophisticated solutions are presented in actuarial textbooks since the problem occurs frequently in the development of mortality and disability tables (e.g., Batten, 1978; Spiegelman, 1973).

The approach which we have used in our analysis consists of interpolating the q_x values in the defined intervals 0.5–1.5, 1.5–2.5, ..., by means of the following equivalent formulae:[2]

$$_{1-t}q_{y+t} = (1 - t)_1q_y \quad 0 \leq t \leq 1 \quad (1)$$

and

$$_tq_y = \frac{t_1q_y}{1 - (1 - t)_1q_y}, \qquad 0 \leq t \leq 1. \quad (2)$$

Equation (1) [or (2)] is known as the Balducci hypothesis, after the Italian actuary, although its first published use seems to be by Ansell (1874). Under the Balducci hypothesis, the hazard function (μ_x) is a decreasing hyperbola over each interval (Batten, 1978). Although not satisfactory for many kinds of mortality analyses, a decreasing hazard function is applicable to most demographic processes which are concentrated at early durations of exposure—e.g., fecundability, infant mortality and divorce.[3]

According to the law of total probability, the probability of a pregnancy occurring in the interval x to $x + 1$ equals

$$_1q_x = {}_{.5}q_x + (1 - {}_{.5}q_x)\ {}_{.5}q_{x+.5} \quad (3)$$

For the first interval, we have

$$_1q_0 = {}_{.5}q_0 + (1 - {}_{.5}q_0)\ {}_{.5}q_{.5} \quad (3a)$$

Note that $_{.5}q_0$ is obtainable directly from our life table calculation and that $_{.5}q_{.5}$ can be estimated from our calculated value of $_1q_{.5}$ by evaluation of (2) for $y = t = 0.5$. The resulting equation for $_1q_0$ is:

$$_1q_0 = {}_{.5}q_0 + \frac{(1 - {}_{.5}q_0)(1/2\ {}_1q_{.5})}{(1 - 1/2\ {}_1q_{.5})} \quad (4)$$

For subsequent intervals, we require values of both $_{.5}q_x$ and $_{.5}q_{x+.5}$, neither of which are calculated directly in our life table. Expressing these two quantities in terms of our calculated values $_1q_{x+.5}$, for integer x, via equations (1) and (2) respectively, we obtain the following equation for $_1q_x$, $x = 1, 2, 3 \ldots$

$$_1q_x = 1/2\ {}_1q_{x-.5} + \frac{(1 - 1/2\ {}_1q_{x-.5})(1/2\ {}_1q_{x+.5})}{(1 - 1/2\ {}_1q_{x+.5})} \quad x \geqslant 1 \quad (5)$$

Table 1 shows q_x values obtained directly for the monthly intervals (0,.5), (.5,1.5), (1.5,2.5) . . . and q_x values interpolated according to equations (4) and (5) for the standard intervals. The values are the probabilities of conception by month since first marriage from the Sri Lanka Fertility Survey (WFS, 1975). Note that once the interpolated $_1q_x$ values have been obtained, the remaining life table values can be calculated in the usual manner.

RESULTS

Many demographers have developed techniques for dealing with exposure-related problems in life table construction, such as the calculation of exposures

Table 1.—Monthly Probabilities of Conception, as Derived Directly from Reported Data, and as Interpolated from Equations (4) and (5): Sri Lanka Fertility Survey.

Duration since First Marriage (months)	Monthly Probability of Conception
Reported	
0.0 - 0.5	.079
0.5 - 1.5	.140
1.5 - 2.5	.155
2.5 - 3.5	.178
3.5 - 4.5	.133
4.5 - 5.5	.149
5.5 - 6.5	.071
6.5 - 7.5	.095
7.5 - 8.5	.078
8.5 - 9.5	.093
9.5 - 10.5	.057
10.5 - 11.5	.080
11.5 - 12.5	.134
Interpolated	
0 - 1	.148
1 - 2	.142
2 - 3	.160
3 - 4	.148
4 - 5	.136
5 - 6	.106
6 - 7	.082
7 - 8	.084
8 - 9	.084
9 - 10	.073
10 - 11	.068
11 - 12	.106

associated with withdrawals, censoring and competing risks (e.g., Potter, 1969; Trussell and Menken, 1982). However, to our knowledge, only Smith (1980) and Trussell and Menken[4] (1980) have explicitly dealt with the problems arising from the use of monthly dates—i.e., the fact that the first duration interval is based on half the exposure of the remaining intervals.

Because there is little recognition of this problem, it is often unclear how researchers have calculated life tables based on data coded in monthly form. Many have ignored the problem entirely. Consequently their calculations follow one of two schemes: (a) events which occur in the same month as the start of exposure are classified in the interval (0,1); events which occur in the month subsequent to the start of exposure are classified in (1,2) etc.; (b) events which occur in the same month or the month subsequent to the start of exposure are classified in the interval (0,1); events which occur two months subsequent to the start of exposure are classified in (1,2), etc. Note that in both of these schemes, all intervals have lengths of one month, in contrast to our procedure. The two schemes described above can be thought of more easily in terms of annual probabilities. In fact, years are the most common length of intervals for demographic studies (although not for studies of fecundability). The first scheme is equivalent to attributing all events for which the duration (century month of event minus century month of start of exposure) is less than or equal to 11, to the first year. Such a procedure actually gives 11 and one-half months worth of events to the first year. The second scheme attributes all events for which the duration is less than or equal to 12 to the first year, and, hence, gives 12 and one-half month's worth of events to the first year.

In Table 2, we compare these two alternative schemes—Method 1 and Method 3, respectively—with the interpolated version—Method 2. The procedures are numbered in this manner since our calculation yields estimates roughly midway between those of Methods 1 and 3. Estimates of the probability of conception within the first, second, third, sixth and twelfth month since first union are shown in Table 2 as derived from all three procedures, for four countries—Colombia, Jordan, Panama, and Sri Lanka. All data come from the World Fertility Survey in the respective countries.[5] The estimates are based on pregnancies which occurred during the two-year peri-

Table 2.—Proportion of Women Having a Conception by Successive Durations since the Onset of First Marriage, and Estimated Fecundability, According to Three Methods of Calculation.

	Duration (in months) since First Marriage					Estimated Fecundability[a]
	1	2	3	6	12	
			Colombia			
Method 1	0.20	0.34	0.50	0.63	0.80	0.33
Method 2[b]	0.28	0.44	0.51	0.65	0.81	0.34
Method 3	0.34	0.50	0.53	0.68	0.81	0.35
			Jordan			
Method 1	0.10	0.22	0.31	0.51	0.70	0.16
Method 2[b]	0.16	0.27	0.36	0.54	0.71	0.19
Method 3	0.22	0.31	0.41	0.57	0.73	0.23
			Panama			
Method 1	0.19	0.39	0.48	0.63	0.70	0.30
Method 2[b]	0.30	0.44	0.52	0.64	0.73	0.35
Method 3	0.39	0.48	0.55	0.65	0.75	--[c]
			Sri Lanka			
Method 1	0.08	0.21	0.33	0.59	0.75	0.17
Method 2[b]	0.15	0.27	0.40	0.61	0.77	0.22
Method 3	0.21	0.33	0.45	0.62	0.79	0.27

NOTE: See text (p.7) for a description of the three methods of calculation.

[a]Derived by the procedure described in Bongaarts (1975).

[b]Method 2 calculations are interpolated.

[c]Estimate is beyond the range given by Bongaarts' procedure (i.e., > 0.35).

od prior to each survey and are restricted to women who were married at the time of the survey and had not experienced a marital disruption during the two-year period. In addition, the estimates are based entirely on noncontraceptive intervals, and hence, can be used to extract estimates of fecundability.

As expected, the three procedures produce very different values for the first month. If one were to estimate fecundability directly from the probability of conception in the first month of marriage (a procedure rarely applied because of reporting and sampling errors), recognition of the fact that exposure in the calendar month of marriage is half as long as exposure in subsequent months is crucial. Three months' duration is an important cut-off point because a well known procedure for estimating fecundability uses the proportion of women giving birth in months 9, 10, and 11 of marriage (roughly equivalent to becoming pregnant in the first three months of marriage) to estimate fecundability (Bongaarts, 1975). Based on this procedure,[6] we obtain the estimates shown in the last column of Table 2. On the assumption that the data are accurate, the difference between our estimate of fecundability and that produced by either Methods 1 or 3 is approximately 3 percent in Colombia, 14 percent in Panama, 16 percent in Jordan, and 23 percent in Sri Lanka. As noted at the beginning of this paper, the corrections associated with life tables tabulated from monthly data are not trivial when events are concentrated at early durations and when one's focus is on these short durations. If we were concerned only with the probability of conception within a year of marriage, the corrections would be insignificant—i.e., roughly 2 to 3 percent.

The reader may have noted an unexpected substantive result in Table 2. The estimates of fecundability for Jordan and for Sri Lanka are about 50 percent less than those for the two Latin American countries, Colombia and Panama. At least part of the difference appears to be due to the inclusion of some premarital pregnancies in the estimates for the latter countries, as a consequence of dating errors. The resulting estimates of about 0.34 are higher than those usually obtained for fecundability. The difference between estimates for Asian and for Latin American countries are discussed in Goldman et al. (1984).

CONCLUSION

Although the examples in Tables 1 and 2 refer only to the estimation of fecundability, the procedures introduced in this paper would be likely to affect estimates of any event concentrated at low durations or at young ages. For example, if information on infant deaths had been collected via a question on date (month and year) of death, life table estimates of infant mortality would be affected in a manner similar to estimates of fecundability: i.e., if infants died in the calendar month subsequent to month of birth, we would not know if the death occurred at a neonatal age (less than one month) or at an age between one and two months. Without resorting to exact dates (days), the problems theoretically could be avoided by asking respondents to supply directly information on age at death. (This strategy would be unlikely to work for estimates of fecundability since waiting times to conception are rarely known.) Unfortunately, several WFS survey questionnaires[7] elicited information on month of death rather than on age at death. In these cases, a correct estimate of neonatal or of infant mortality would have to be based on the interpolation equations described here. However, since mortality rates decline so steeply in the first few months of life, a proper solution would also have to incorporate exposures based on actual distributions of age at infant death.

In conclusion, the reader should bear in mind that in many surveys in developing countries, reporting errors would be of an order of magnitude that is larger than the correction factors introduced here. Moreover, our procedure is not really necessary when the focus of interest is on durations greater than six months. However, it is important for demographers to be aware of the technical problems associated with lack of information on exact dates and the fact that seemingly trivial corrections for what is essentially a "half-month problem" can have a considerable effect on life table estimates based on events which are concentrated in the first few duration intervals.

NOTES

[1] For the purpose of this analysis we have included only those conceptions which result in a live birth.

[2] Equations (1) and (2) are equivalent since one can be derived from the other by the law of total probability described in equation (3).

[3] David Smith (1983) has performed several calculations which suggest that when the underlying hazard function is exponential, a quadratic interpolation formula for l_x performs better than Balducci's hyperbolic formula.

[4] Trussell and Menken (1980) use a procedure which is based on the number of *calendar* months of exposure rather than on actual months of duration since the onset of exposure. For example, a woman married in January and conceiving the following March would contribute exposure to three intervals in the Trussell and Menken calculation but to only two intervals in our procedure. As a result, the two procedures are not comparable for early intervals, but they become indistinguishable by twelve months' duration.

[5] Colombia National Fertility Survey (1976); Jordan Fertility Survey (1976); Panama Fertility Survey (1976); Sri Lanka Fertility Survey (1975).

[6] If fecundability is heterogeneous across a population, estimates of fecundability are unbiased only if they are obtained from the proportion of women who conceive in the first month of marriage, since women with high probabilities of conception are successively removed from the pool of susceptible women with increasing duration. Bongaarts' procedure uses an underlying distribution of heterogeneity to convert reported proportions into estimates of fecundability from birth interval information.

[7] Among the WFS surveys, Korea and Portugal asked for only the date (month and year) of death, and Malaysia, Senegal, Guyana, Jamaica and Trinidad and Tobago initially asked for the date of death, but probed respondents who could not supply dates for information on either age at death or how many years ago the child had died (Singh, 1983).

ACKNOWLEDGMENTS

The authors are grateful to Griffith Feeney, Andrew Foster, Jane Menken, David Smith, James Trussell, Barbara Vaughan, and anonymous reviewers for their comments and suggestions and to Lois Paul for her programming assistance. Noreen Goldman was funded by NICHD Grant RO1 HD-17968 and the Andrew W. Mellon Foundation; Anne Pebley was funded by the Andrew W. Mellon Foundation.

REFERENCES

Ansell, C. 1874. On the Rate of Mortality. London: National Life Assurance Society.

Batten, R. W. 1978. Mortality Table Construction. New Jersey: Prentice Hall.

Bongaarts, J. 1975. A Method for Estimation of Fecundability. Demography 4:645–660.

Goldman, N., C. F. Westoff, and L. E. Paul (1984). Measurement of Fecundability from Survey Data: A Comparative Analysis. Paper presented at the Annual Meeting of the Population Association of America, Minneapolis.

Potter, R. G. 1969. Use-effectiveness of Intrauterine Contraception as a Problem in Competing Risks. In R. Freedman and J. Y. Takeshita (eds.). Family Planning in Taiwan. Princeton, N.J.: Princeton University Press.

Singh, S. 1983. Comparability of Questionnaires. World Fertility Survey Cross National Summaries. WFS/TECH 2127.

Smith, D. P. 1983. Personal communication.

———1980. Life Table Analysis. World Fertility Survey Technical Bulletin No. 6.

Spiegelman, M. 1973. Introduction to Demography. Cambridge: Harvard University Press.

Trussell, J. and J. Menken. 1980. The Calculation of Gross Rates of Continuation for Contraceptive Methods: Single and Multiple Increment Life Tables. Appendix in Contraceptive Efficacy Among Married Women Aged 15–44 Years, Vital and Health Statistics Series 23, Number 5, National Center for Health Statistics.

Trussell, J. and J. Menken. 1982. Life Table Analysis of Contraceptive Failure. In Albert I. Hermalin and Barbara Entwisle (eds.). The Role of Surveys in the Analysis of Family Planning Programs. Liege: Ordina Editions.