FERTILITY AND MORTALITY ESTIMATION FROM SINGLE-ROUND SURVEY DATA

Hilary J. Page, Princeton University

A majority of the world's population lives in areas in which universal birth and death registration systems are either absent or else so incomplete as to yield records that are practically useless for fertility and mortality estimation. Modernizing countries will doubtless need in the long run to develop efficient civil registration systems; but for most developing countries the attainment of this ideal is many years, if not decades, away. In the meantime, attempts are being made to develop specifically demographic data-collection and analysis procedures that will produce more reliable estimates of vital rates than are permitted by existing civil registration systems. Strategies to improve direct recording of vital events include intensive registration schemes in sample areas; multi-round surveys, or repeated recording of the individual elements in sample populations; or some combination of these as in dual-records systems of the PGE type, in which events recorded by one observation system are matched against those recorded by an independent system and an estimate of the total events (including those missed by both systems) is obtained from the results of the matching process. This type of strategy is the focus of Seltzer's paper presented later in this session, evaluating PGE studies [15].

An alternative tactic in the absence of adequate registration records is to avail oneself of the indirect evidence on fertility and mortality that can be gleaned from single-round censuses or surveys. The principle forms of evidence are the reported age-distributions and retrospective reports of reproductive and bereavement experiences. Such data are typically fragmentary and defective, besides being indirect, and they certainly cannot be taken at face value. However, in the last two decades a large body of estimation techniques has been developed to extract the most reliable and informative components of these data and to derive estimates of vital rates from them, by exploiting the relationships that must hold between various demographic variables.

This paper reviews several of the indirect estimation techniques that are most widely applicable in the developing world at the present time - that is, techniques appropriate when the only data available are those from one or very few single-round censuses or surveys and are more or less defective in quality. The papers by Cho [3] and Zachariah [21] that follow discuss special procedures that are applicable under less inauspicious circumstances: situations in which the data either exhibit exceptionally high quality with respect to age-reporting (in this case, Korea) or form a series extending over a comparatively long time-period (several Arab countries).

It is obviously impossible to present a fully comprehensive review of the state of the art in a short paper. The literature on indirect estimation of vital rates is substantial; for a detailed description of the basic principles and procedures the reader can be referred to a U.N. manual devoted to the subject [18]. The present paper is restricted to little more than a sketch of some of the possibilities that are based on analyses of aggregate age-distribution data, and concentrates on use of the more detailed information gathered from retrospective fertility and mortality reports. My practical assessment of these techniques derives largely from experience with their application to materials from Africa south of the Sahara: major analyses of these data have been carried out at the Institut National des Statistiques et des Etudes Economiques in Paris [11] and at Princeton [2]. Many of the problems that arise are not peculiar to these particular data sets but are commonly found, with greater or less intensity, in other developing areas.

Estimation from age-distributions

Given age-distributions from two censuses or surveys (and the assumption that the population is closed to migration) one can readily estimate apparent intercensal survival rates for each age-group and hence build up a life table for the intercensal period. provided there are no inequalities of coverage between the two censuses to bias the estimates. systematic patterns of age-misreporting often give rise to wild fluctuations between the estimates for adjacent age-groups (or even reported survival rates exceeding unity) making the results unusable as they stand. Mortality estimation under these conditions is greatly facilitated by our knowledge of the strong tendency for death rates at different ages in any population to be intercorrelated. Thus if we can determine the mortality level for one age-group it is possible to estimate upper and lower limits for mortality rates at other ages. Exact determination of mortality rates at other ages is not feasible because the inter-correlations are not perfect; in fact, several distinct age-structures of mortality have been detected, and from these various sets of model life tables have been prepared incorporating each age-structure at different levels of mortality. Three sets of model tables have been developed to date. The first set was prepared by the United Nations in 1955 [17], and has been widely used; this set has now been largely replaced by subsequent sets because it does not embody variations in the age-structure of mortality at the same overall level as do the later tables. The later models consist of the four 'families' (age-structures of mortality) prepared by Coale and Demeny at Princeton [7], and the various combinations of mortality patterns incorporated in the tables prepared by Lederman at the Institut National des Etudes Démographiques in

Paris. The most striking difference between the various observed age-structures of mortality lies in the relationship between infant and early childhood mortality (under age 5) on the one hand, and adult mortality on the other. These differences imply that it is useful to have separate estimates of adult and of early childhood mortality.

Adult mortality rates can be estimated fairly reliably by comparison of the reported intercensal survival rates and those embodied in model life tables. In practice it is necessary to use cumulated sections of the age-distribution rather than individual cohorts, in order to minimise the effects of age-misreporting. The technique of forward projection can be used to project successive cumulated segments of the age-distribution recorded in the first census (for example, the population aged over 10 years, over 15 years, etc.) according to the survival rates of the most likely model life tables; the resulting projected populations can be compared with the population actually recorded in the later census. The model life table producing the best fit can then be taken to represent mortality rates over all age-groups - except the very youngest age-group, for which no survival ratio is reported. If we are prepared to assume that childhood and adult mortality bear the same relation to each other in the population as they do in the model life table selected, then the overall average intercensal death rate can be estimated by applying the model's age-specific mortality rates to the mid-period population. Since the intercensal growth rate is also known, the overall birth rate can be obtained by subtraction. However, the assumption about childhood mortality is questionable: it would be preferable to make separate estimates of this using other forms of data and other estimation techniques.

Recent fertility levels can also be estimated from one age-distribution by reverse projection of the population under age 5, assuming that this number is accurately reported (which is unlikely) and that childhood mortality conditions are known from some other source. Alternatively, since the age-distribution of a closed population depends only on its recent history of fertility and mortality, if a stable situation can be assumed then one has the range of stable population theory at one's disposal for estimation [18,19]. If either the rate of population increase is known or an index of mortality is available, then fertility levels can be inferred from an age-distribution. Tabulation of stable age-distributions by Coale and Demeny [7] has greatly simplified this sort of estimation; also tabulation of approximate adjustment factors for quasi-stable situations in which mortality has been declining [18]. However, the assumption of stability (or some conjecture as to the nature and extent of departure from stability) limits successful application of the method. Furthermore, choice of a particular age-structure of mortality may introduce errors. Fertility estimates from model stable populations are not severely affected by this. Mortality estimates are very sensitive to it, and consequently cannot be accorded much credibility unless there is independent evidence as to the appropriate age-structure to use. Finally, there is the problem of age-misreporting: birth rate estimates, for example, vary widely depending upon which index of the age-distribution is taken as the most reliable; detailed analysis of the age-distribution data is really required in each individual case to select this index.

Retrospective fertility data

Given highly accurate information from women regarding the birth dates of both themselves and each of their children, one can determine the age-specific fertility rates that have been experienced by each cohort. If we then assumes that the fertility experience of those who have died was not exceptional, then we can reconstruct in detail the population's fertility experience over roughly half a century. For most developing countries we must rule out the possibility of obtaining large numbers of individual reproductive histories of sufficiently high quality. In societies where respondents usually know neither their own nor their children's chronological ages, the intensive interviewing and highly trained enumerators required to produce full and accurate records are beyond the scope of most large-scale demographic enquiries. Bogue [1] has developed techniques of recording and analysing full pregnancy histories, but these would generally appear to be limited in application to small and very intensive studies or to the younger and better educated segments of the population.

The data that can be, and have been, more widely collected are much less ambitious. They consist of mean parity (average number of children ever born per woman up to the date of the survey) and current fertility (average number of children born per owoman during the year preceding the survey) - both sets of data being tabulated typically for 5-year age-groups. Both types of data are subject to severe weaknesses and cannot be used directly.

Current fertility data would permit straightforward calculation of the prevailing age-specific fertility schedule if they were accurate; but accuracy is unlikely on several grounds. Patterns of age-misreporting systematically related to physical maturity and preconceptions about age at puberty, to marital status, or to apparent fertility are all likely, since age is frequently estimated on the basis of these more readily observable characteristics. As one official report expressed it (Central African Republic), a married woman's age is too often estimated by rule of thumb as being '14 years plus twice the number of children she has borne' [6, p. 13]. Such biases may account for the improbably high mean age of the fertility schedule reported in the 1960 Ghana Census ([10, pp. 7-9] and also noted in the 1962 Kenya Census [2, p. 173], for example. Even estimates of the overall population birth rate from the current fertility data are highly dubious because of misreporting of the occurence or non-occurence of a birth in the preceding twelve months. There seems to be a tendency in many populations for either over- or under-reporting of these events - the reason, it is usually suggested, being a misperception of the specified reference period. In addition, there is the problem of straightforward omission of some births: infants that died very shortly after birth are probably particularly likely to be omitted, especially if the birth occurred a long time before the survey. Analysis of the births reported for a one year reference period in Upper Volta provide an illustration of typical data defects. When tabulated by month of reported occurence, the monthly total fell off quite regularly and rapidly with increasing remoteness from the survey date [19, p.246]. This feature might indicate either a propensity to misperceive time periods, or a tendency to omit births with extension of the recall period. Whatever the cause, the observation enhances one's levels of skepticism with respect to the raw data.

Accurate age-specific parity data could be used directly to provide a summary measure of the average fertility to date of each cohort - a measure of completed fertility for those cohorts that have passed the childbearing age. Strictly, it would be a measure for the surviving portion of the cohort, but the effects of mortality differentials between relatively fertile and relatively infertile women are presumed to be negligible. The age-reporting problem is at least as severe here as with current fertility data, if not more so. The parity data are likewise subject to omission; this usually appears to be of considerable magnitude in older women's reports, which consequently have little credibility.

The simple fact that the average parity of any cohort of women corresponds to the cumulation of the age-specific fertility rates they have experienced can be exploited in a number of ways; though in practice, difficulties often arise from the curvilinearity of typical age-specific fertility schedules combined with the fact that data are frequently available only for 5-year age-groups, which necessitates some form of curve fitting. Straightforward differencing of the parity data to obtain estimates of age-specific fertility rates is one possibility if data are available by single years of age. This requires, though, that reporting of both age and parity be very accurate; moreover, unless fertility has been extremely stable in the past, the results confound the effect of the rise in parity for each cohort as it passes through its reproductive span with the effect of differences in fertility experience between successive cohorts. The conditions required for this easy procedure to work are rarely encountered.

Where moderately complete registration records exist and where age is reported quite accurately, the more reliable portions of the parity data (those relating to younger women, who are less likely to omit a significant proportion of births) can be used to provide some check and ajustment of the registration data. Comparison of the younger women's reported parity with the average parity levels implied by the cumulated age-specific fertility rates recorded for the same cohorts in the registers provides an adjustment factor for these registration data. If we can assume that the level of registration completeness is uniform at all ages and for all cohorts, then this adjustment factor can be applied to all the registered births, permitting estimates of cohort fertility rates and hence of both current levels and of past trends (if the registration series covers any length of time). Registration data are rarely of sufficiently high quality to be amenable to this form of adjustment, however.

With the assumption that fertility rates have been constant in the recent past, the cumulation of current fertility data from a cross-sectional survey corresponds to the past experience of the various cohorts of women, summarised in their average parity. Under these conditions we can estimate prevailing age-specific fertility rates from the current fertility data in a manner analogous to that in which both prevailing and past rates can be obtained from adjusted registration data. In this case the current fertility data are assumed correct in age-structure though not necessarily in level; that is, it is assumed that any tendency to over- or under-report current births is the same for all age-groups. Brass has developed an ingenious procedure on the basis of this, by which the current fertility rates by 5-year age-groups can be converted into the average parity levels they imply for the same 5-year age-groups, taking into account the curvilinearity of fertility schedules. [18, chapters 2 and 7; 2, chapter 3]. The procedure utilises a model age-structure of fertility for populations not practising fertility control; it is allowed to vary only in its starting age. The location of the model that is appropriate for any particular population is determined from the ratio of the current fertility rates for the two youngest reproductive age-groups. A model that was allowed to vary in form as well as starting point might better describe the variety of fertility schedules. Coale [8] has recently shown that the age-pattern of entry into marriage follows a standard pattern that varies in both starting point and form; since the rising portion of the fertility schedule in populations not practising fertility control is determined largely by the age-pattern of entry into marriage, both parameters should probably be included in model fertility schedules. However, in practice this point is very minor: the improvement any such refinement might introduce is probably paltry in relation to the errors that arise from defects in the data.

The Brass method is very appealing and has been used extensively. But it is sensitive to fertility trends and to data defects - especially it is very sensitive to massive and systematic age-misstatements like those common to data from tropical Africa [2, chapter 3, appendix B], and its practical value is therefore limited.

Despite the availability of retrospective fertility data in many areas, we are often thrown back to inferring fertility from age-distribution data.

Retrospective mortality data

Current mortality data from single-round surveys - that is, deaths by age during the preceding twelve months - are of notoriously poor quality. Errors and omissions are rife and the data are rarely usuable.

Mortality at early ages can, however, be estimated from data on the number of children ever born and the number of these surviving, tabulated by age of mother. The basic method, again developed by Brass [18, chapters 2 and 7 ; 2, chapter 3], involves conversion of the proportions surviving into conventional life table functions. The essential idea is that, if the time distribution of births that have occurred to a group of women is known, then the proportion surviving for these women at any point in time can be viewed as a function of the mortality schedule to which the children have been subject. Roughly speaking, the proportion dead for women 20-24 years of age corresponds to the probability of dying before age 2; that for women aged 25-29 years to the probability of dying before age 3, and so forth. The correspondence is only approximate because of the effects of different age-structures of mortality and of different time distributions of births. The time distribution of births that have occurred to a cohort of women is given quite simply by the age-specific fertility schedule they have experienced; if we knew this then we could determine the age-distribution the children would have in the absence of mortality. If we also knew the probability of dying to which the children have been subject we could calculate the proportion of them that would have died before the survey; conversely, given the proportion dead, we can calculate the probability of dying before a certain age, for any particular age-structure of mortality. The method involves the input of two models, therefore; one for the assumed age-structure of fertility and one for that of early mortality.

Brass used the same model age-structure of fertility to determine the time-distribution of births as is used in his fertility estimation method; here the ratio of parity levels reported for the two youngest reproductive age-groups is used as an index of the appropriate starting age. Unfortunately, this index is rather sensitive to the forms of misreporting that are probably common. Sullivan [16] has developed a modification of the Brass basic procedure using an alternative way of determining the time distribution of births. This is based on empirical single-year age-specific schedules rather than on a standard schedule, and hence to some extent circumvents the problem that the standard does not allow for variations in form as well as starting age. Moreover, the modification uses the ratio of parity levels for the second and third reproductive age-groups, which is probably sless sensitive to data errors than the ratio for the first and second groups. The modification is based on regression analysis (for various levels and age-patterns of mortality) of the relationships between the proportion dead of children ever born for women in each age-group and this index of the fertility schedule. Sullivan found very close relationship between this index on the one hand and the conversion factor needed to transform the proportion dead into the corresponding life table function. He found this over the entire range of mortality levels within each of the Coale-Demeny families of model life tables, but slightly different relationships between these families. Regression coefficients for the North and West families are given in: -

North

$$ {}_2q_0/D_2 = 1.30 - .63(P_2/P_3) $$
$$ {}_3q_0/D_3 = 1.17 - .50(P_2/P_3) $$
$$ {}_5q_0/D_4 = 1.15 - .42(P_2/P_3) $$

West

$$ {}_2q_0/D_2 = 1.30 - .54(P_2/P_3) $$
$$ {}_3q_0/D_3 = 1.17 - .40(P_2/P_3) $$
$$ {}_5q_0/D_4 = 1.13 - .33(P_2/P_3) $$

where ${}_xq_0$ is the proportion dying before exact age x in the life table;
P_i is reported parity for women in the i'th 5 year age-group (i=1 for age-group 15-19);
D_i is the proportion of children dead as reported by women in the i'th age-group

Comparison of results of the two procedures for African data [13] shows that the estimated probabilities of dying before ages 2, 3, and 5 for any given population are similar, but that they are more consistent with the model life tables on which their derivation is based when calculated from Sullivan's modification than from the original procedure. Consistency is not necessarily a virtue: if mortality had been declining then the estimated probabilities of dying before older ages should be too high relative to the estimated probabilities of dying before the youngest ages, since the former refer to mortality conditions of a less recent period. Such a tendency is found very faintly in the Sullivan estimates for those populations where the estimated probabilities of dying before age 2 are particularly low: one would scarcely expect to find such a tendency where early mortality is still extremely high.

Several problems remain with this procedure. Firstly there are the problems of misreporting; both omissions and age-misstatement. Secondly, the assumption that the cross-sectional parity data truly reflect past experience of the younger cohorts of women - that is, that fertility has not changed during the last ten to fifteen years - may be unwarranted. Thirdly, the estimates refer to past mortality conditions rather than prevailing ones, and so are incapable of providing up-to-date information. The estimated probability of dying before age 2, for example, is based on average conditions over the preceding 4-5 years. Finally, the choice of a particular model age-structure of mortality may introduce uncertainty, for the choice must be made on the basis of limited clues and alternative models yield rather different results. Determination of an appropriate age-structure is a problem throughout most of the developing world; it seems to be of particular severity in Africa where there are indications that early child mortality between ages one and five may be exceptionally high relative to infant mortality. One can question whether any of the Coale-Demeny or U.N. models represent early mortality patterns adequately. Recent multi-round surveys by Cantrelle and his associates in Senegal [4, 5] for example, suggest that the pattern may be far more extreme than any of the Coale-Demeny models. Examination of the estimates obtained from the Sullivan procedure using the two most likely Coale-Demeny families (North and West) shows that the estimates of proportions surviving to age five are almost the same whichever family is used, but that the implied levels of infant mortality are divergent [14]. Clearly, whilst retrospective data of this kind can indicate roughly the levels of early mortality, we can have very little confidence in our estimates until we know more about the age-structure of early mortality. For this, multi-round surveys and intensive registration schemes are probably the only answer, though national resources in most countries are not adequate to develop these on a large-scale.

With respect to adult mortality patterns and levels we know even less. The possibility exists of estimating adult mortality from single-round surveys by exploiting data on the proportion of people in each age-group whose mothers and fathers have died. Again we owe the development of a technique to Brass, but in this case the procedure is considerably more complex (especially for males) and has not been fully worked out and tested. The essential idea is analogous to that used in estimation from proportions of children surviving, and again it involves the input of model age-structures of fertility and mortality; the former is rarely available for males. If we can assume that the structure of fertility is known for some time in the past - say five years ago - then we can calculate the approximate age-distribution of the women who gave birth during that year. If we knew also the age-structure and level of adult female mortality, we could then calculate what proportion of those women would have died during the five years before the survey. Conversely, given the proportion of children whose mothers have died, and making some assumption as to the age-structure of mortality, we could estimate the level of that mortality. The procedure suffers similar problems as that for estimating childhood mortality: in addition, the assumption that survival of the children is independent of survival of the mother (analogous to the assumption that survival of the mother is independent of her fertility in the childhood mortality case) may introduce a bias. To date this procedure has not been widely applied, even for females: in Africa, data are available for some five or so countries, and are currently being worked on by Brass.

Conclusion

Indirect estimation techniques have been applied with varying degress of success, their usefulness being restricted most frequently by one or more of three sets of limitations.

Firstly, the sensitivity of most procedures to age-misstatement. This problem bedevils any detailed demographic analysis in much of the developing world, and will continue to plague demographers for a long time. It even prevents reliable indirect estimation of such simple summary measures as the overall population birth and death rates inasmuch as estimation of these must depend at present on techniques sensitive to age mis-statement: overal current fertility and mortality reports cannot be used because of over- or under-reporting.

Secondly, the use of population models to compensate for incompleteness of data. Model age-structures of fertility and mortality, and model stable age-distributions permit estimation in cases where otherwise it would be impossible; but they must all be selected on the basis of fragmentary information, and the appropriateness of any particular model is often open to question. Only intensive attempts to gather reliable direct data on fertility and mortality age-structures can ease this problem.

Thirdly, the assumption in some procedures that the population has experienced constant fertility and mortality in the recent past. A broad theoretical insight has been gained on the relations between age-distribution and trends in vital rates, notably by Coale [9], but at present we are able to make only limited practical application of this: we can estimate current vital rates using age-distribution data provided that the tempo and magnitude of the trends are known and have been comparatively regular. Procedures for detecting and measuring trends are clearly going to become of increasing importance in the future as mortality conditions improve and as fertility levels may start to decline.

It is clear that any estimates obtained from single-round surveys should be viewed with some degree of skepticism; also that we should use and compare all possible pieces of information and methods of estimation, rather than relying on a single item or technique. For many countries, however, they are the only sort of estimates presently available. For those countries where direct recording of vital events is more nearly adequate or where intensive efforts are being made to obtain accurate direct observation on a sample of the population, single-round survey estimates will provide a useful check on, and complement to, the direct data for some time to come.

REFERENCES

[1] Bogue, D.J. and Bogue, E., A Model Interview for Fertility Research and Family Planning Evaluation, Family Planning Manual No. 3; and Techniques of Pregnancy History Analysis, Family Planning Manual No. 4, Community and Family Study Center, University of Chicago, 1970.

[2] Brass, W. et al., The Demography of Tropical Africa, Princeton University Press, 1968.

[3] Cho, L.J., "On Estimating Annual Birth Rates from Census Data on Children," paper presented at the Annual Meeting of the American Statistical Association, Fort Collins, August, 1971.

[4] Cantrelle, P. Etude démographique dans la région du Siné-Saloum (Sénégal), 1963-1965. O.R.S.T.O.M. Travaux et Documents, No. 1, Paris, 1969.

[5] Cantrelle, P. et al., Mortalité de l'enfant dans la région de Khombol-Thiénaba (Sénégal), 1964-1968, O.R.S.T.O,M. Cahiers, Série Sciences Humaines, Vol. VI, Paris, 1969.

[6] Central African Republic, Mission Socio-Economique Centre-Oubangui; France, Secrétariat d'Etat aux Relations avec les Etats de la Communauté, Enquête démographique Centre-Oubangui, 1959. Méthodologie, résultats provisoires, Paris, 1969

[7] Coale, A.J. and Demeny, P., Regional Model Life Tables and Stable Populations, Princeton University Press, 1966.

[8] Coale, A.J., "Age Patterns of Marriage," Population Studies, Vol. XXV (July, 1971), pp. 193-214.

[9] Coale, A.J., Fertility, Mortality and Age Composition, Princeton University Press, forthcoming.

[10] Gaisie, S.K., Dynamics of Population Growth in Ghana, University of Ghana Demographic Unit, Ghana Population Studies No. 1, Legon, undated.

[11] I:N.S.E.E., Afrique noire, Madagascar, Comores: démographie comparée, Paris, 1967.

[12] Lederman, S., Nouvelles tables-types de mortalité, I.N.E.D. Travaux et Documents No. 53, Presses Universitaires de France, Paris, 1969.

[13] Page, H.J. and Coale, A.J., "Estimates of Fertility and Child Mortality in Africa south of the Sahara," in S.H.Ominde (ed.), Studies in the Population of Africa, 1971.

[14] Page, H.J., "Infant and Child Mortality," paper prepared for the African Regional Population Conference, Accra, December,1971.

[15] Seltzer, W., "PGE Studies: Costs and Effectiveness," paper presented at the Annual Meeting of the American Statistical Association, Fort Collins, August, 1971.

[16] Sullivan, J.M., Estimates of Childhood Mortality Conditions from Childhood Survival Statistics, unpublished Ph.D. dissertation, Princeton University, 1970.

[17] United Nations, Age and Sex Patterns of Mortality, ST/SOA/Series A/22, New York,

[18] United Nations, Methods of Estimating Basic Demographic Measures from Incomplete Data, prepared by A.J.Coale and P.Demeny, ST/SOA/ Series A/42, New York,1967.

[19] United Nations, The Concept of a Stable Population: Application to the Study of Countries with Incomplete Demographic Statistics, prepared by J. Bourgeois- Pichat, ST/SOA/Series A/39, New York, 1968.

[20] Upper Volta, Service de Statistique et de la Mécanographie; France, Service de Coopération, I.N.S.E.E., Enquête démographique par Sondage en République de Haut-Volta, 1960-1961, Tome II, Paris,1970.

[21] Zachariah, K.C., "Application of Stable Population Techniques for Estimating Demographic Measures for Arab Countries," paper presented at the Annual Meeting of the American Statistical Association, Fort Collins, August, 1971.