50140

UNITED NATIONS
ECONOMIC AND SOCIAL COUNCIL

Distr.
LIMITED
E/CN.14/CAS.4/VS/7
18 November 1964
Original: ENGLISH

ECONOMIC COMMISSION FOR AFRICA
African Seminar on
Vital Statistics
Addis Ababa, 14-19 December 1964.
Item 10 on the provisional agenda

USES OF CENSUS OR SURVEY DATA FOR THE ESTIMATION OF VITAL RATES

By W. Brass
Department of Statistics, University of Aberdeen

Table of Contents

		Paragraphs
I.	Introduction	1 - 2
II.	Estimation of Fertility	3 - 18
III.	Estimation of Mortality	19 - 29
IV.	Use of a Model Life Table System	30 - 37
V.	Checks from Quasi-stable Population Theory	38 - 43
	Note	44
	References	

64-4505

I. Introduction

1. Recently, extensive reviews of methods for extracting basic demographic measures from limited or defective records have been published or prepared (International Union for the Scientific Study of Population (1963), Brass (1964)). Instead of covering the same ground with the sketchiness made necessary by the limitations of space, this paper will describe an integrated and systematic procedure for analysing census or survey recrods of a particular type. The techniques were developed at the Office of Population Research, Princeton University, consequent on studies of the nature of errors and deficiencies in African data. Their purpose is the estimation of basic rates in which biases are eliminated or reduced by internal cross checks of the records and deficiencies overcome by the use of demographic models. It will only be possible to give an account of how the methods are applied with brief indications of their characteristics and justification.

2. The censuses and surveys with which we are concerned are those in which data of two kinds have been collected. The two types of data are vital events, i.e. births to each mother and deaths, in a current period and the number of children ever born and died to each woman. It is assumed that ages of the living and of those dying in the current period are also reported, either in single years or intervals of moderate length. The methods can be modified for less complete records but this will not be discussed.

II. Estimation of Fertility

3. In this section we will discuss methods of analysing the two types of data to obtain plausible estimates of the level of fertility even when both sets are subject to error. The information on current births is obtained by questioning the women about the number of children born to them in a short period preceding the Census. For definiteness it

will be assumed that the period is the most usual one of a year. Age specific fertility rates can be derived by division of the number of births to mothers by the corresponding total women in the population for each age group. Measures based on such specific rates will be referred to as "current" and the indices of fertility obtained from the mean number of children ever born to women of each age will be called "retrospective".

4. The two types of measures can be used to detect and allow for errors in the data because of the logical relationship between them. As a cohort of women moves through life the mean number of children ever born at each exact age equals the cumulative total of age specific fertility rates to that age, if it can be assumed that the women dying have the same fertility as those surviving. If the fertility of the population is constant the age specific rates will be the same as the "current" ones and the relationship will hold for all ages of women.

5. The mean children ever born per woman by each age can be calculated from the current data and compared with the corresponding retrospective ratios. If the two sets of indices agree the evidence for accuracy is strong. Application of the procedure to records for many African populations has revealed systematic discrepancies. A rate based on responses to a question about births occurring during the year preceding a census or survey has sometimes been so clearly wrong as to be unusable, mostly because the level is far too low but sometimes because it is grossly inflated. On the other hand the reported mean numbers of children ever born often increase too gradually with age of woman, especially above 30 or 35 years, and sometimes even decline at ages beyond the reproductive period, due to the omission of an increasing proportion of the children born from the reports as women become older.

6. Because the deficiencies are different a technique can be developed for estimating total fertility from these parts of the two sets of measurements which are likely to be most reliable. The technique depends on the following propositions:

1. The most important source of error in the recalled number of births in the year preceding the census is imprecision in the <u>reference period</u>. The respondents may report events which occurred, on average, in the past 8 months or (with a different culture or framing of questions) in the past 15 months. The same average reference period, however, can be expected to hold for each age of mother, particularly in surveys of illiterate populations, where the interviewer tries to check the consistency of the responses by the development of the children.

2. The number of children ever born is reported with good accuracy by younger women . The events which they are asked to recall have happened recently; the total births to each are typically not more than two or three so that the difficulties of counting large numbers in a non-numerate society do not arise.

7. On the basis of these propositions a technique for estimating fertility has been developed. The <u>age pattern</u> of specific rates obtained from the reports of current births is accepted but the <u>level</u> of fertility is estimated from the mean children ever born, reported by the younger women. In the application of the procedure the mean children ever born implied by the current rates is compared at each age with the observed retrospective value. The ratios of the retrospective to the cumulated measures for the young women give a factor which is applied to the current specific rates at all ages to adjust the fertility to the required level.

8. If consistent data on births by individual years of age were available the calculations for making the comparisons would be straight forward. Usually, however, the tabulations are by five year groups of women, typically 15-19, 20-24, etc., to reduce the effects of age errors due to digit preference and chance fluctuations of small numbers. It must also be noted that the specific fertility experience is for ages about six months less than the tabulated ones. This displacement arises because the births reported occurred over the preceding year; on average the mothers were half a year younger when giving birth than at the time of the census.

9. The specific fertility rate for an interval is an average value per year. Multiplication by the number of years in the interval gives the mean births which would occur to women passing through the interval. These mean values for the seven five year groups which effectively cover the reproductive period will be denoted by f_1, f_2 and so on. The cumulation of the current f values from the lowest age group gives the mean children ever born at the boundaries of the intervals. The total of $f_1 + f_2$ up to f_{i-1}, i.e. to the lower boundary of f_i will be symbolised by F_i. The retrospective ratios of mean children ever born per woman for the same five year age groups will be denoted by r_1, r_2 and so on. These are not measures at exact ages but averages over the age group. It is therefore, necessary to devise some procedure for calculating from the f values, the corresponding ratios of mean children ever born for the age intervals. The calculated ratios will be referred to as the A set and the observed retrospective means as the B set.

10. An estimation procedure which can be applied very easily has been developed. The basis for the procedure is a model distribution of age specific fertilities in which the shape is fixed but the location may vary, i.e. the curve describing the shape can move along the age axis so that the lower and upper ages of child-bearing and the mean of the distribution are displaced by the same amount. For evenly spaced locations of the model the following values were calculated:

1 - the mean of the age specific fertility distribution;

2 - the ratio of f_1 to f_2, i.e. of the mean births in the first age interval to those in the second; 3 - for each age group, the factor k by which the f_i value had to be multiplied in order that the formula $F_i + k f_i$ should give r_i, the corresponding mean children ever born, exactly when there was a half year age displacement in the reported current fertility rates.

Table 1. Multiplying factors for the derivation of mean children ever born per woman from f ratios of births in the year preceding the census.*

Age group of women in years								
15-19	.224	.262	.323	.390	.461	.528	.585	.634
20-24	.511	.538	.556	.568	.578	.585	.592	.597
25-29	.585	.592	.597	.602	.607	.611	.615	.619
30-34	.611	.615	.619	.624	.628	.633	.638	.643
35-39	.633	.638	.643	.649	.657	.665	.675	.687
40-44	.665	.675	.687	.702	.722	.748	.783	.830
45-49	.728	.779	.830	.879	.926	.968	.997	1.000
f_1/f_2	.036	.113	.213	.330	.460	.605	.764	.939
m (years)	31.7	30.7	29.7	28.7	27.7	26.7	25.7	24.7

* It is assumed that (in accordance with common practice) births to women before fifteen years were re-allocated to the age group 15-19.

11. The mathematical form of the fertility model is described in the appendix note and the k multiplying factors are shown in table 1. The routine for using the factors is the following. Multiply the specific rates obtained from the observed data on current births by the age interval to derive the f values. From these calculate the mean age m of the distribution, the ratio f_1/f_2 and the cumulated sums of the f's, F_i up to the lower boundary of the i'th age group. Choose the column of k factors which corresponds to the observed f_1/f_2, interpolating linearly between the columns of the table if necessary. The factors from this column for the <u>first three</u> age groups are used to estimate r_i from the expression $F_i + k\ f_i$. Similarly, the mean m determines a column of factors which is used to derive the r values for the remaining four age groups.

12. In effect, by this procedure, the observed specific fertility distribution is fitted firstly by the model with the same f_1/f_2 ratio and secondly by the one with the same mean. The k factors of the models are taken as the estimates for the observed data. Fitting by the equation of f_1/f_2 ratios ensures that the observed and model distributions are in good agreement at the younger ages and equality of means leads to a similar outcome over the middle range of reproduction.

13. Since in applications we are seeking for a proportional correction to the current rates to bring them into agreement with the retrospective fertility level of the younger women it is most useful to make comparisons in terms of the ratios of the observed (B) to the estimated (A) set of values. If the assumptions held and the model was exact all the B/A ratios would equal one. Calculations for a number of sets of records have shown that the B/A ratio for the first age group is sensitive to the exact shape of the specific fertility distribution in the first few years of the reproductive period but that errors due to deviations from the model at later ages are very slight. The fertility level must be determined by the mean children ever born to the youngest group of women for

which the results have acceptable accuracy, to reduce the effects of memory failures and time trends. As a rule, therefore, the B/A ratio for the age group 20-24 years should be used to adjust the current rates for fertility level unless it is inconsistent with the pattern of values at later ages.

14. To illustrate the characteristics of the method in applications to census records which are deficient, the calculations for a hypothetical set of observations will be outlined. The example was constructed from the specific fertility rates at single years for the Ukraine, 1926-27, by making the following changes: (a) the specific fertilities were adjusted to what the observed rates would have been if recording was of births in the year preceding a census and one quarter were not reported at all ages of mothers, (b) a proportion of the children, varying linearly from zero in the age group 15-19 years to 24% for the group 45-49 years, was omitted from the retrospective values of mean children per woman, which were constructed to agree with the original specific rates.

15. Table 2 shows how the computations are made.

Table 2. Example to illustrate the B/A ratio method for estimating the level of fertility

Age group of women in years	f*	F	k	F+kf =A	B	B/A	Adjusted f*
15-19	0.105	-	.269	0.028	0.038	1.36	0.140
20-24	0.850	0.105	.539	0.563	0.747	1.33	1.130
25-29	0.975	0.955	.593	1.533	1.892	1.23	1.297
30-34	0.858	1.930	.617	2.459	2.884	1.17	1.141
35-39	0.618	2.788	.640	3.184	3.560	1.12	0.822
40-44	0.335	3.406	.681	3.634	3.868	1.06	0.446
45-49	0.110	3.741	.804	3.829	3.868	1.01	0.146
Total fertility	3.851	3.851	-	3.851	3.868	1.00	5.122

$f_1/f_2 = 0.124$; $m = 30.22$

* For age intervals half a year younger than shown

The mean children ever born per woman, recorded for this population, are in the column headed B. The F values are obtained by the summation of the current f's to the lower boundary of the age group, f_1/f_2 and m are also calculated from the f distribution and are shown at the foot of the table. Reference is then made to table 1. The observed f_1/f_2 ratio of 0.124 lies between the second and third column of factors, and is a proportion .011/.100 = 0.11 of the column interval from the first of these, linear interpolation between the k factors, using this proportion, gives the values for the first three age groups in table 2. Similarly the mean 30.22 is 0.46 of the interval along from the second to the third column and the k values for the fourth to the seventh age groups are obtained by interpolation. When the estimated k factors are entered in the expression F + kf, the A set of mean children ever born per woman, corresponding to the current rates, is derived.

16. The B/A ratios fall consistently with age with a value of 1.33 for the age group 20-24. The level of fertility implied by the retrospective reports of children ever born to this age groups is , therefore, 33% higher than that indicated by the current rates. Multiplication of the recorded f values and total fertility leads to the adjusted measures of the final column. The estimate of total fertility is 5.122 compared with the true level in the hypothetical population of 5.135. It is worth noting that the derived total fertility rate is about one third higher than the values obtained from either the observed current rates or the mean children ever born to women of completed fertility. The near equality of these two values, which occurs because the proportion of births in the preceding year not reported was made about the same as the omissions of children born by the older women, is no evidence of accuracy.

17. The adjustment factor for the level of fertility is derived from the mean children ever born to women aged 20-24 years. By this age group only a small proportion of the women who entered the reproduction period have died and the possible effects of differential fertility of the dead is slight. Since the great majority of the births to these women will

have occured within a few year of the census the estimate of fertility level will apply to a very recent experience even when there is a trend in rates. No allowance is made, however, for any class of births which is omitted equally by women of all ages in current and retrospective records. A possible example is children who die very young.

18. It can be shown that the B/A ratios are insensitive to age misstatements even although the retrospective and current mean births are badly distorted; the errors in the two components of the ratio tend to be compensating. If too many or few women are reported to be in the reproductive period, however, the calculated specific rates and the total fertility ratio will be wrong even if the number of births is correct. Reported ages in many African censuses show a tendency towards overstatement in the years near the beginning of the reproductive period and understatement at the end, with a characteristic crowding towards the centre. The correction factor from the B/A ratio for the age group 20-24 years can still be used but it should not now be applied directly to the total fertility but to the number of births reported for the preceding year. A corrected birth rate can then be calculated or (more conveniently for analysis purposes) a female rate, where the birth and population numbers are for one sex only. If the female population is quasi-stable (see below) the relation between the gross reproduction ratio and the female birth rate depends strongly on the mean m of the fertility distribution but little on the shape of the mortality curve. The approximate translation from one to another of these measures can be made by the use of table 3.

Table 3. Gross reproduction ratios Rg. corresponding to female birth rates b in a quasi-stable population

m in	b(rates per thousand)								
years	15	20	25	30	35	40	45	50	55
26.2	1.07	1.33	1.59	1.87	2.16	2.47	2.80	3.16	3.54
28.2	1.07	1.34	1.63	1.94	2.26	2.61	2.99	3.41	3.87
30.2	1.07	1.35	1.67	2.00	2.37	2.77	3.21	3.69	4.22

III. Estimation of Mortality

19. We will be concerned in this section with the extraction of information on the incidence of deaths from the census records. In particular a method is developed for the estimation of mortality in childhood and early adult years from reports by mothers of the number of their children who have died. It will be assumed that specific fertility and death rates have remained constant for the required age range and time period; the experience of the surviving women will also be taken to be effectively that of the total numbers exposed to the risk of births and deaths of children. The consequences of deviations from these assumptions will be examined later. Throughout the presentation the description will be in terms of vital events for children of both sexes but the results hold if observations for males and females are treated separately.

20. We will denote the proportion of children ever born who have died previous to the census by D_1, D_2 and so on for the age groups of women 15-19, 20-24 etc., and the function $q(t)$ will be defined as the probability that a child will be dead by age t. $q(t)$ is equal to $1-l(t)$ where $l(t)$ is the proportion of persons surviving to age t in the life table representing the mortality of the population. The women in an age group at the census will have borne their children at various lengths of time previously but the bulk of the births will have taken place around the ages when specific fertility rates were highest; the interval from this age region to that of the group at the date of recording will be approximately the average period from birth for which the children have been exposed to the risk of death. For each D_i, therefore, there will be an interval of time t_i such that $D_i = q(t_i)$, i.e. t_i is the equivalent period of exposure to risk for children of this age group of mothers. D_i will fall and t_i increase as the mothers, and consequently their children on average, grow older.

21. The locations of specific fertility distributions and, therefore, the equivalent exposed to risk period of the children of mothers in any age group, vary greatly among populations. This must be taken into account in the determination of the t_i. For a fixed location of the fertility curve alterations in the shape will modify the distributions of the exposed to risk intervals; variations in the curve of mortality will have a similar effect. However, the changes in the distributions arising from plausible alternative curves are not large. In addition the very heavy mortality of the first year or two of life is followed by much lower and only gradually changing death rates. At these later ages small variations in the distribution of children by exposed to risk period have little relative effect on the values of D. It can be surmised that a method of estimation of equivalent exposed to risk periods which allows only for population differences in the location of the fertility distribution will give adequate results.

22. As a result of these considerations the following procedure was devised. A standard life table, giving mortality measures at single years of age, was selected. This life table, which is an "average" of observations for populations with high death rates, forms the basic pattern for the model system, which is described in the next section. Fertility was taken to conform with the polynomial function model used above for graduation purposes. The D values for the usual five year age groups of mothers, corresponding to the standard life table and locations of the fertility model at one year intervals, were computed directly by numerical methods. Each equivalent exposed to risk period t_i was then determined by noting the age on the q(t) curve at which the proportion dead was equal to D_i.

23. The values of t_i obtained in this way were inconvenient for direct application because they were not, in general, whole numbers of years. It was noted, however, that (a) for each age group of mothers there was a $\bar{t}_i$ interval in whole years, corresponding to a location of the fertility distribution within the range of recorded values and (b) except for the youngest age group the rate of change in the relation of q $(\bar{t}_i)$ to D_i with location was small. As a result it was possible to adjust the D_i by relatively small amounts to equal $q(\bar{t}_i)$. For each age group of mothers the value of $q(\bar{t}_i)$ for the appropriate whole number of years was taken from the standard life table. The ratio $q(\bar{t}_i)/D_i$ was computed for the different locations of the model fertility distribution to give sets of multiplying factors. These are presented in table 4. By the same procedure factors for ten year groups were derived and are also shown in the table.

24. Multiplication of the proportion of children born who had died previous to the census by the appropriate factor for the age group of mothers and location of the fertility distribution gives an estimate of $q(\bar{t}_i)$, the life table probability of dying before the specified whole number age. Each row of the table refers to the age group shown in the first column and the q measure in the second. The factor in the row, which corresponds best to the location of the fertility distribution estimated from the population records, is chosen. Two measures specifying the location are shown in the middle of each column of the table. One is m, the mean age of the specific fertility distribution, which has already been used for the same purpose in table 1. The other is the ratio r_1/r_2, where r_1 and r_2 are the total children ever born per woman in the age groups 15-19 years and 20-24 years respectively. For some populations only a value for r_1/r_2

Table 4. Multiplying factors for the derivation of life table q values from proportions of children dead by age group of mothers

Age group of mother in years	Mortality measure +				Five year age groups				
15-19*	$_{1}q_{0}$	0.859	0.890	0.928	0.977	1.041	1.129	1.254	1.425
20-24	$_{2}q_{0}$	0.938	0.959	0.983	1.010	1.043	1.082	1.129	1.188
25-29	$_{3}q_{0}$	0.948	0.962	0.978	0.994	1.012	1.033	1.055	1.081
30-34	$_{5}q_{0}$	0.961	0.975	0.988	1.002	1.016	1.031	1.046	1.063
35-39	$_{10}q_{0}$	0.966	0.982	0.996	1.011	1.026	1.040	1.054	1.069
40-44	$_{15}q_{0}$	0.938	0.955	0.971	0.988	1.004	1.021	1.037	1.052
45-49	$_{20}q_{0}$	0.937	0.953	0.969	0.986	1.003	1.021	1.039	1.057
50-54	$_{25}q_{0}$	0.949	0.966	0.983	1.001	1.019	1.036	1.054	1.072
55-59	$_{30}q_{0}$	0.951	0.968	0.985	1.002	1.020	1.0[illegible]	1.058	1.076
60-64	$_{35}q_{0}$	0.949	0.965	0.982	0.999	1.016	1.034	1.052	1.070
r_1/r_2		0.387	0.330	0.268	0.205	0.143	0.090	0.045	0.014
m(years)		24.7	25.7	26.7	27.7	28.7	29.7	30.7	31.7
					Ten year age groups				
15-24*	$_{2}q_{0}$	0.982	1.000	1.021	1.045	1.072	1.[illegible]	1.144	1.193
25-34	$_{5}q_{0}$	0.990	1.004	1.01[illegible]	1.033	1.048	1.064	1.081	1.099
35-44	$_{15}q_{0}$	0.977	0.993	1.009	1.024	1.040	1.056	1.071	1.086
45-54	$_{25}q_{0}$	0.990	1.008	1.025	1.043	1.062	1.080	1.099	1.118
55-64	$_{35}q_{0}$	0.990	1.007	1.025	1.043	1.061	1.080	1.099	1.119

* It has been assumed in the calculations that births and deaths of children of mothers under fifteen years have been re-allocated to these age groups but the effect on the estimated mortalities is very small.

+ The standard notation $_{t}q_{0}$ has been used for q(t) or m will be available or acceptable. When both can be used the factors obtained from the two measures will not usually agree. For the older women the mean m specifies most accurately the age region in which most of their children were born and, therefore, the equivalent exposed to risk interval. For the young women, the shape of the fertility curve at the beginning of the reproductive period, measured by r_1/r_2, is more important. The rule proposed is that the observed r_1/r_2 should be used to estimate the factors for the first three age groups and m for the older women.

Table 5. Example to illustrate the estimation of mortality measures from the proportion of children dead: by age group of mother: Republic of Guinea sample census 1954-55.

Age group of mothers in years	Proportion of children dead	Multiplying factors from: r_1/r_2	m	Exact age t	$q(t)$	$l(t)$
15-19	.224	0.903	0.958	1	.202	.798
20.24	.299	0.967	1.000	2	.289	.711
25.29	.354	0.967	0.988	3	.342	.658
30.34	.379	0.979	0.997	5	.378	.622
35.39	.401	0.987	1.005	10	.403	.597
40.44	.429	0.960	0.982	15	.421	.579
45.49	.448	0.958	0.980	20	.439	.561
50.54	.578	0.972	0.994	25	.475	.525
55.59	.484	0.974	0.996	30	.482	.518
60.64	.505	0.971	0.993	35	.501	.499

25. The use of the factors is illustrated by application to records collected in the 1954-55 sample census of the Republic of Guinea. The proportions of children ever born reported as dead, by age of mothers, are shown in the second column of table 5. The retrospective births per woman in the two youngest age groups, r_1 and r_2, were 0.54 and 1.75 giving 0.309 for r_1/r_2. The mean age of the specific fertility distribution, calculated from the rates found from births in the year preceding

the survey, was 27.32. From table 4 the multiplying factors corresponding to 0.309 for r_1/r_2 are found by interpolating linearly between the second and third columns of entries and those corresponding to the mean of 27.32 by interpolating between the third and fourth columns, with the results shown. The differences between the two sets of factors in the region of transition from one to the other are less than 2%. The application of the first three factors from the r_1/r_2 column and the rest from the m column to the recorded D proportions gives the q values in the table. Subtraction of these fromone leads to the estimates of the life table survivorship ratios in the final column. The first eight ratios show a normal mortality pattern with little evidence of systematic error but thelast two are inconsistently high.

26. Extensive investigations have shown that the factors give satisfying estimates of mortality, if the assumptions hold, despite variations of the true fertility and mortality patterns from the model ones; the method is also insensitive to age mis-statements. There are, however, several important types of error or deviations from assumptions whose effects are not only difficult to isolate but also to detect at all.

(a) Children whose mothers were in the relevant cohorts but died before the survey date are not represented in the records. The estimates would be biased if the mortality of this category of children differed from the average.

(b) If mortality is changing with a consistent trend the children of mothers in a particular age group will have experienced death rates which are some kind of average of those in operation from the birth date of the oldest to the time of the survey. Because a very high proportion childhood deaths occur in the first year or two of life the level of mortality in the period when the bulk of the births took place will be dominant.

(c) Commonly in censuses the proportion of ommissions in retrospective reports of children ever born increases with age of woman. It has been suggested that most of the omissions are of dead children but for many populations, this conclusion can not be reconciled with the observations. However, if there is differential reporting the proportions recorded as dead will be biased.

(d) If any class of dead children was left out equally by all women (e.g. births surviving for only a short time) there would be a roughly constant absolute error in the estimated q probabilities of dying.

27. Without external checks on the accuracy of the data it seems impossible to establish a systematic procedure for assessing which, if any, of these effects are appreciably falsifying the estimates. The materials for establishing the level and structure of mortality over a range of ages are weak. It is important to note, however, that few of the women in the younger cohorts will have died between the start of reproduction and the census, their children have been born very recently, and omissions are less likely. Of the possible effects listed the only one which could have an important influence on the estimate of early mortality is the under reporting of children who only survived a short time. There is good reason to suppose, therefore, that these estimates will be near minimum. The best guide will come from the observations for women aged 20-34 years; the probability of dying under one, calculated from the data for the youngest group, is unreliable because of sensitivity to variations in the patterns of fertility and mortality.

28. From the census records of the number of deaths in the year preceding, current mortality rates can be derived and compared critically with the retrospective measures. The reported deaths are treated as if they were the number registered in the period and the population at risk taken from the census. Because of errors in the age distributions of both the population and deaths it is best to calculate the specific rates

for five or ten year age groups. A current abridged life table can then be constructed by standard procedures (e.g. Barclay (1958)).

29. The probabilities of dying by different ages, or the complementary survivorship ratios l(t), from the current abridged life tables are compared with the corresponding retrospective values over the range which is common to both. If agreement is good the current life table (perhaps after smoothing) can be accepted with some confidence. The divergences of the two sets of measures are often substantial. Their reconciliation and the use of the joint evidence to estimate a life table raise problems. Profitable comparisons can only be made at a few points since the values at one year are unreliable and there are normally no current measures directly corresponding to the key retrospective estimates at ages two and three. Random and erratic fluctuation confuse the results. When the mortality levels estimated from the two types of records do not agree the deviations at different ages can only be assessed, in relation to possible determining errors, by reference to the expected patterns of death rates. For these reasons, the measures can best be studied with the help of model life tables. A system for this purpose, and for the application of quasi-stable population theory is developed in the next section.

IV. Use of a Model Life Table System

30. Whenever information is lacking or suspect about a basic demographic measure of a population it is natural to try to overcome the difficulty by a "guess" based on the indices of other countries. Demographic models are a systematisation of this approach, in which functions are constructed to represent the "average" experience of a number of populations. Models may be expressed as tabulated values, computed from actual observations, by mathematical relations or, as here, by a combination of the two methods. They can be classified by the number of parameters, or degrees of freedom, needed to specify them.

The best known system of model life tables is that produced by the United Nations. The mortality schedules of this system for each sex are a one parameter set, i.e. fixed by a single measurement, computed from the average experience of a large number of life tables. The measures of retrospective and current mortality calculated from the observations by the procedures of the previous section could be reconciled and graduated by the use of the United Nations model tables. However, recent studies have shown that in many applications a one parameter model is not good enough since, for the same general level of mortality, patterns of death by age may differ in important ways.

31. In the development of the mortality model described here, the requirements for the analysis of records from censues were a major consideration. The system has two parameters and is, therefore, more flexible than the United Nations model. In addition, the description is in terms of mathematical functions which are very convenient for applications. The construction of the system stems from the observation that, for varying ages t, the survivorship ratios of two separate life tablesare approximately linearly related on a logit scale; the logit is defined as $Y(t)=\frac{1}{2}\log_e \frac{1-l(t)}{l(t)}$, i.e. half the natural logarithm of the quotient of the proportions dying and surviving. There are several published tables for obtaining the logits of proportions, including one given by Fisher and Yates (1963). There is nothing very strange in the use of the logit transformation for the analysis of mortality data. As Cox (1958) has pointed out the logit scale has much the same relation to proportions as the usual linear scale has to measurements which are not bounded. When $l(t)$ varies from one to zero Y (t) ranges from minus to plus infinity. A straight line relationship in logits is thus an obvious possibility for study when the measures are proportions.

32. If one life table with survivorship ratios $l_s(t)$ is taken as a standard, the linear relation is $Y(t) = \alpha + \beta \quad Y_s(t)$. Here $Y_s(t)$ and $Y(t)$ are the logits for the standard and any other life table respectively with and as two constants. Roughly the parameter α can be taken to measure the level of childhood mortality and β the "steepness" of the increases of death rates with age. As α and β vary the life tables of the model system are generated. If the standard life table measures $l_s(t)$ are chosen to be an average of observed ratios a set of mortality schedules which broadly apply to all populations is obtained. A general standard of this kind has been constructed. However, as recent papers have shown (Coale (1962), United Nations (1962)) particular groups of life tables show distinctive mortality patterns. Greater accuracy can, therefore, be achieved if it is possible to choose a standard which has the same detailed characteristics as the mortality configuration of the population which is studied. An "African Standard" life table has been developed based on experience in the use of the model system for the analysis of mortality records of African communities. In this lif table the infant mortality is relatively lower but subsequent childhood mortality higher than in the general standar. The survivorship ratios and corresponding logits of the African standard mortality schedule are shown for selected ages in table 6.

33. In application of the model system the aim is to estimate and from the observations by a procedure which is dictated by the nature and accuracy of the data. When and are known $Y(t)$, and hence $l(t)$, can be calculated at the specified ages from the linear equation with $Y_s(t)$. The complete life table is, therefore, constructed by the estimation of two parameters.

Table 6. Survivorship ratios $l_s(t)$ and corresponding logits at specified ages of the African Standard life table

Exact age in years t	$l_s(t)$	Logit $Y_s(t)$ (negative)	Exact age in years t	$l_s(t)$	Logit $Y_s(t)$ (negative)	Exact age in years t	$l_s(t)$	Logit $Y_s(t)$ (positive)
1	.8322	0.9970	20	.7130	0.4551	55	.4585	0.0832
2	.8335	0.8052	25	.6826	0.3829	60	.3965	0.2100
3	.8101	0.7252	30	.6525	0.3150	65	.3210	0.3746
4	.7964	0.6819	35	.6223	0.2496	70	.2380	0.5818
5	.7863	0.6615	40	.5898	0.1817	75	.1500	0.8673
10	.7502	0.5498	45	.5535	0.1073	80	.0760	1.2490
15	.7362	0.5131	50	.5106	0.0212	85	.0310	1.7211

34. The justification of the model must be in the power and elegance of its application; these can not be examined in this brief paper. It is important, however, to note some of the conclusions which have been drawn from the uses of the model in the analysis of African census records. To the degree of accuracy of the observations and the evidence of basic patterns the same standard schedule can be adopted for males, females and the two sexes together although the estimated values of the parameters will, in general, differ. The system is designed to describe mortality patterns in childhood and up to middle adult years as accurately as possible; information on African death rates in old age is almost non-existent. It appears that in most populations β is not very different from its central value of 1.0 and generally lies between 0.8 and 1.2. If β is fixed at 1.0 for all populations we obtain a reduced model, with α as the one variable parameter, which is in good agreement with the United Nations system.

35. To reconcile and graduate the survivorship ratios, calculated from the retrospective and current observations by the methods of the last section, their logits $Y(t)$ are first computed. These values are plotted on the y axis of a graph against the corresponding $Y_s(t)$ measurements for the same age t, which form the scale of the x axis. If the retrospective and current logits lie approximately on the same straight line (even if individual points are erratic) there is satisfactory agreement. A straight line can be fitted by eye or more elaborate methods; the required values of α and β are estimated from the intercept on the y axis and the slope of the line respectively.

36. In general, systematic differences between the retrospective and current logits will be apparent from the graph. It is not possible to lay down absolute rules for reconciling these since the best procedure will depend on the nature of the deviations and other relevant information about the population and methods of collecting the data. The main principle is that the retrospective probabilities of dying at ages 2, 3 and 5 are likely to be the most accurate measures and that any fitted line should be very close to them. Commonly, it is found that the current logits are displaced from the retrospective at early ages because the child death rates of the former are substantially lower. The current rates in childhood should then be increased to make the corresponding logit at age five years, accord with level for the retrospective values at 2, 3 and 5.

37. To correct the current measures at later ages it is necessary to make some reasonable assumptions. One possibility is to accept the adjustment made to rates under five years as an index of under-recording and to increase deaths at all ages to the same extent. However, the accuracy and completeness of death recording may be very much influenced by the importance of an individual in the community and hence may vary with sex and age. It might be reasonable to assume that the errors in current reports were confined to young children and that the specific death rates after age five were approximately correct. In general, the best

41. Stable age distributions, which are in approximate agreement, have been calculated for very diverse mortality schedules of the model system by suitable choice of the intrinsic rates of natural increase. Characteristic measurements of the stable populations and percentages under specified ages are shown in table 7. The values of bl(2) for the different populations are in close agreement; this is a general finding for quasi-stable age distributions with approximately the same configuration. It can be concluded that the check of the stable against the observed age distribution gives good evidence about the accuracy of the estimates of the birth rate and childhood mortality, except for the possible omission of children who died very young and were not reported either as births or deaths. It gives no guidance on the level of adult mortality or natural increase.

Table 8. Corresponding deviations in percentages and years of age from fitted stable age distributions with given values of b l(2)

Age in years	bl(2) per thousand: 15 %	15 yrs.	20 %	20 yrs.	25 %	25 yrs.	30 %	30 yrs.	35 %	35 yrs.	40 %	40 yrs.
5	2.44	1.67	2.82	1.53	3.07	1.40	3.23	1.28	3.34	1.17	3.40	1.09
10	4.45	3.04	4.96	2.78	5.28	2.55	5.42	2.33	5.47	2.15	5.46	2.00
15	6.09	4.15	6.58	3.83	6.83	3.55	6.82	3.29	6.77	3.05	6.59	2.84
20	7.43	5.04	7.73	4.67	7.79	4.37	7.63	4.09	7.37	3.82	7.01	3.58
25	8.31	5.67	8.44	5.35	8.27	5.05	7.90	4.77	7.44	4.50	6.91	4.34
30	8.90	6.12	8.72	5.83	8.32	5.57	7.75	5.29	7.12	5.03	6.46	4.78
35	9.15	6.35	8.63	6.09	7.99	5.85	7.28	5.63	6.53	5.40	5.79	5.16
45	8.44	6.15	7.49	6.03	6.55	5.93	5.67	5.78	4.84	5.63	4.10	5.47
55	6.49	5.30	5.40	5.34	4.44	5.32	3.66	5.28	2.97	5.22	2.40	5.13
65	3.82	4.14	2.94	4.19	2.28	4.25	1.7[illegible]	4.27	1.37	4.27	1.05	4.21

42. In many African censuses the reported age distributions have been so erratic that the range of stable populations which can be claimed to be in reasonable agreement is wide. The direct comparison of the percentages below specified ages in the stable and observed distributions is not very illuminating. A method for examining the results which is more immediately comprehensible can be applied by the use of table 8. For a range of values of the index b1(2) the table shows the deviations in years corresponding to differences in the percentages of the population under the specified ages. The table can be used in two ways. The columns for the b1(2) value of the constructed stable population are chosen, with interpolation if necessary. The differences between the stable and observed distributions in the percentages under each age are calculated. The ratios of the percentage differences to those shown in the chosen column for the given ages are multiplied by the corresponding deviation in years. The values obtained are the average biases in the reported age statements at these points if the stable distribution is correct. For example, a positive deviation of 2.1 years at age five means that the children reported as under five were really on average under 7.1 years if the stable distribution is correct. The tables can also be used to find the deviations for another alternative stable population which is taken to differ from the constructed one by a given percentage at some age. The ratio of the percentage to the appropriate entry at the age is found in the same way as above and applied to <u>all</u> the yearly deviations in the column. The differences between these values and the corresponding ones from the reported distribution, taking account of sign, give the deviations for the alternative stable population. The calculations are approximate but provide a quick and vivid method for examining the errors in age reports implied by different assumptions.

43. The application of these preocedures to records for Guinea is illustrated in table 9. Columns 2, 3 and 4 show the observed and constructed stable distribution percentages under the ages in the first column and their differences. In the fifth and sixth columns are the percentage and age deviations, for the bl(2) value of 32.5 per thousand of the stable population, found from table 8 by linear interpolation. The second last column gives the age deviations for the differences between observed and stable percentages, calculated to be in the same ratios to the entries in columns 6 as the percentages in column 4 bear to those in column 5. The maximum absolute value of the latter ratios is approximately one half (positive at age ten and negative at 20 years). The last column shows the yearly deviations corresponding to this ratio, i,e, the values for the most extreme stable distributions which agree with the observed percentages at some age. If the deviations in the last two columns are plotted in a graph against age the implications of variations in the stable population in terms of age mis-statements are clearly seen.

Table 9. Example to illustrate the calculation of age deviations of observed and possible distributions from the constructed stable population: Republic of Guinea sample census 1954-55

	Percentage under age			bl(2)=32.5		Yearly deviations of	Maximum yearly
Age	Observed	Stable	Differences	%	Years	observations	deviation
5	17.10	16.00	1.10	3.28	1.23	0.41	0.62
10	31.40	28.69	2.71	5.44	2.24	1.12	1.12
15	38.90	39.90	-1.00	6.80	3.17	-0.47	1.59
20	46.20	49.97	-3.77	7.51	3.96	-1.99	1.99
25	56.00	58.68	-2.68	7.68	4.64	-1.62	2.32
30	65.80	66.26	-0.46	7.44	5.16	-0.32	2.58
35	73.10	73.01	0.09	6.91	5.52	0.07	2.76
45	85.30	83.95	1.35	5.26	5.71	1.47	2.86
55	92.60	91.77	0.83	3.32	5.25	1.31	2.62
65	96.90	96.74	0.16	1.59	4.27	0.43	2.13

Note

44. Polynomial functions are often used to graduate observations because of the ease with which the mathematical manipulations can be performed. The number of unknown parameters was reduced by the imposition of appropriate

restrictions to give the following function which is satisfactory for the applications in this study:

$$f(t) = C(t-s)(s + 33 - t)^2, \quad s \leq t \leq s + 33$$

where f(t) is the specific fertility rate of women aged t years, s is the start of the reproductive period and C is a constant which varies with the level of fertility; f(t) is taken as zero when t is outside the range s to s + 33 years. The mean of the function is s + 13.2 years and the 33 year range was chosen to make the variance (43.6) close to the average for observed distributions.

References

Barclay, G.W., Techniques of Population Analysis, John Wiley and Sons, New York (1958).

Bourgeois-Pichat, J., Utilisation de la notion de population stable pour mesurer la mortalité et la fécondité des populations des pays sous-développées, Bulletin of the International Statistical Institute, 36(2), 94-121 (1958).

Brass, W., Methods of obtaining basic demographic measures where census and vital statistics registration systems are lacking or defective, background paper prepared for 1965 World Population Conference (1964).

Coale, A.J., The effects of changes in mortality and fertility on age composition, Milbank Memorial Fund Quarterly, 34(1), 79-114 (1956).

Coale, A.J., Estimates of various demographic measures through the quasi-stable age distribution, in Emerging Techniques in Population Research, Proceedings of the annual conference of the Milbank Memorail Fund, 175-193 (1962).

Cox, D.R., The regression analysis of binary sequences, Journal of the Royal Statistical Society, B,20(2), 215-242 (1958).

Fisher, R.A. and Yates, F., Statistical Tables for Biological, Agricultural and Medical Research (6th edition), Oliver and Boyd, London (1963).

International Union for the Scientific Study of Population, Vital Statistics from limited data, Proceedings of 1963 International Population Conference, 19-244, Liège (1964).

United Nations, Factor analysis of sex-age-specific death rates: a contribution to the study of the dimensions of mortality, Population Bulletin, 6, 147-201 (1962).